Making Use of Human Knowledge in Machine Learning

Image-based Classification from Tissue to Trees

Ute Schmid
with Bettina Finzel and Christoph Wehner
Cognitive Systems Group, University of Bamberg

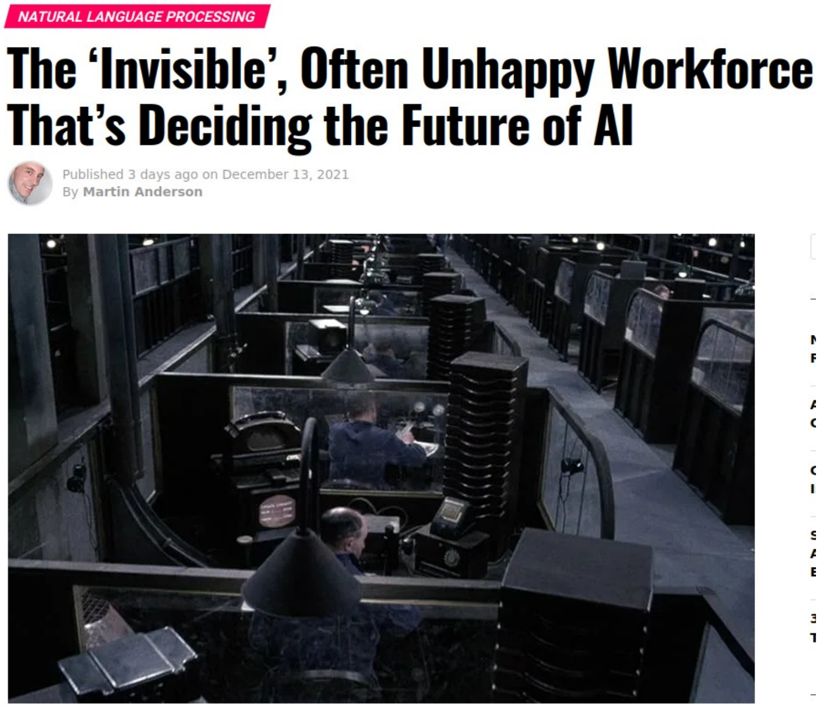
HCAI Lab Symposium
“Digital Transformation in Smart Farm and Forest Operations”

BOKU Wien, Aug 22 2022

The Three Waves of AI

- **1st Wave:** Focus on explicit representation of knowledge
  - Powerful algorithms with provable characteristics
  - But: A large amount of human knowledge is implicit, i.e. not available to inspection and verbalisation (Polyani's Paradox)
- **2nd Wave:** Focus on data-intensive machine learning
  - Impressive results, e.g. for image classification (mostly implicit perceptual knowledge)
  - But: high demands on amount and quality of data ("garbage in garbage out")
  - Labeling of training data in specialized domains demands high expertise (medical diagnostics, quality control)

Data Engineering Bottleneck – the next AI winter?


NATURAL LANGUAGE PROCESSING

The 'Invisible', Often Unhappy Workforce That's Deciding the Future of AI

Published 3 days ago on December 13, 2021
By Martin Anderson

Nuremberg Funnel, 1910

https://de.wikipedia.org/

Polanyi's Revenge

"Human, grant me the serenity to accept the things I cannot learn, data to learn the things I can, and wisdom to know the difference."

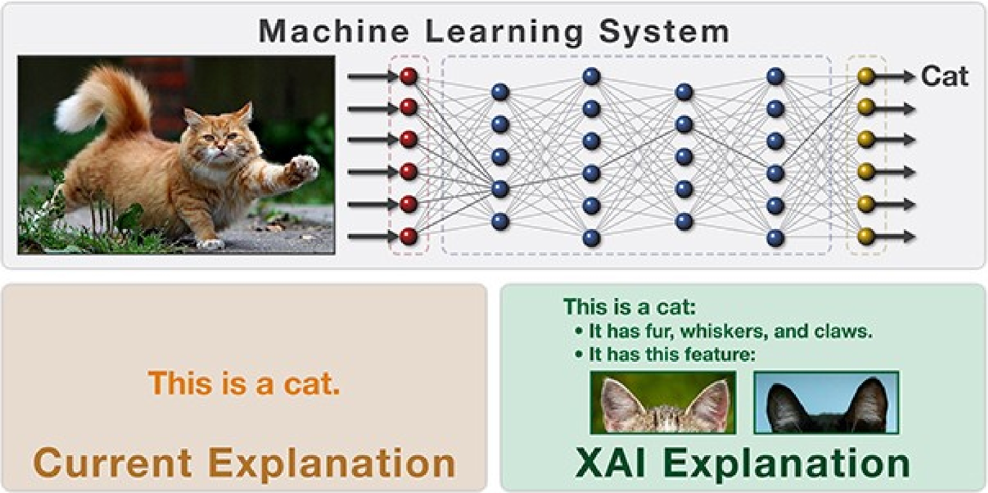
(Subbarao Kambhampati, Communications of the ACM, February 2021)

3rd Wave of AI: XAI

But also:

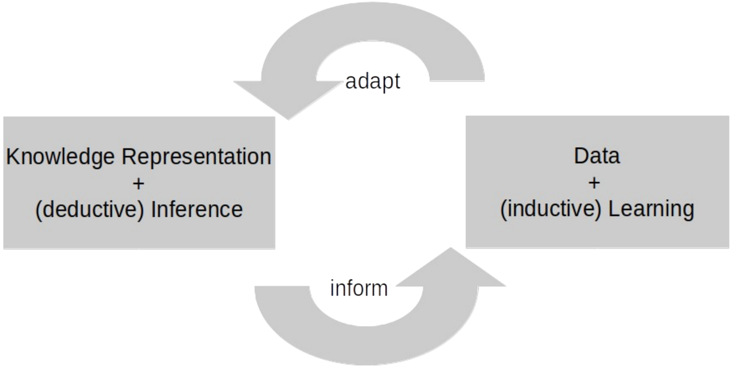
- Hybrid approaches
- Interactive ML
  - Recent advances have made AI synonymous with learning from massive amounts of data, even in tasks for which we do have explicit theories and hard-won causal knowledge!
  - Knowledge is injected in deep learning through architectural biases and carefully manufactured examples

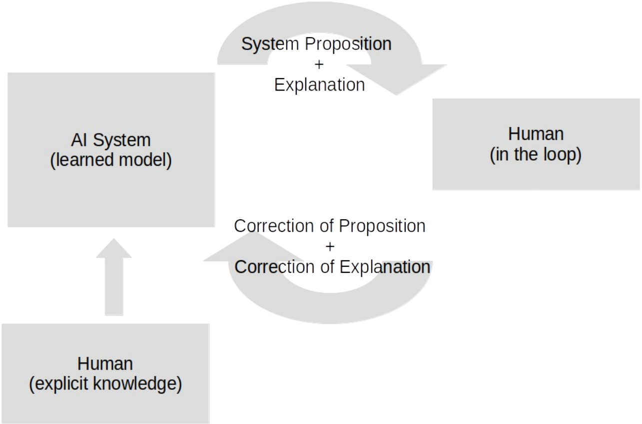
3rd Wave of AI: Explainable AI (XAI)
Hybrid, explanatory, interactive, human-centric



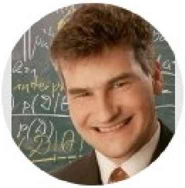
David Gunning, IJCAI 2016

http://www.darpa.mil/program/explainable-artificial-intelligence





Pioneer of Interactive Machine Learning



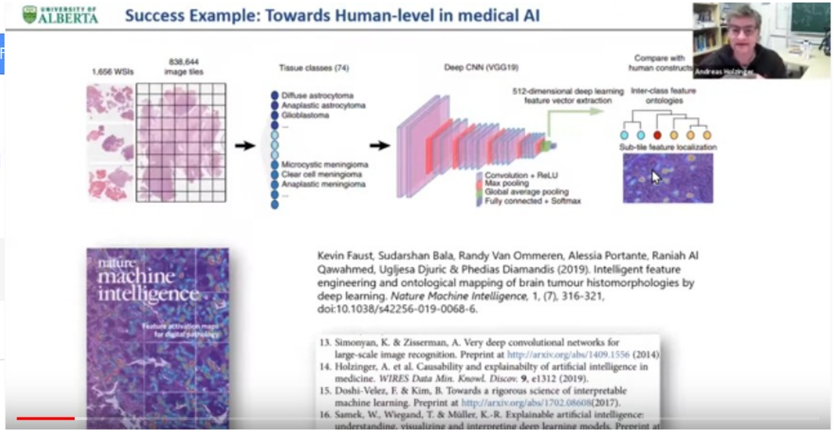
Andreas Holzinger

Human-Centered AI Lab, University of Natural Resources and Life Sciences, Vienna, Austria
Bestätigte E-Mail-Adresse bei boku.ac.at - Startseite

Human-Centered AI   Explainable-AI   interactive Machine Learning   Decision Support   trustworthy AI

ARTIKEL   ZITIERT VON   ÖFFENTLICHER ZUGRIFF   KOAUTOREN

TITEL	ZITIERT VON	
Usability engineering methods for software developers A Holzinger Communications of the ACM 48 (1), 71-74	1314	
Successful implementation of user-centered game based learning in higher education: An example from civil engineering M Ebner, A Holzinger Computers & education 49 (3), 873-890	891	2007
Interactive machine learning for health informatics: when do we need the human-in-the-loop? A Holzinger Brain Informatics 3 (2), 119-131	675	2016
Causability and explainability of artificial intelligence in medicine A Holzinger, G Langs, H Denk, K Zatloukal, H Müller Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 9 (4 …	584	2019
What do we need to build explainable AI systems for the medical domain? A Holzinger, C Biemann, CS Pattichis, DB Kell arXiv preprint arXiv:1712.09923	556	2017



In the medical domain

Pioneer of Interactive Machine Learning

Network module detection from multi-modal node features with a greedy decision **forest** for actionable explainable AI
[PDF] arxiv.org
B Pfeifer, A Saranti, A **Holzinger** - arXiv preprint arXiv:2108.11674, 2021 - arxiv.org
Network-based algorithms are used in most domains of research and industry in a wide variety of applications and are of great practical use. In this work, we demonstrate subnetwork …
☆ Speichern 99 Zitieren Zitiert von: 4 Ähnliche Artikel Alle 4 Versionen

Federated Random **Forests** can improve local performance of predictive models for various healthcare applications
[HTML] oup.com
…, T Frisch, O Zolotareva, A **Holzinger**… - …, 2022 - academic.oup.com
Motivation Limited data access has hindered the field of precision medicine from exploring its full potential, eg concerning machine learning and privacy and data protection rules. Our …
☆ Speichern 99 Zitieren Ähnliche Artikel Alle 5 Versionen

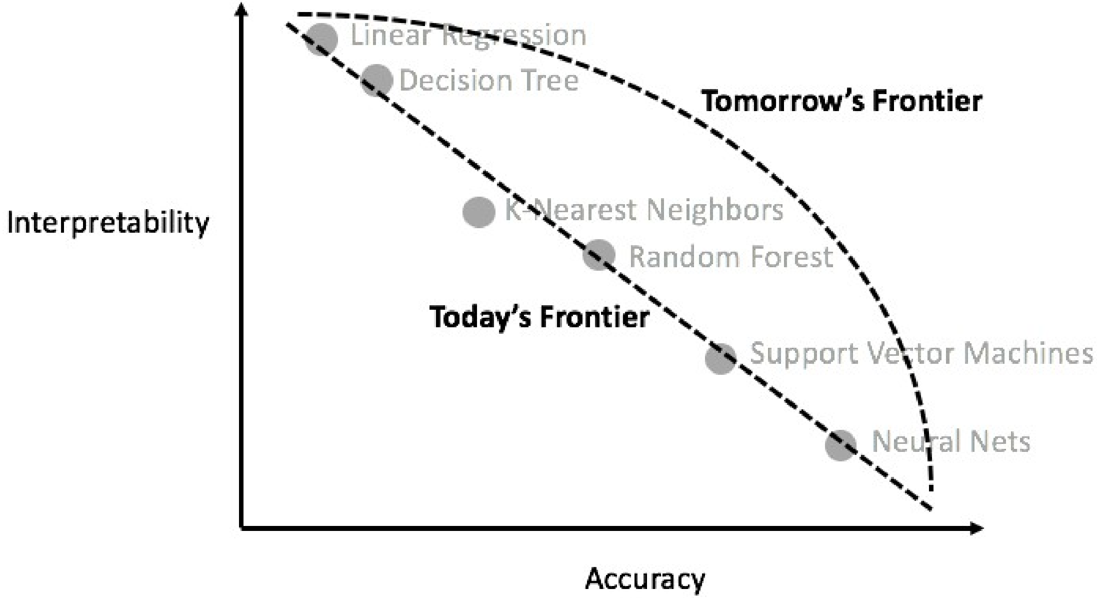
[PDF] Digital Transformation in Smart Farm and **Forest** Operations Needs Human-Centered AI: Challenges and Future Directions
[PDF] mdpi.com
A **Holzinger**, A Saranti, A Angerschmid, CO Retzlaff… - Sensors, 2022 - mdpi.com
The main impetus for the global efforts toward the current digital transformation in almost all areas of our daily lives is due to the great successes of artificial intelligence (AI), and in …
☆ Speichern 99 Zitieren Zitiert von: 3 Ähnliche Artikel Alle 11 Versionen

Machine Learning and Knowledge Extraction to Support Work Safety for Smart **Forest** Operations
…, A Nothdurft, P Kieseberg, A **Holzinger**… - … -Domain Conference for …, 2022 - Springer
… Random **Forest**. A random **forest** was also trained in a way similar to the decision tree as far as … Random **forests** are not considered interpretable, but have generally better performance …
☆ Speichern 99 Zitieren

Biodiversity of *Klebsormidium* (Streptophyta) from alpine biological soil crusts (Alps, Tyrol, Austria, and Italy)
[PDF] wiley.com
T Mikhailyuk, K Glaser, A **Holzinger**… - Journal of …, 2015 - Wiley Online Library
… (<1,800 m above sea level; asl) in the pine-**forest** zone. Strains of clades B/C, D, and F … soil

For smart farm and forest operations

Predictive Accuracy & Comprehensibility of Models/Decisions



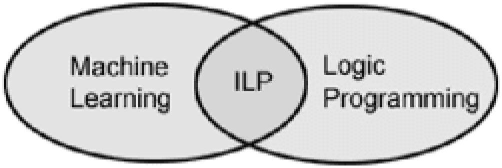

PERSPECTIVE
https://doi.org/10.1038/s42256-019-0048-x
nature machine intelligence

Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead

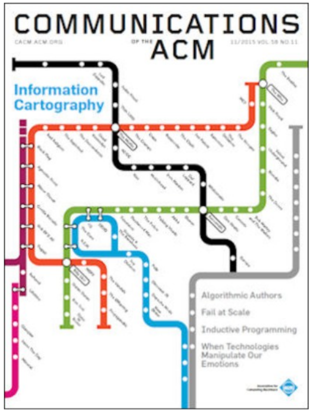
Cynthia Rudin

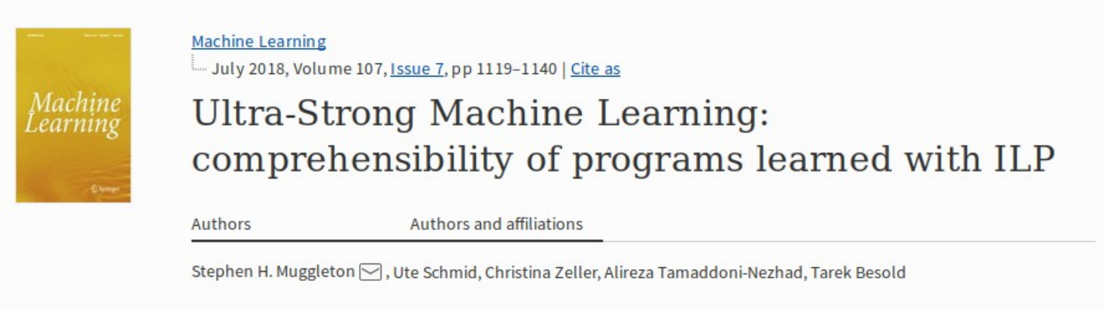
Inductive Logic Programming (ILP)

- Training examples, background knowledge, learned models are all represented as Horn clauses



Gulwani, Hernandez-Orallo, Kitzelmann, Muggleton, Schmid, Zorn, Inductive Programming meets the real world, *CACM 58(11)*, 2015




Machine Learning

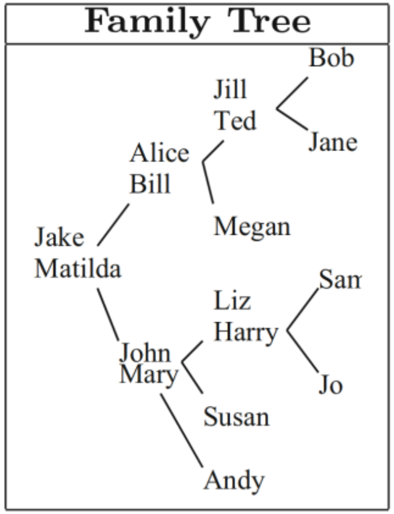
July 2018, Volume 107, Issue 7, pp 1119–1140 | Cite as

Ultra-Strong Machine Learning: comprehensibility of programs learned with ILP

Authors | Authors and affiliations

Stephen H. Muggleton, Ute Schmid, Christina Zeller, Alireza Tamaddoni-Nezhad, Tarek Besold

Example Family Tree




```
% Background Knowledge
father(jake,bill).    mother(matilda,bill).
father(jake,john).    mother(matilda,john).
father(bill,ted).     mother(alice,jill).
father(bill,megan).   mother(alice,ted).
father(john,harry).   mother(alice,megan).
father(john,susan).   mother(mary,harry).
father(ted,bob).      mother(mary,susan).
father(ted,jane).     mother(mary,andy).
father(harry,san).    mother(jill,bob).
father(harry,jo).     mother(jill,jane).
mother(liz,san).      mother(liz,jo).
```

```
% Examples
grandparent(matilda,megan).    not grandparent(megan,matilda).
grandparent(matilda,harry).    not grandparent(jake,jake).
grandparent(jake,susan).       not grandparent(matila,alice).

% Learned hypothesis    (parent can be background theory or invented)
grandparent(X,Y) :- parent(X, Z), parent(Z,Y).
parent(X,Y) :- father(X,Y).
parent(X,Y) :- mother(X,Y).
```

ILP Algorithms

Given a tuple (B, E^+, E^-) where:

- B denotes background knowledge
- E^+ denotes positive examples of the concept
- E^- denotes negative examples of the concept

An ILP algorithm returns a hypothesis $H \in \mathcal{H}$ such that:

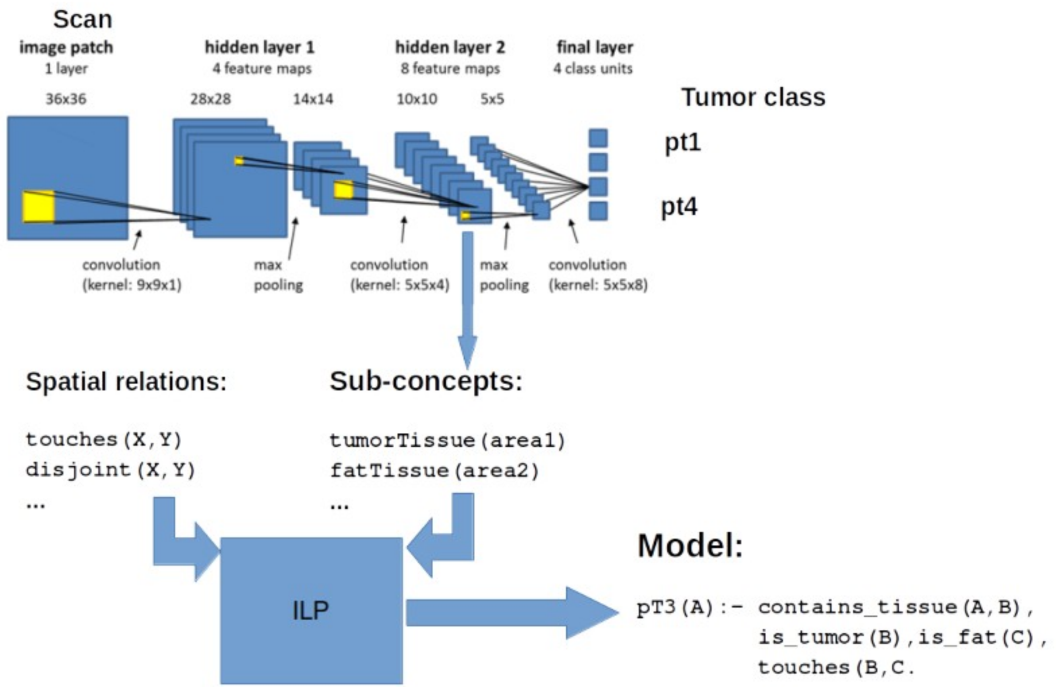
$$\forall e \in E^+, H \cup B \vdash e \text{ (i.e. H is complete)}$$

$$\forall e \in E^-, H \cup B \not\vdash e \text{ (i.e. H is consistent)}$$

- FOIL (Quinlan, 1990): Generate-and-test, sequential covering (ID3, C4.5, simulteneous covering by the same author)
- Golem, Progol, Aleph, Metagol (Muggleton, since 1990ies): learning from entailment in different variants
- Igor (Kitzelmann & Schmid, JMLR 2006; Schmid & Kitzelmann, CSR 2011): Inductive (functional) programming
- ProbLog (de Raedt, 2007): combining logical and statistical learning

Neuro-symbolic Integration

- Many recent approaches (de Raedt et al., IJCAI 2020 Survey)
- Combining learning for perceptual domains and interpretable ML
- Blackbox classifiers as sensors, whitebox classifieres as surrogate models



Picasso Faces

Table 1.
Results for ensemble embeddings with set IoU (sIoU), mean cosine distance to the runs (Cos.d.), and index of conv layer or block (L) (cf. Fig. 3).

		L	sIoU	Cos.d.
AlexNet	NOSE	2	0.228	0.040
	MOUTH	2	0.239	0.040
	EYES	2	0.272	0.058
VGG16	NOSE	7	0.332	0.104
	MOUTH	6	0.296	0.154
	EYES	6	0.350	0.197
ResNeXt	NOSE	6	0.264	0.017
	MOUTH	5	0.237	0.020
	EYES	7	0.302	0.020

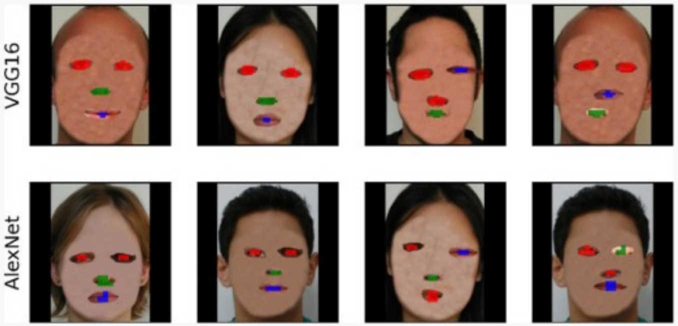


Fig. 4.
Ensemble embedding outputs of NOSE (green), MOUTH (blue), EYES (red). (Color figure online)

Table 2.
Learned rules for different architectures and their fidelity scores (accuracy and F1 score wrt. to the original model predictions). Learned rules are of common form `face(F):- contains(F, A), isa(A, nose), contains(F, B), isa(B, mouth), distinctPart`

Arch.	Accuracy	F1	Distinct rule part
VGG16	99.60%	99.60%	`top_of(A, B), contains(F, C), top_of(C, A)`
AlexNet	99.05%	99.04%	`contains(F, C), left_of(C, A), top_of(C, B), top_of(C, A)`
ResNext	99.75%	99.75%	`top_of(A, B), contains(F, C), top_of(C, A)`

Rabold, Schwalbe, Schmid, Expressive Explanations of DNNs by Combining Concept Analysis with ILP, KI 2020

Ultra-Strong Machine Learning

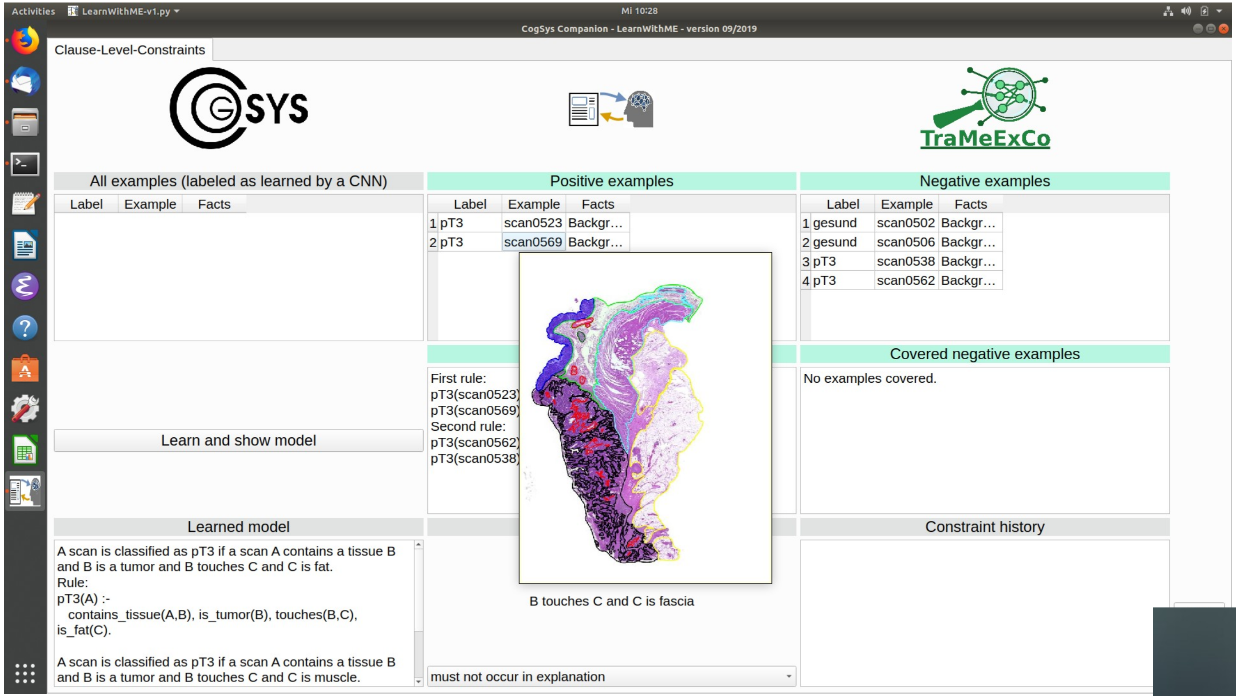
Michie (1988):

- Weak ML: machine learner produces improved predictive performance with increasing amounts of data
- Strong ML: additionally requires the learning system to provide its hypotheses in symbolic form (interpretable machine learning, e.g. Rudin, Nature ML, 2019)
- Ultra-strong ML: extends the strong criterion by requiring the learner to teach the hypothesis to a human, whose performance is consequently increased to a level beyond that of the human studying the training data alone

Human-AI Partnerships

- Keeping humans in-the-loop is not only ***ethical***
  - Transparence of AI decision making (white box)
  - Appreciation of human labor
- But also ***practical and necessary***
  - Providing explicit knowlege which contraints ML
  - Correcting model decisions for model adaptation (combining strengths of humans and AI methods)

Example 1: Decision Making in Medicine



GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

Schmid, U., & Finzel, B. (2020). Mutual explanations for cooperative decision making in medicine. KI-Künstliche Intelligenz, 34(2), 227-233.

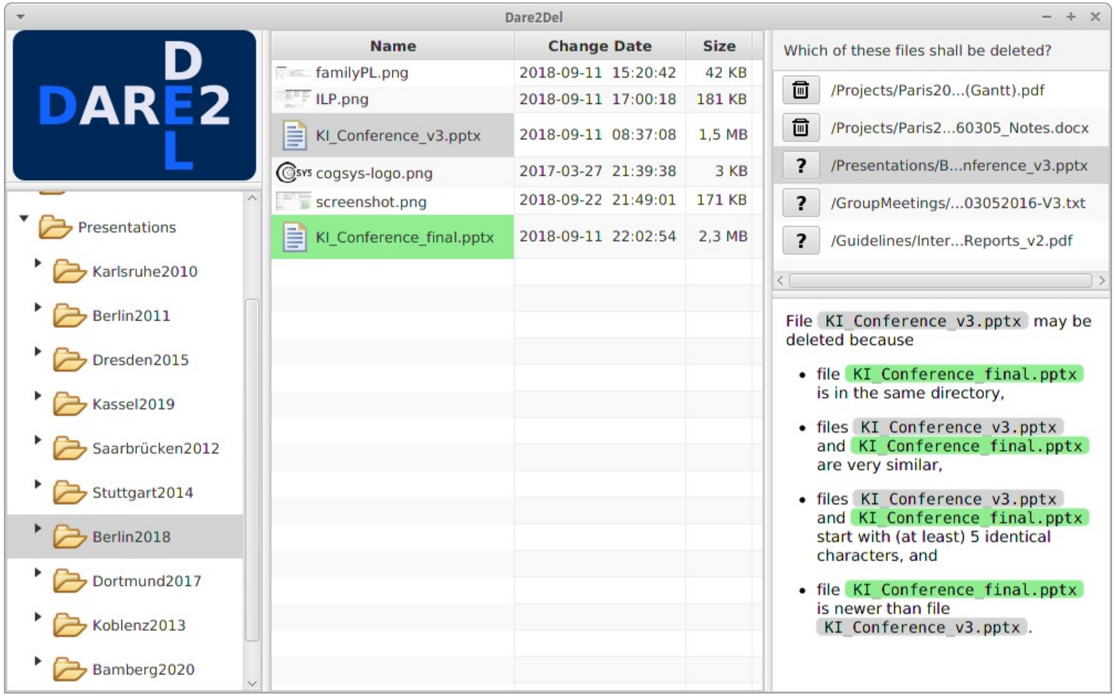
New Project BaKIM

Bambergs Baumbestand wird dokumentiert. Nun wurde die dafür angeschaffte Drohne von den Projektleitern und Andreas Starke (3. v. li.) der Öffentlichkeit vorgestellt. Foto: Stadtarchiv Bamberg, Sina Schraudner

Fränkischer Tag, 13.08.2022

- cooperative project of the City of Bamberg and the University of Bamberg
- improve the care of city trees and wooded areas
- a fixed-wing drone gathers RGB, multispectral, and thermal data, which are evaluated with different AI approaches
- support city arborists and foresters via a web application
- Human-in-the-loop ML: continuously enhances and expands the database

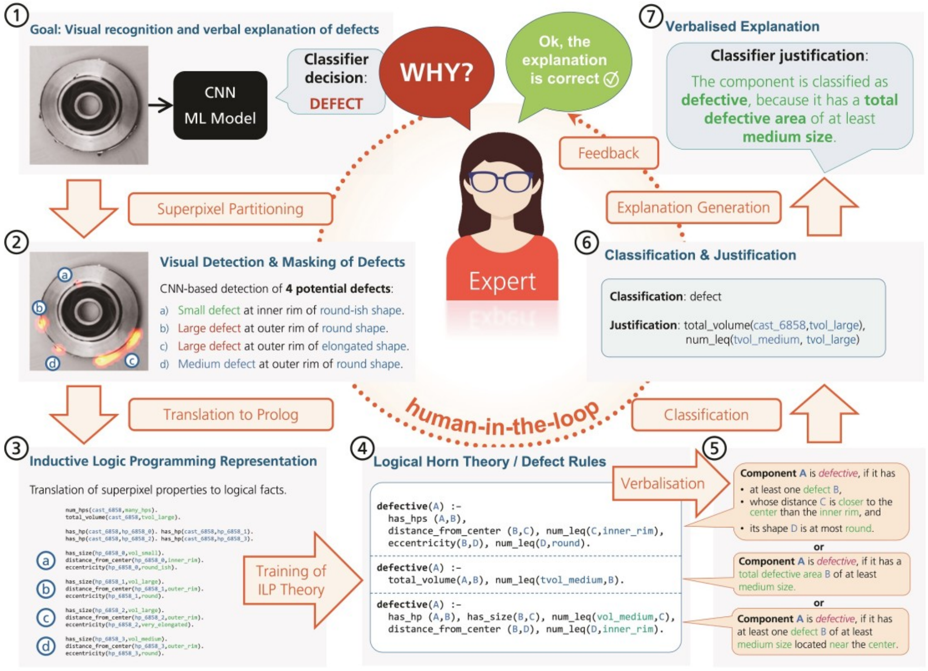
Example 2: Deleting Irrelevant Files/Data



Schmid, U. (2021). Interactive learning with mutual explanations in relational domains. In: S. Muggleton and N. Chater, Human-like Machine Intelligence,(chap.~17). 338-354, OUP.

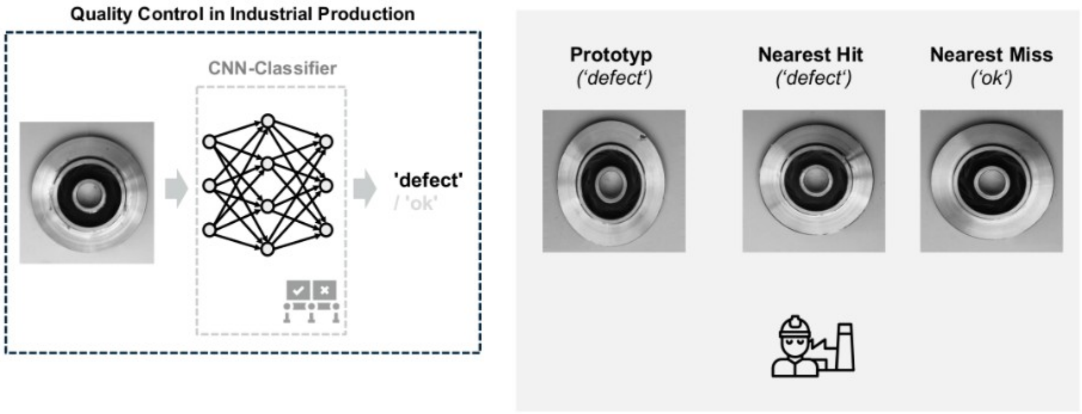
DFG Deutsche Forschungsgemeinschaft

Example 3: Human-in-the-loop ML for Quality Control



Müller, D., März, M., Scheele, S., & Schmid, U. (2022). An Interactive Explanatory AI System for Industrial Quality Control. IAIA@AAIP 2022 preprint arXiv:2203.09181.

Explaining with Near Misses



Re-implementation of Kim, Khanna, Koyejo: Examples are not Enough – Learn to Criticize! Criticism for Interpretability, NeurIPS 2016

$$\mathrm{MMD}^2(X, Y) := \frac{1}{|X|^2} \sum_{x_1, x_2 \in X} k(x_1, x_2) + \frac{1}{|Y|^2} \sum_{y_1, y_2 \in Y} k(y_1, y_2) - \frac{2}{|X| \cdot |Y|} \sum_{x \in X, y \in Y} k(x, y)$$

Maximum mean discrepancy between two distributions

(Fraunhofer IIS CAI, Herchenbach, Müller, Scheele, Schmid: Explaining Image Classifications with Near Misses, Near Hits and Prototypes – Supporting Domain Experts in Understanding Decision Boundaries, ICPRAI 2022)

Example 4: Interactive Root Cause Detection

Task:

Detect root-causes of **faulty parts** in the **production line** of electric vehicles

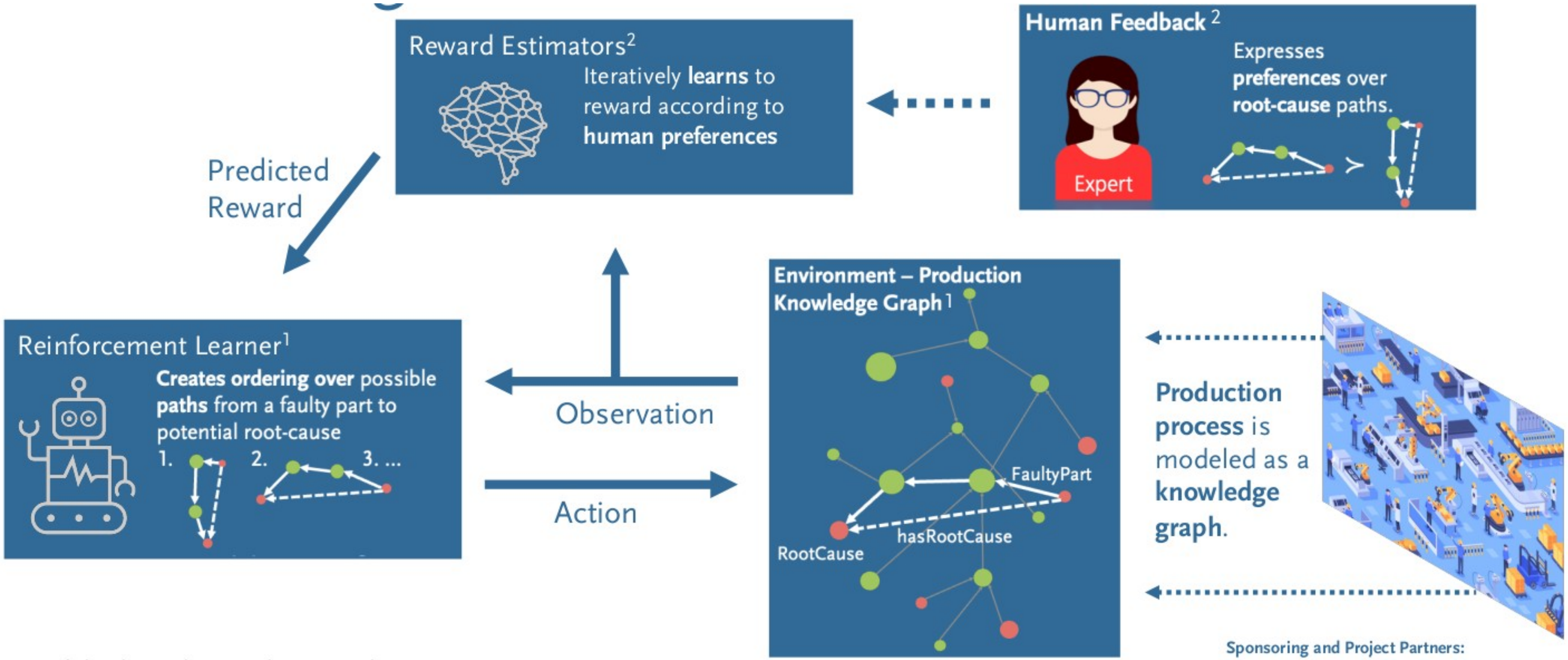
Scientific Contribution:

Combining path-based **link prediction method** for knowledge graphs and **human-in-the-loop reinforcement learning** to detect root causes

KIProQua

Bayerisches Staatsministerium für Wirtschaft, Landesentwicklung und Energie

Example 4: Interactive Root Cause Detection



1. Das, R., Dhuliawala, S., Zaheer, M., Vilnis, L., Durugkar, I., Krishnamurthy, A., Smola, A., McCallum, A. (2017). Go for a walk and arrive at the answer: Reasoning over knowledge bases with reinforcement learning, In 6th Workshop on Automated Knowledge Base Construction at NIPS 2017.

2. Christiano, P.F., Leike, J., Brown, T., Martic, M., Legg, S., Amodei, D.: (2017) Deep reinforcement learning from human preferences. vol. 30. Curran Associates, Inc.

KIProQua

Bayerisches Staatsministerium für Wirtschaft, Landesentwicklung und Energie

Example 5: AI in Education

- Predictive analytics – data-intensive, blackbox, danger of biases (behavioristic perspective on human learning)
- Intelligent Tutor Systems: explanatory, hybrid
- For specialised domains such as medicine, quality control, or farming and forest operations
  - Explanations not (only) for MLOps but also for
    - domain experts (allowed to correct models)
    - novices – teaching

Explanatory Dialogs

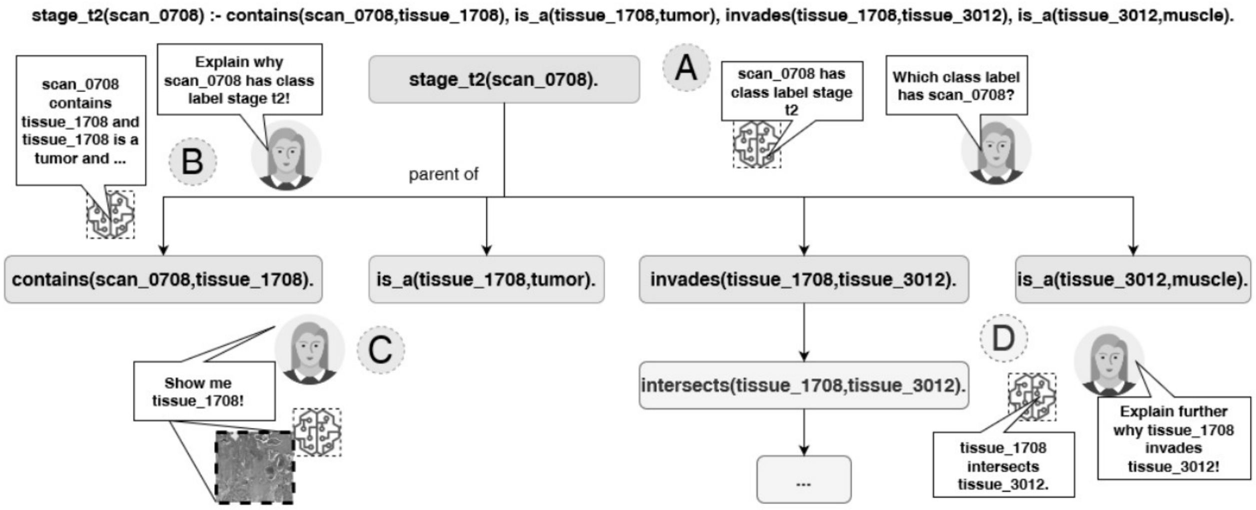


Figure 2: An explanatory tree for *stage_t2(scan_0708)*, that can be queried by the user to get a local explanation why scan_0708 is labeled as T2 (steps A and B). A dialogue is realized by further requests, either to get more visual explanations in terms of prototypes (step C) or to get more verbal explanations in a drill-down manner (step D).

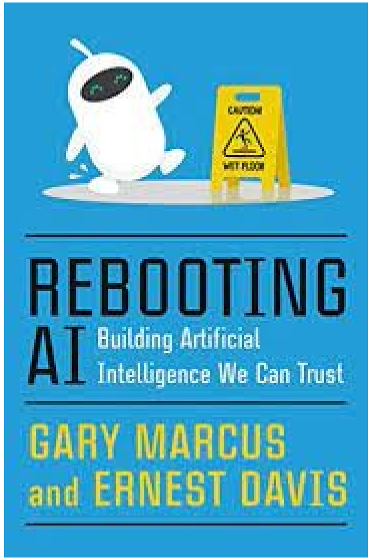
Multimodal explanations to address specific information needs

- Verbal for complex relational information
- (visual) Prototypes
- Near miss examples

Take Away

- Research on methods for Human-centred AI for complex applications
- HCAI – 3rd wave of AI: explainable and corrigible/interactive and allows to incorporate human knowledge
- Causality, of why an AI-decision has been made, paving the way towards verifiable machine learning and ethical responsible AI
- Adequate explanations and interaction interfaces require interdisciplinary cooperation

  The Holzinger Group is one of the world's leading experts on HCAI – All the best for you in Tulln!



The Holzinger Group is one of the world‘s leading experts on HCAI

All the best for you in Tulln!

From the Bamberg CogSys Group