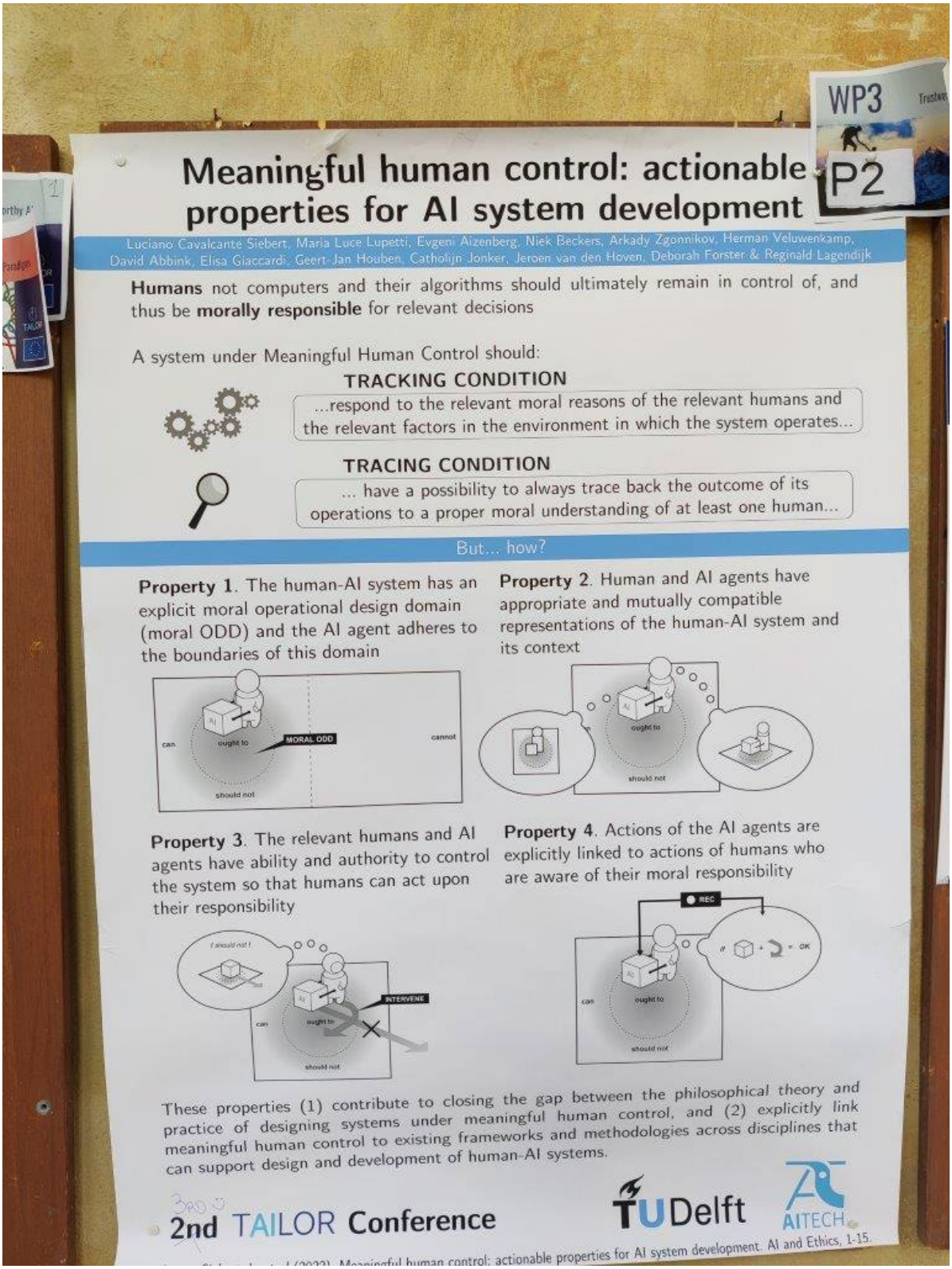
TAILOR

**Foundation of Trustworthy AI:**
Integrating Learning, Optimisation and Reasoning

# POSTER FROM 3RD TAILOR CONFERENCE

5-6 June 2023, Siena, Italy

# List of posters

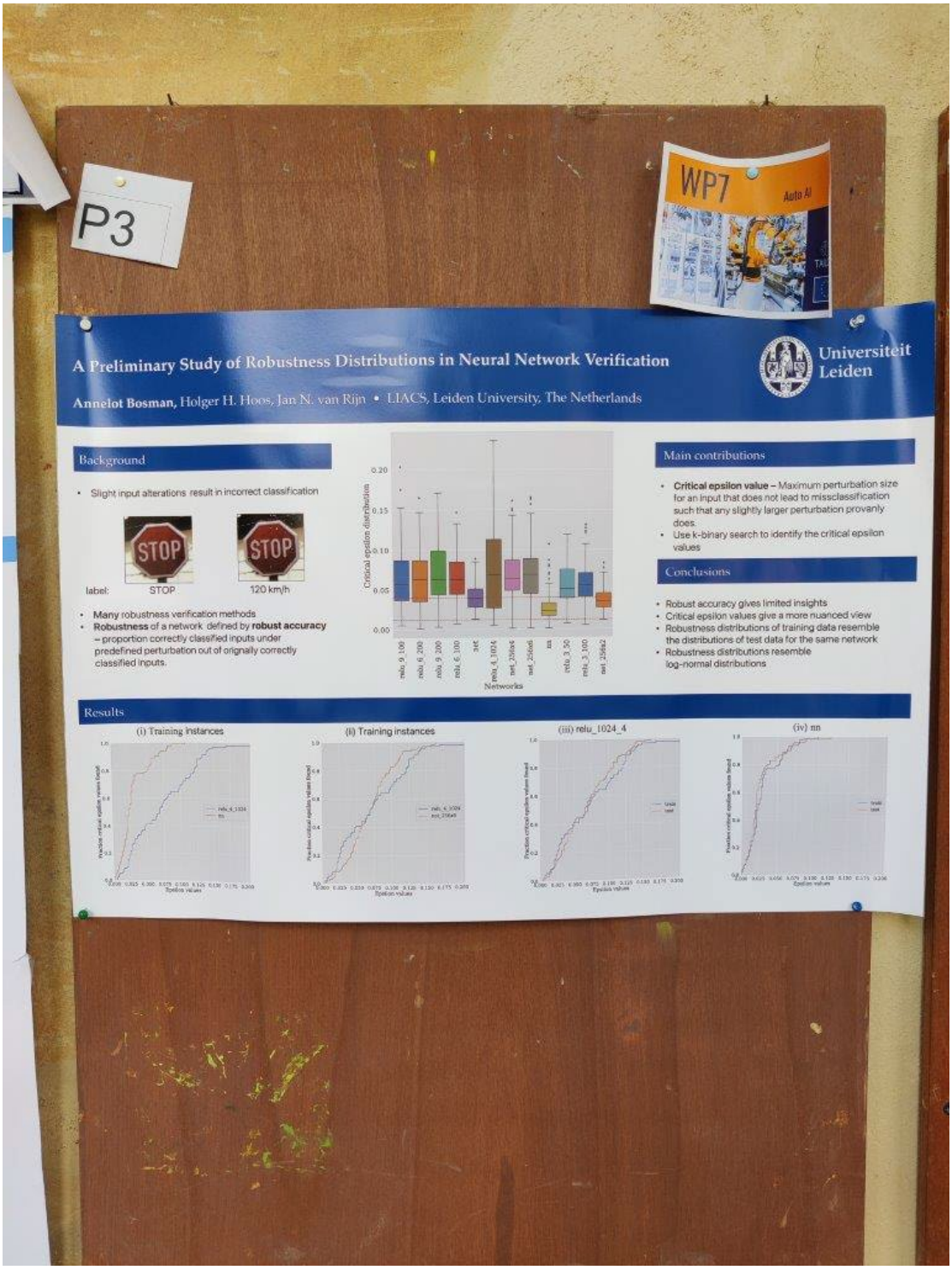

WP3
P2
Meaningful human control: actionable properties for AI system development
Luciano Cavalcante Siebert, Maria Luce Lupetti, Evgeni Aizenberg, Niek Beckers, Arkady Zgonnikov, Herman Veluwenkamp, David Abbink, Elisa Giaccardi, Geert-Jan Houben, Catholijn Jonker, Jeroen van den Hoven, Deborah Forster & Reginald Lagendijk
Humans not computers and their algorithms should ultimately remain in control of, and thus be morally responsible for relevant decisions
A system under Meaningful Human Control should:
TRACKING CONDITION
...respond to the relevant moral reasons of the relevant humans and the relevant factors in the environment in which the system operates...
TRACING CONDITION
... have a possibility to always trace back the outcome of its operations to a proper moral understanding of at least one human...
But... how?
Property 1. The human-AI system has an explicit moral operational design domain (moral ODD) and the AI agent adheres to the boundaries of this domain
can
ought to
MORAL ODD
cannot
should not
Property 2. Human and AI agents have appropriate and mutually compatible representations of the human-AI system and its context
ought to
should not
Property 3. The relevant humans and AI agents have ability and authority to control the system so that humans can act upon their responsibility
INTERVENE
can
ought to
should not
Property 4. Actions of the AI agents are explicitly linked to actions of humans who are aware of their moral responsibility
REC
OK
can
ought to
should not
These properties (1) contribute to closing the gap between the philosophical theory and practice of designing systems under meaningful human control, and (2) explicitly link meaningful human control to existing frameworks and methodologies across disciplines that can support design and development of human-AI systems.
2nd TAILOR Conference
TUDelft
AITECH
Meaningful human control: actionable properties for AI system development. AI and Ethics, 1-15.

# Meaningful human control: actionable properties for AI system development

Luciano Cavalcante Siebert, Maria Luce Lupetti, Evgeni Aizenberg, Niek Beckers, Arkady Zgonnikov, Herman Veluwenkamp, David Abbink, Elisa Giaccardi, Geert-Jan Houben, Catholijn M. Jonker, Jeroen van den Hoven, Deborah Forster & Reginald L. Lagendijk

Contact : l.cavalcantesiebert@tudelft.nl

How can humans remain in control of artificial intelligence (AI)-based systems designed to perform tasks autonomously? Such systems are increasingly ubiquitous, creating benefits - but also undesirable situations where moral responsibility for their actions cannot be properly attributed to any particular person or group. The concept of meaningful human control has been proposed to address responsibility gaps and mitigate them by establishing conditions that enable a proper attribution of responsibility for humans; however, clear requirements for researchers, designers, and engineers are yet inexistent, making the development of AI-based systems that remain under meaningful human control challenging. In this paper, we address the gap between philosophical theory and engineering practice by identifying, through an iterative process of abductive thinking, four actionable properties for AI-based systems under meaningful human control, which we discuss making use of two applications scenarios: automated vehicles and AI-based hiring. First, a system in which humans and AI algorithms interact should have an explicitly defined domain of morally loaded situations within which the system ought to operate. Second, humans and AI agents within the system should have appropriate and mutually compatible representations. Third, responsibility attributed to a human should be commensurate with that humans ability and authority to control the system. Fourth, there should be explicit links between the actions of the AI agents and actions of humans who are aware of their moral responsibility. We argue that these four properties will support practically minded professionals to take concrete steps toward designing and engineering for AI systems that facilitate meaningful human control.


P3
WP7
Auto AI
A Preliminary Study of Robustness Distributions in Neural Network Verification
Annelot Bosman, Holger H. Hoos, Jan N. van Rijn • LIACS, Leiden University, The Netherlands
Universiteit Leiden
Background
Slight input alterations result in incorrect classification
STOP
STOP
label: STOP
120 km/h
Many robustness verification methods
Robustness of a network defined by robust accuracy – proportion correctly classified inputs under predefined perturbation out of originally correctly classified inputs.
Critical epsilon distribution
Networks
Main contributions
Critical epsilon value – Maximum perturbation size for an input that does not lead to misclassification such that any slightly larger perturbation provably does.
Use k-binary search to identify the critical epsilon values
Conclusions
Robust accuracy gives limited insights
Critical epsilon values give a more nuanced view
Robustness distributions of training data resemble the distributions of test data for the same network
Robustness distributions resemble log-normal distributions
Results
(i) Training instances
(ii) Training instances
(iii) relu_1024_4
(iv) nn

# Preliminary work on robustness distributions in neural network verification

## Annelot Bosman, Holger H. Hoos, Jan N. van Rijn

Contact : a.w.bosman@liacs.leidenuniv.nl

Neural networks are vulnerable to slight alterations to correctly classified inputs, leading to incorrect predictions. To rigorously

assess the robustness of neural networks against such perturbations, exact verification techniques are employed. Robustness is generally measured in terms of adversarial accuracy, based on an upper bound on the magnitude of perturbations commonly denoted as epsilon. For each input in a given set, it is then determined whether a perturbation up to magnitude epsilon can deceive the network. In this work, we demonstrate that by refining the notion of a single bound epsilon as well as the analysis of neural network robustness, interesting insights can be gained. We introduce the concept of critical epsilon values, defined as the maximum amount of perturbation for which a given input is provably correctly classified, such that any larger perturbations can cause misclassification. To effectively estimate the critical epsilon values for each input in a given set, we utilise a variant of the binary search algorithm that we call k-binary search. We then analyse the distribution of critical epsilon values over a given set of inputs for 12 classifiers that have been used widely in the literature on neural network verification. Using a Kolmogorov-Smirnov test, we found support for the hypothesis that the critical epsilon values of all of the networks follow a log-normal distribution. We also analyse the distribution of epsilons over a set of previously unseen inputs (testing data) and the Kolmogorov-Smirnov test finds no statistically significant difference compared to seen inputs (training data). Interestingly, we find that input that is easily corrupted to deceive one network may require a considerably larger perturbation to deceive another.

# AI and Official Statistics

## Barteld Braaksma

Contact : b.braaksma@cbs.nl

The poster describes projects, approaches and considerations related to AI, from the perspective of Statistics Netherlands (CBS) and official statistics.

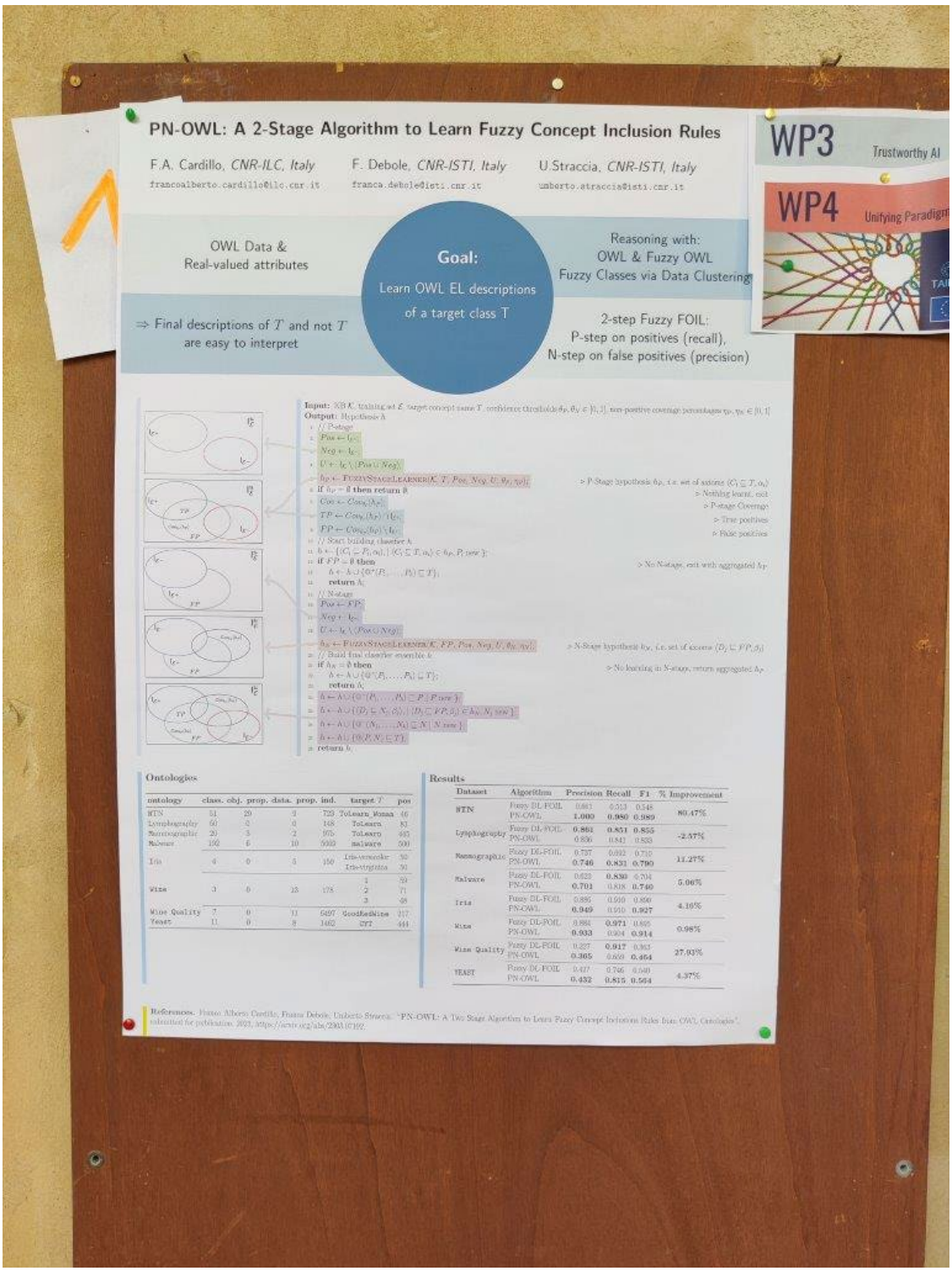

P21
WP3
Trustworthy AI
TAILOR
Physics-informed Machine Learning in Precipitation Nowcasting
3rd TAILOR Conference Poster Session
Peter Pavlík, Viera Rozinajová, Anna Bou Ezzeddine
slovak.AI
in collaboration with
KINIT
Kempelen Institute of Intelligent Technologies
TAILOR
Long-term effects of climate change will include [...] an increase in heat waves and heavy precipitation, and decreased water resources in semi-arid regions.
– United States Geological Survey
Neural Networks in Precipitation Nowcasting
Precipitation Nowcasting
Usage of Convolutional Neural Networks
Precipitation Nowcasting U-Net
Physics-Informed Machine Learning
Objectives of physics-informed machine learning
How to make a model physics-informed?
Prior knowledge, if incorporated properly, can provide rich information not existing or hard to extract in limited training data.
– When Physics Meets Machine Learning: A Survey of Physics-Informed Machine Learning
Classic Nowcasting
Motion Field-based CNN Extrapolation Nowcasting
Initial Exploratory Research
Emulate the motion field algorithm
Lukas-Kanade M.F.
U-Net emulated M.F.
Learn directly from Data
M.F. learned from data
Extrapolation

# Physics-informed Machine Learning in Precipitation Nowcasting

## Peter Pavlík

Contact : peter.pavlik@kinit.sk

Accurate forecasting of future weather is inseparable from modeling and simulation of underlying physical processes. Data-driven deep learning emulators are a promising alternative to complex numerical approaches, showing many advantages in some areas but also lacking in others. Machine learning methods incorporating existing domain knowledge about the modeled physical processes can combine the best aspects of both paradigms. The forecasts produced by machine learning models that enforce physical laws are physically consistent and increase the trust in these models. We show an example where we developed a convolutional neural network model based on existing domain knowledge, integrating the Lagrangian extrapolation scheme.

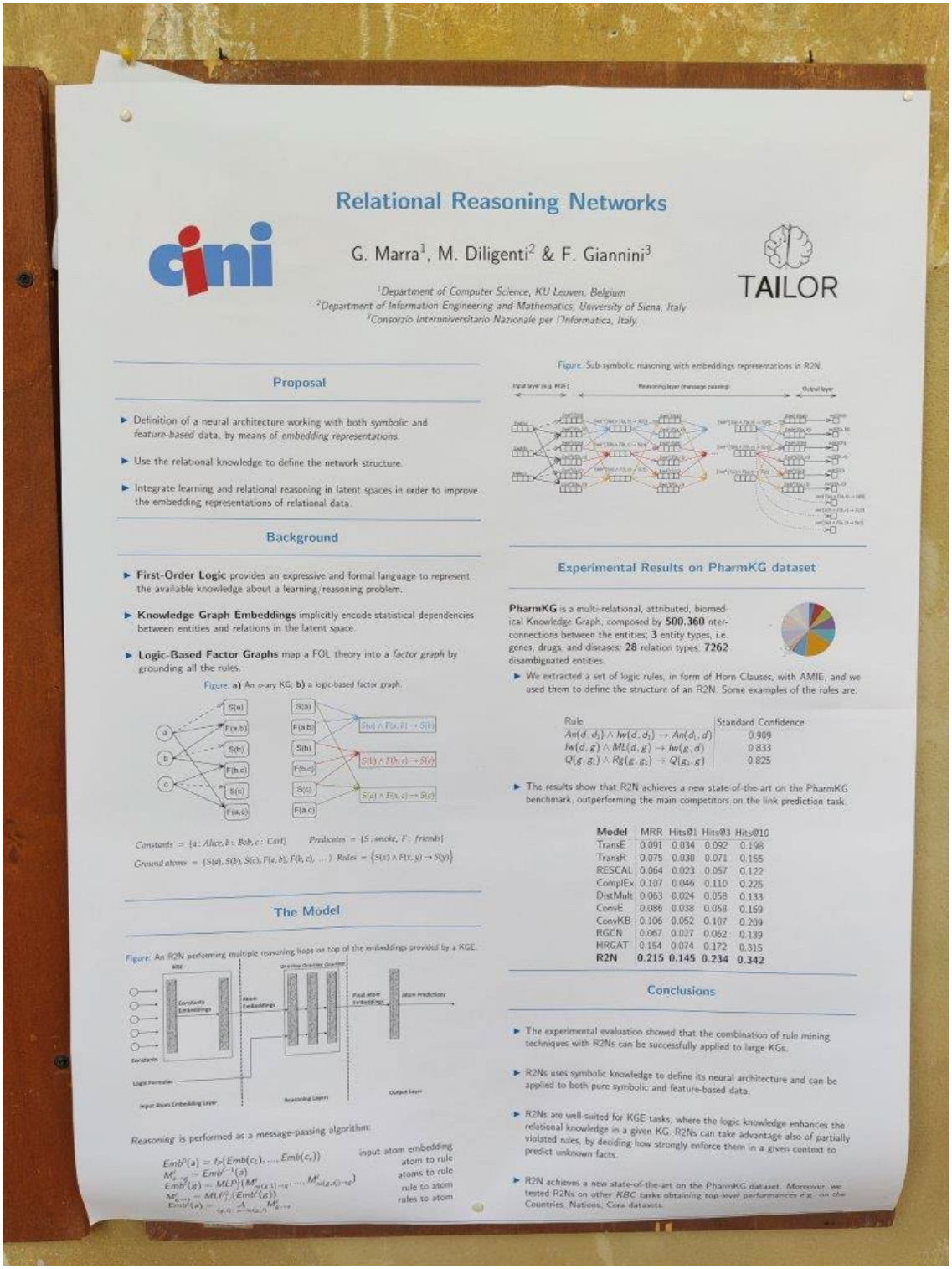

PN-OWL: A 2-Stage Algorithm to Learn Fuzzy Concept Inclusion Rules
F.A. Cardillo, CNR-ILC, Italy
francoalberto.cardillo@ilc.cnr.it
F. Debole, CNR-ISTI, Italy
franca.debole@isti.cnr.it
U.Straccia, CNR-ISTI, Italy
umberto.straccia@isti.cnr.it
WP3
Trustworthy AI
WP4
Unifying Paradigm
OWL Data &
Real-valued attributes
Goal:
Learn OWL EL descriptions
of a target class T
Reasoning with:
OWL & Fuzzy OWL
Fuzzy Classes via Data Clustering
⇒ Final descriptions of T and not T
are easy to interpret
2-step Fuzzy FOIL:
P-step on positives (recall),
N-step on false positives (precision)
Ontologies
Results
References.

# PN-OWL: A Two Stage Algorithm to Learn Fuzzy Concept Inclusions Rules from OWL Ontologies

Franco Alberto Cardillo, Franca Debole, Umberto Straccia

Contact : francoalberto.cardillo@ilc.cnr.it

In this paper, given a target class T of an OWL ontology (a W3C standard to define classes, relations among classes and class instances), we address the problem of learning graded fuzzy concept inclusion rules with the aim of describing sufficient conditions for being an individual classified as instance of the class T. To do so, we present PN-OWL that is a two-stage learning algorithm made of a P-stage and an N-stage. Roughly, in the P-stage the algorithm tries to cover as many positive examples as possible increase recall), without compromising too much precision, while in the N-stage, the algorithm tries to rule out as many false positives, covered by the P-stage, as possible. PN-OWL then aggregates the fuzzy inclusion rules learnt at the P-stage and at the N-stage by combining them via aggregation functions to allow for a final decision whether an individual is instance of T or not. We also illustrate its effectiveness by means of an extensive experimentation.



# Relational Reasoning Networks

G. Marra[1], M. Diligenti[2] & F. Giannini[3]

[1]Department of Computer Science, KU Leuven, Belgium
[2]Department of Information Engineering and Mathematics, University of Siena, Italy
[3]Consorzio Interuniversitario Nazionale per l'Informatica, Italy

## Proposal

- Definition of a neural architecture working with both *symbolic* and *feature-based* data, by means of *embedding representations*.
- Use the relational knowledge to define the network structure.
- Integrate learning and relational reasoning in latent spaces in order to improve the embedding representations of relational data.

## Background

- **First-Order Logic** provides an expressive and formal language to represent the available knowledge about a learning/reasoning problem.
- **Knowledge Graph Embeddings** implicitly encode statistical dependencies between entities and relations in the latent space.
- **Logic-Based Factor Graphs** map a FOL theory into a *factor graph* by grounding all the rules.

Figure: **a)** An n-ary KG; **b)** a logic-based factor graph.



Constants = {a : Alice, b : Bob, c : Carl}  Predicates = {S : smoke, F : friends}
Ground atoms = {S(a), S(b), S(c), F(a, b), F(b, c), ...}  Rules = $\{S(x) \wedge F(x, y) \rightarrow S(y)\}$

## The Model

Figure: An R2N performing multiple reasoning hops on top of the embeddings provided by a KGE.



Reasoning is performed as a message-passing algorithm:

$$Emb^0(a) = f_0(Emb(c_1), ..., Emb(c_n))$$ input atom embedding
$$M^t_{a\to g} = Emb^{t-1}(a)$$ atom to rule
$$Emb^t(g) = MLP^t_1(M^t_{a(g,1)\to g}, ..., M^t_{a(g,K)\to g})$$ atoms to rule
$$M^t_{g\to a} = MLP^t_2(Emb^t(g))$$ rule to atom
$$Emb^t(a) = \bigwedge_{g:a\in a(g)} M^t_{g\to a}$$ rules to atom

Figure: Sub-symbolic reasoning with embeddings representations in R2N.



## Experimental Results on PharmKG dataset

**PharmKG** is a multi-relational, attributed, biomedical Knowledge Graph, composed by **500.360** interconnections between the entities; **3** entity types, i.e. genes, drugs, and diseases; **28** relation types; **7262** disambiguated entities.



- We extracted a set of logic rules, in form of Horn Clauses, with AMIE, and we used them to define the structure of an R2N. Some examples of the rules are:

| Rule | Standard Confidence |
|---|---|
| $An(d, d_1) \wedge Iw(d, d_1) \rightarrow An(d_1, d)$ | 0.909 |
| $Iw(d, g) \wedge Mt(d, g) \rightarrow Iw(g, d)$ | 0.833 |
| $Q(g, g_1) \wedge Rg(g, g_2) \rightarrow Q(g_2, g)$ | 0.825 |

- The results show that R2N achieves a new state-of-the-art on the PharmKG benchmark, outperforming the main competitors on the link prediction task.

| Model | MRR | Hits@1 | Hits@3 | Hits@10 |
|---|---|---|---|---|
| TransE | 0.091 | 0.034 | 0.092 | 0.198 |
| TransR | 0.075 | 0.030 | 0.071 | 0.155 |
| RESCAL | 0.064 | 0.023 | 0.057 | 0.122 |
| ComplEx | 0.107 | 0.046 | 0.110 | 0.225 |
| DistMult | 0.063 | 0.024 | 0.058 | 0.133 |
| ConvE | 0.086 | 0.038 | 0.058 | 0.169 |
| ConvKB | 0.106 | 0.052 | 0.107 | 0.209 |
| RGCN | 0.067 | 0.027 | 0.062 | 0.139 |
| HRGAT | 0.154 | 0.074 | 0.172 | 0.315 |
| **R2N** | **0.215** | **0.145** | **0.234** | **0.342** |

## Conclusions

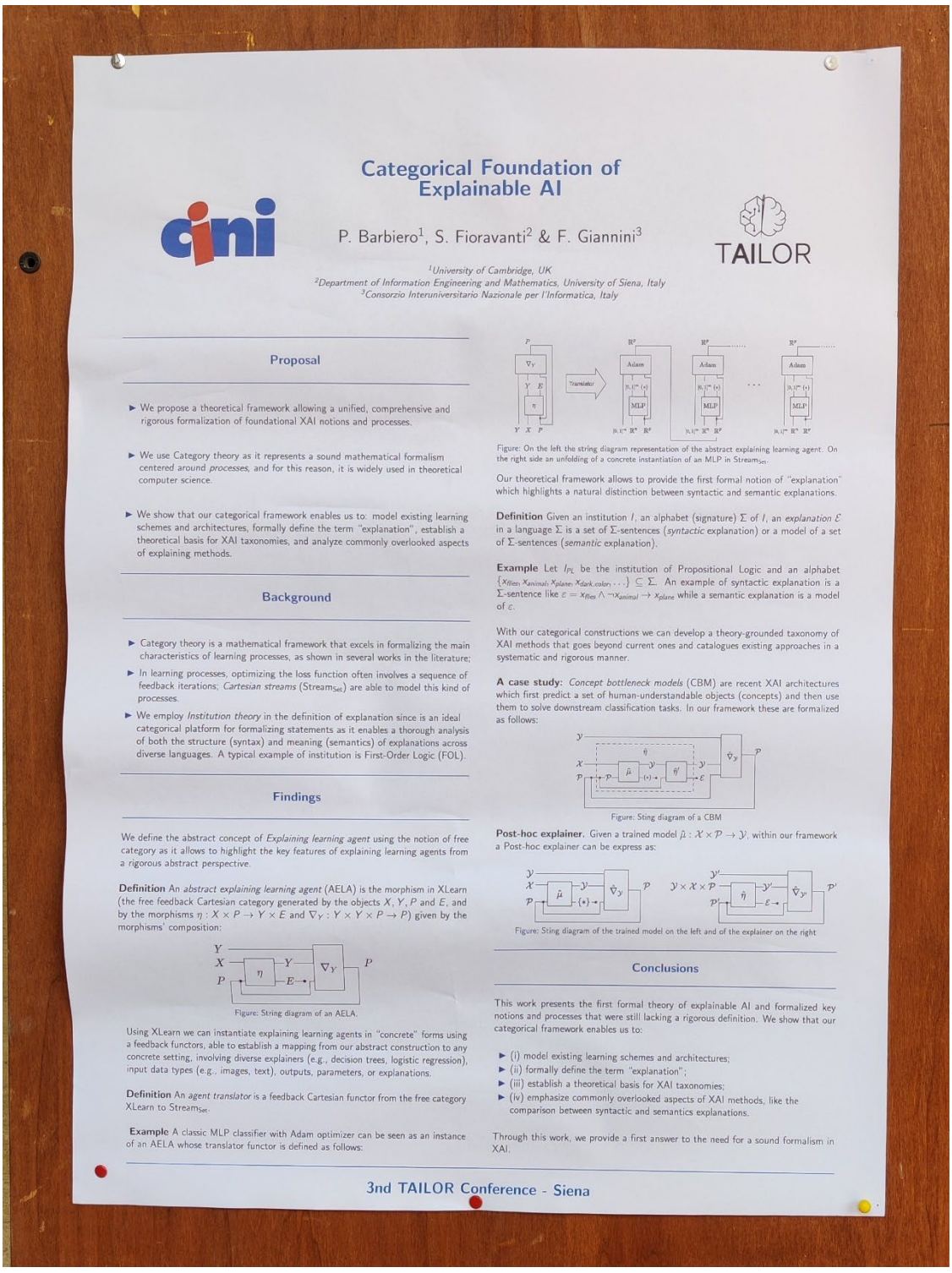
- The experimental evaluation showed that the combination of rule mining techniques with R2Ns can be successfully applied to large KGs.
- R2Ns uses symbolic knowledge to define its neural architecture and can be applied to both pure symbolic and feature-based data.
- R2Ns are well-suited for KGE tasks, where the logic knowledge enhances the relational knowledge in a given KG. R2Ns can take advantage also of partially violated rules, by deciding how strongly enforce them in a given context to predict unknown facts.
- R2N achieves a new state-of-the-art on the PharmKG dataset. Moreover, we tested R2Ns on other *KBC* tasks obtaining top-level performances e.g. on the Countries, Nations, Cora datasets.

# Relational Reasoning Networks

## Giuseppe Marra, Michelangelo Diligenti, Francesco Giannini

Contact : francesco.giannini@unisi.it

This poster introduces Relational Reasoning Network (R2N) a neural architecture that combines symbolic and feature-based data through embedded representations. The architecture utilizes relational knowledge to define the structure of the network. R2N further incorporates learning and relational reasoning in latent spaces to enhance the embedding representations of relational data.



# Categorical Foundation of Explainable AI

P. Barbiero[1], S. Fioravanti[2] & F. Giannini[3]

[1]University of Cambridge, UK
[2]Department of Information Engineering and Mathematics, University of Siena, Italy
[3]Consorzio Interuniversitario Nazionale per l'Informatica, Italy

## Proposal

- We propose a theoretical framework allowing a unified, comprehensive and rigorous formalization of foundational XAI notions and processes.
- We use Category theory as it represents a sound mathematical formalism centered around *processes*, and for this reason, it is widely used in theoretical computer science.
- We show that our categorical framework enables us to: model existing learning schemes and architectures, formally define the term "explanation", establish a theoretical basis for XAI taxonomies, and analyze commonly overlooked aspects of explaining methods.

## Background

- Category theory is a mathematical framework that excels in formalizing the main characteristics of learning processes, as shown in several works in the literature;
- In learning processes, optimizing the loss function often involves a sequence of feedback iterations; *Cartesian streams* ($\mathrm{Stream}_{\mathrm{Set}}$) are able to model this kind of processes.
- We employ *Institution theory* in the definition of explanation since is an ideal categorical platform for formalizing statements as it enables a thorough analysis of both the structure (syntax) and meaning (semantics) of explanations across diverse languages. A typical example of institution is First-Order Logic (FOL).

## Findings

We define the abstract concept of *Explaining learning agent* using the notion of free category as it allows to highlight the key features of explaining learning agents from a rigorous abstract perspective.

**Definition** An *abstract explaining learning agent* (AELA) is the morphism in XLearn (the free feedback Cartesian category generated by the objects $X, Y, P$ and $E$, and by the morphisms $\eta : X \times P \to Y \times E$ and $\nabla_Y : Y \times Y \times P \to P$) given by the morphisms' composition:



Figure: String diagram of an AELA.

Using XLearn we can instantiate explaining learning agents in "concrete" forms using a feedback functors, able to establish a mapping from our abstract construction to any concrete setting, involving diverse explainers (e.g., decision trees, logistic regression), input data types (e.g., images, text), outputs, parameters, or explanations.

**Definition** An *agent translator* is a feedback Cartesian functor from the free category XLearn to $\mathrm{Stream}_{\mathrm{Set}}$.

**Example** A classic MLP classifier with Adam optimizer can be seen as an instance of an AELA whose translator functor is defined as follows:



Figure: On the left the string diagram representation of the abstract explaining learning agent. On the right side an unfolding of a concrete instantiation of an MLP in $\mathrm{Stream}_{\mathrm{Set}}$.

Our theoretical framework allows to provide the first formal notion of "explanation" which highlights a natural distinction between syntactic and semantic explanations.

**Definition** Given an institution $I$, an alphabet (signature) $\Sigma$ of $I$, an *explanation* $\mathcal{E}$ in a language $\Sigma$ is a set of $\Sigma$-sentences (*syntactic* explanation) or a model of a set of $\Sigma$-sentences (*semantic* explanation).

**Example** Let $I_{PL}$ be the institution of Propositional Logic and an alphabet $\{x_{flies}, x_{animal}, x_{plane}, x_{dark.color}, \dots\} \subseteq \Sigma$. An example of syntactic explanation is a $\Sigma$-sentence like $\varepsilon = x_{flies} \wedge \neg x_{animal} \to x_{plane}$ while a semantic explanation is a model of $\varepsilon$.

With our categorical constructions we can develop a theory-grounded taxonomy of XAI methods that goes beyond current ones and catalogues existing approaches in a systematic and rigorous manner.

**A case study:** *Concept bottleneck models* (CBM) are recent XAI architectures which first predict a set of human-understandable objects (concepts) and then use them to solve downstream classification tasks. In our framework these are formalized as follows:



Figure: Sting diagram of a CBM

**Post-hoc explainer.** Given a trained model $\hat{\mu} : \mathcal{X} \times \mathcal{P} \to \mathcal{Y}$, within our framework a Post-hoc explainer can be express as:



Figure: Sting diagram of the trained model on the left and of the explainer on the right

## Conclusions

This work presents the first formal theory of explainable AI and formalized key notions and processes that were still lacking a rigorous definition. We show that our categorical framework enables us to:

- (i) model existing learning schemes and architectures;
- (ii) formally define the term "explanation";
- (iii) establish a theoretical basis for XAI taxonomies;
- (iv) emphasize commonly overlooked aspects of XAI methods, like the comparison between syntactic and semantics explanations.

Through this work, we provide a first answer to the need for a sound formalism in XAI.

3nd TAILOR Conference - Siena

# Categorical Foundation of Explainable A

## Pietro Barbiero, Stefano Fioravanti, Francesco Giannini, Alberto Tonda, Pietro Li\'o, Elena Di Lavore

Contact : stefano.fioravanti66@gmail.com

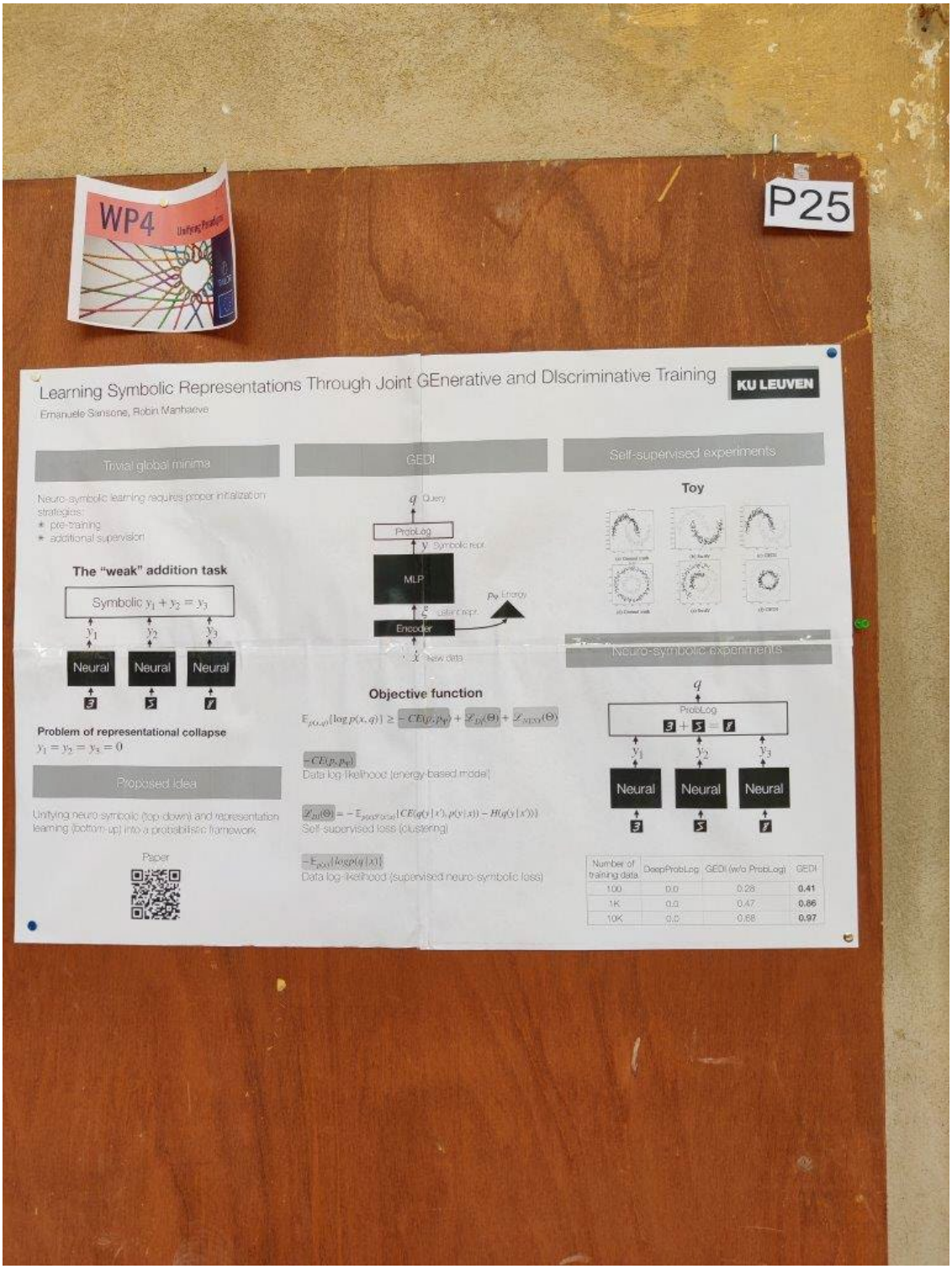
Explainable AI (XAI) aims to address the human need for safe and reliable AI systems. However, numerous surveys emphasize the absence of a sound mathematical formalization of key XAI notions. To bridge this gap, this paper presents the first mathematically rigorous definitions of key XAI notions and processes, using the well-funded formalism of Category theory. We show that our categorical framework allows to: (i) model existing learning schemes and architectures, (ii) formally define the term ``explanation'', (iii) establish a theoretical basis for XAI taxonomies, and (iv) analyze commonly overlooked aspects of explaining methods. As a consequence, our categorical framework promotes the ethical and secure deployment of AI technologies as it represents a significant step towards a sound theoretical foundation of explainable AI.

# Graph Neural Networks for ECG Classification]{Graph Neural Networks for Topological Feature Extraction in ECG Classification

Kamyar Zeinalipour, Marco Gori

Contact : Kamyar.zeinalipur2@unisi.it

The electrocardiogram (ECG) is a dependable instrument for assessing the function of the cardiovascular system. There has recently been much emphasis on precisely classifying ECGs. While ECG situations have numerous similarities, little attention has been paid to categorizing ECGs using graph neural networks. In this study, we offer three distinct techniques for classifying heartbeats using deep graph neural networks to classify the ECG signals accurately. We suggest using different methods to extract topological features from the ECG signal and then using a branch of the graph neural network named graph isomorphism network for classifying the ECGs. On the PTB Diagnostics data set, we tested the three proposed techniques. According to the findings, the three proposed techniques are capable of making arrhythmia classification predictions with the accuracy of 99.38, 98.76, and 91.93 percent, respectively.

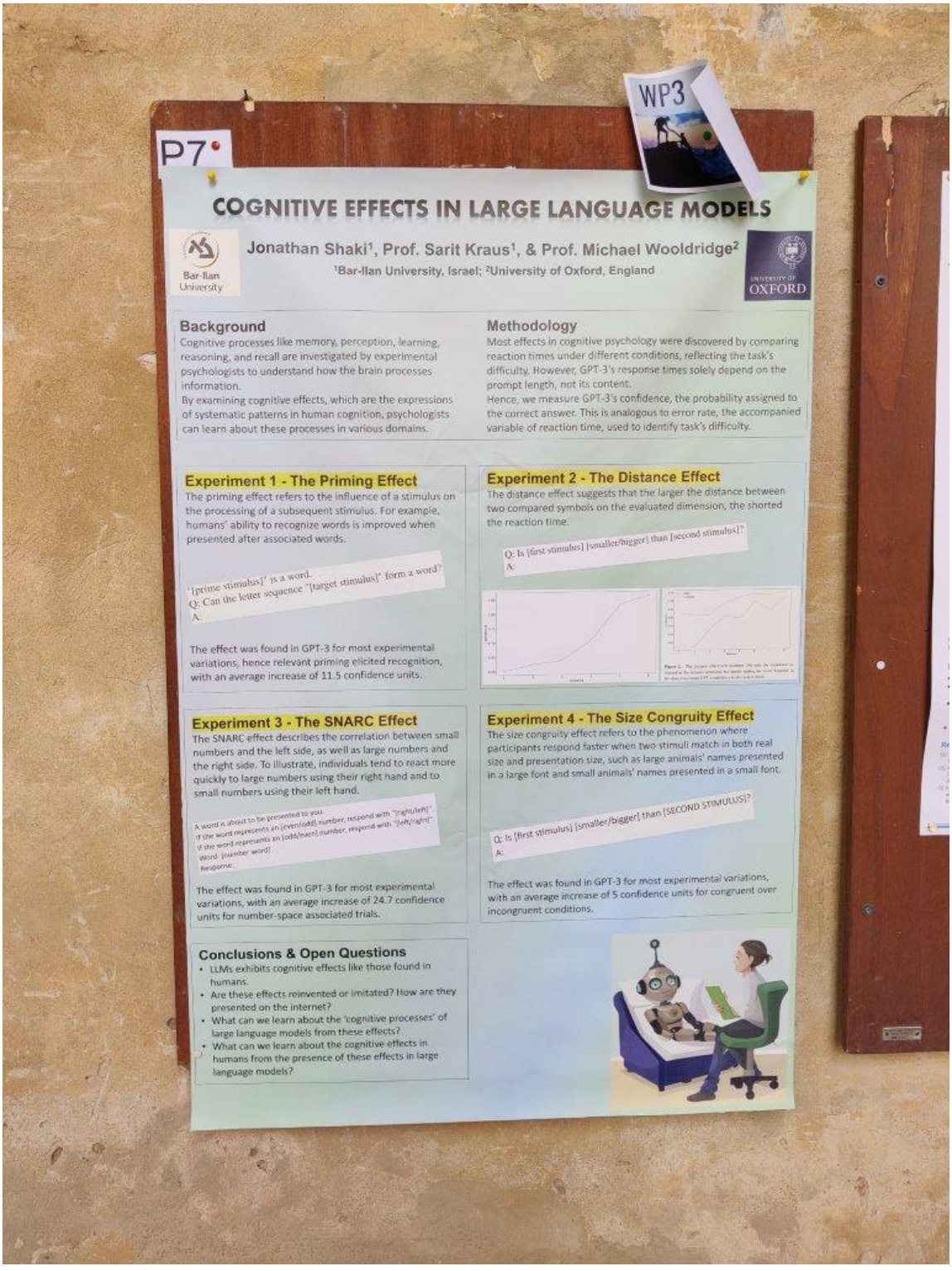

WP4
P25
Learning Symbolic Representations Through Joint GEnerative and DIscriminative Training
KU LEUVEN
Emanuele Sansone, Robin Manhaeve
Trivial global minima
Neuro-symbolic learning requires proper initialization strategies:
* pre-training
* additional supervision
The "weak" addition task
Symbolic $y_1 + y_2 = y_3$
Neural
Problem of representational collapse
$y_1 = y_2 = y_3 = 0$
Proposed Idea
Unifying neuro-symbolic (top-down) and representation learning (bottom-up) into a probabilistic framework
Paper
GEDI
ProbLog
MLP
Encoder
Objective function
Data log-likelihood (energy-based model)
Self-supervised loss (clustering)
Data log-likelihood (supervised neuro-symbolic loss)
Self-supervised experiments
Toy
Neuro-symbolic experiments
ProbLog
Neural
Number of training data
DeepProbLog
GEDI (w/o ProbLog)
GEDI
100
0.0
0.28
0.41
1K
0.0
0.47
0.86
10K
0.0
0.68
0.97

# Learning Symbolic Representations Through Joint GEnerative and DIscriminative Training

## Emanuele Sansone, Robin Manhaeve

Contact : e.sansone@hotmail.it

We introduce GEDI, a Bayesian framework that combines existing self-supervised learning objectives with likelihood-based generative models. This framework leverages the benefits of both GEnerative and DIscriminative approaches, resulting in improved symbolic representations over standalone solutions. Additionally, GEDI can be easily integrated and trained jointly with existing neuro-symbolic frameworks without the need for additional supervision or costly pre-training steps. We demonstrate through experiments on real-world data, including SVHN, CIFAR10, and CIFAR100, that GEDI outperforms existing self-supervised learning strategies in terms of clustering performance by a significant margin. The symbolic component further allows it to leverage knowledge in the form of logical constraints to improve performance in the small data regime.

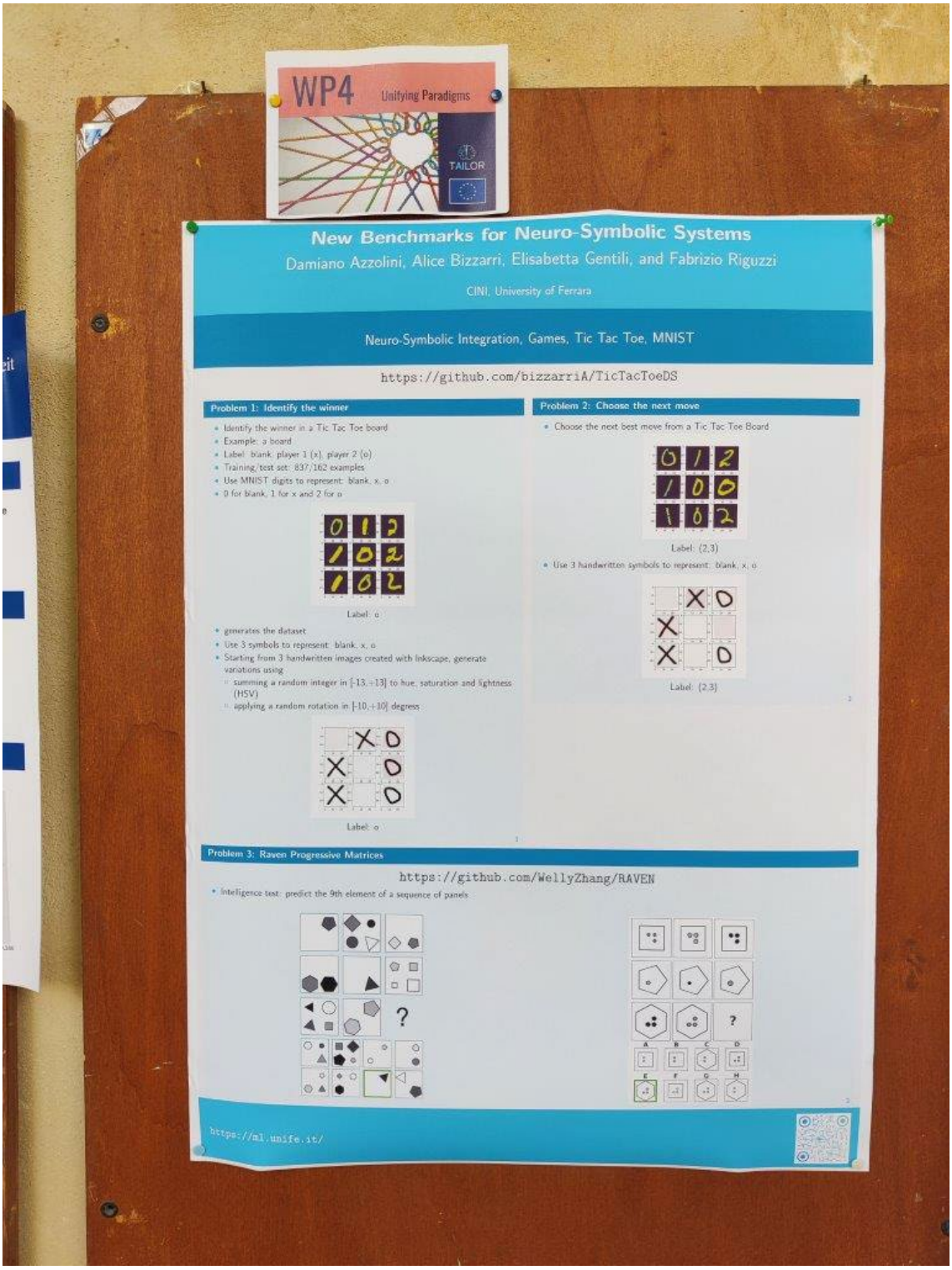

P7
WP3
COGNITIVE EFFECTS IN LARGE LANGUAGE MODELS
Jonathan Shaki[1], Prof. Sarit Kraus[1], & Prof. Michael Wooldridge[2]
[1]Bar-Ilan University, Israel; [2]University of Oxford, England
Bar-Ilan University
OXFORD
Background
Cognitive processes like memory, perception, learning, reasoning, and recall are investigated by experimental psychologists to understand how the brain processes information.
By examining cognitive effects, which are the expressions of systematic patterns in human cognition, psychologists can learn about these processes in various domains.
Methodology
Most effects in cognitive psychology were discovered by comparing reaction times under different conditions, reflecting the task's difficulty. However, GPT-3's response times solely depend on the prompt length, not its content.
Hence, we measure GPT-3's confidence, the probability assigned to the correct answer. This is analogous to error rate, the accompanied variable of reaction time, used to identify task's difficulty.
Experiment 1 - The Priming Effect
The priming effect refers to the influence of a stimulus on the processing of a subsequent stimulus. For example, humans' ability to recognize words is improved when presented after associated words.
'[prime stimulus]' is a word.
Q: Can the letter sequence "[target stimulus]" form a word?
A:
The effect was found in GPT-3 for most experimental variations, hence relevant priming elicited recognition, with an average increase of 11.5 confidence units.
Experiment 2 - The Distance Effect
The distance effect suggests that the larger the distance between two compared symbols on the evaluated dimension, the shorted the reaction time.
Q: Is [first stimulus] [smaller/bigger] than [second stimulus]?
A:
Experiment 3 - The SNARC Effect
The SNARC effect describes the correlation between small numbers and the left side, as well as large numbers and the right side. To illustrate, individuals tend to react more quickly to large numbers using their right hand and to small numbers using their left hand.
A word is about to be presented to you.
If the word represents an [even/odd] number, respond with "[right/left]".
If the word represents an [odd/even] number, respond with "[left/right]".
Word: [number word]
Response:
The effect was found in GPT-3 for most experimental variations, with an average increase of 24.7 confidence units for number-space associated trials.
Experiment 4 - The Size Congruity Effect
The size congruity effect refers to the phenomenon where participants respond faster when two stimuli match in both real size and presentation size, such as large animals' names presented in a large font and small animals' names presented in a small font.
Q: Is [first stimulus] [smaller/bigger] than [SECOND STIMULUS]?
A:
The effect was found in GPT-3 for most experimental variations, with an average increase of 5 confidence units for congruent over incongruent conditions.
Conclusions & Open Questions
• LLMs exhibits cognitive effects like those found in humans.
• Are these effects reinvented or imitated? How are they presented on the internet?
• What can we learn about the 'cognitive processes' of large language models from these effects?
• What can we learn about the cognitive effects in humans from the presence of these effects in large language models?

# Cognitive Effects in Large Language Models

## Jonathan Shaki, Prof. Sarit Kraus, Prof. Michael Wooldridge

Contact : jonathan.shaki@hotmail.com

Large Language Models (LLMs) such as ChatGPT have received enormous attention over the past year, and are now used by hundreds of millions of people every day. The rapid adoption of this technology naturally raises questions about the possible biases that such models might exhibit. In this work, we tested one of these models (GPT-3) on a range of cognitive effects, which are systematic patterns that are usually found in human cognitive tasks. We found that LLMs are indeed prone to a number of human cognitive effects. Specifically, we show that the priming, distance, SNARC, and size congruity effects were presented with GPT-3, while the anchoring effect is absent. We describe our methodology, and specifically the way we converted real-world experiments to text-based experiments. Finally, we speculate on the possible reasons why GPT-3 exhibits these effects, and discuss the question of whether they are imitated or reinvented.


WP4
Unifying Paradigms
TAILOR
New Benchmarks for Neuro-Symbolic Systems
Damiano Azzolini, Alice Bizzarri, Elisabetta Gentili, and Fabrizio Riguzzi
CINI, University of Ferrara
Neuro-Symbolic Integration, Games, Tic Tac Toe, MNIST
https://github.com/bizzarriA/TicTacToeDS
Problem 1: Identify the winner
Identify the winner in a Tic Tac Toe board
Example: a board
Label: blank: player 1 (x), player 2 (o)
Training/test set: 837/162 examples
Use MNIST digits to represent: blank, x, o
0 for blank, 1 for x and 2 for o
Label: o
generates the dataset
Use 3 symbols to represent: blank, x, o
Starting from 3 handwritten images created with Inkscape, generate variations using
summing a random integer in [-13, +13] to hue, saturation and lightness (HSV)
applying a random rotation in [-10, +10] degrees
Label: o
Problem 2: Choose the next move
Choose the next best move from a Tic Tac Toe Board
Label: (2,3)
Use 3 handwritten symbols to represent: blank, x, o
Label: (2,3)
Problem 3: Raven Progressive Matrices
https://github.com/WellyZhang/RAVEN
Intelligence test: predict the 9th element of a sequence of panels
https://ml.unife.it/

# New Benchmarks for Neuro-Symbolic Systems

## Damiano Azzolini, Alice Bizzarri, Elisabetta Gentili, Fabrizio Riguzzi

Contact : fabrizio.riguzzi@unife.it

We present three new benchmarks for neuro-symbolic systems: identifying the winner in a Tic Tac Toe board, predicting the next move in a Tic Tac Toe game and solving Raven Progressive Matrices.

In Tic Tac Toe, to represent the three symbols blank, x and o, we use either MNIST images or synthetic images.

The datasets for winner identification and next move prediction can be generated with the code at https://github.com/bizzarriA/TicTacToeDS

Raven Progressive Matrices are IQ tests where the palyer is presented with a 3x3 array of panels with geometric figures that are arranged according to unkwown rules. Of the 9 panels, the last one is masked and is to be selected among 8 possible answer panels.

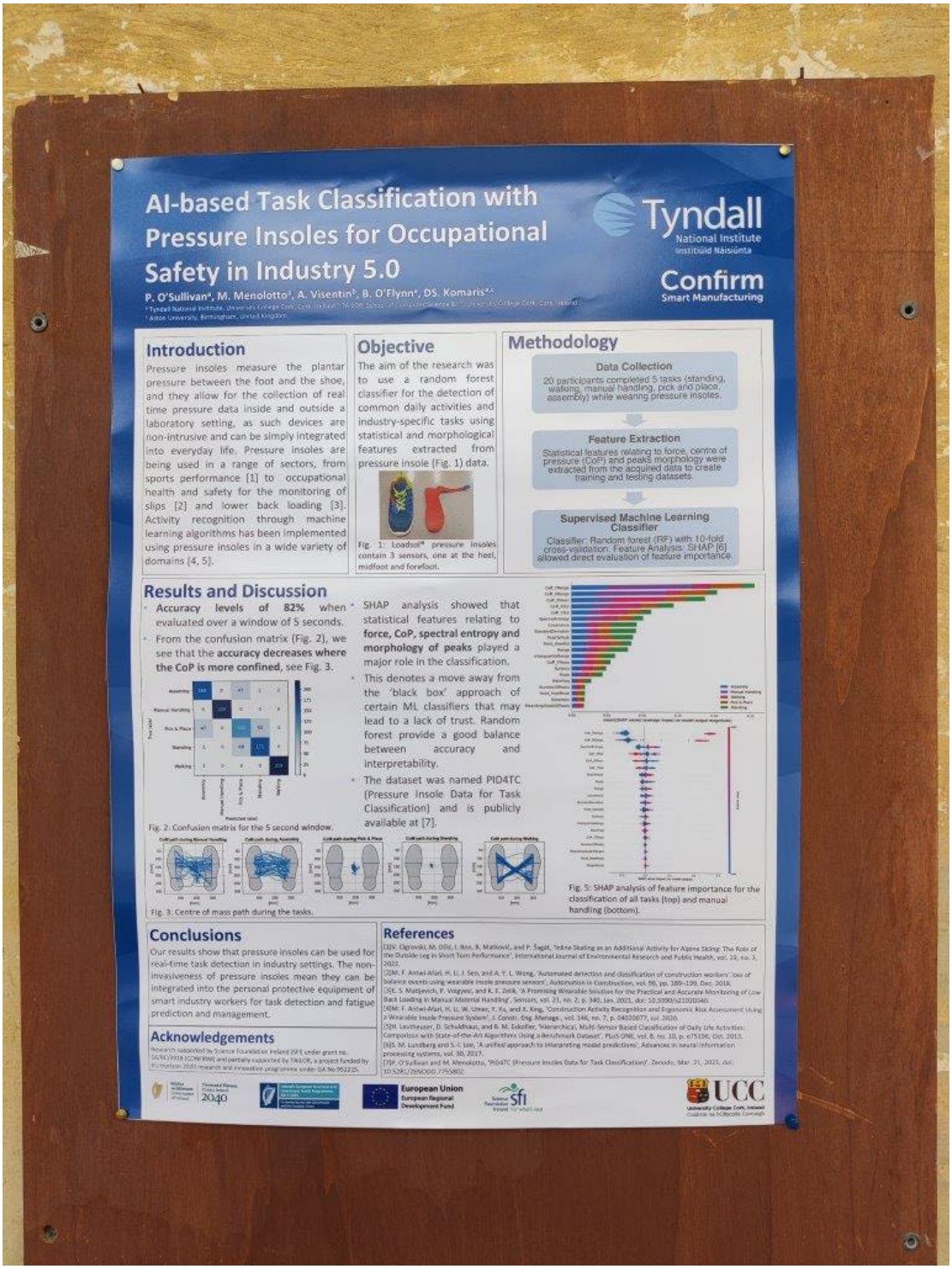

P6
WP3
Trustworthy AI
TAILOR
3rd TAILOR Conference - Siena, June 2023
TAILOR
Multi-criteria approach for selecting an explanation from the set of counterfactuals
Ignacy Stępka, Mateusz Lango, Jerzy Stefanowski
Poznan University of Technology
{ignacy.stepka, mateusz.lango, jerzy.stefanowski}@cs.put.poznan.pl
Counterfactual explanations
They are used to explain predictions of black box ML models. They provide directions on the smallest possible changes to the attributes of a given instance that lead to the desired prediction.
Consider a loan case, where your application was rejected by the software. You may wonder what changes you can make to improve your chances of getting the loan.
Our Contribution
The counterfactuals can be generated by a variety of methods that produce quite different alternatives - not easy for the final analysis. The aims of our work:
We propose the use of an ensemble of explainers to generate a large and diverse set of counterfactuals, which can be analyzed with different quality criteria
We address the need to impose constraints on immutable attributes to render the explanations actionable
We investigate the usefulness of constraints and the dominance relation in multi-criteria analysis to select a concise set of explanations
We generalize the multi-criteria method (the reference point) to support user in selecting the final explanation
Below QRcode to the early version of the paper
Method
The analysis of generated counterfactual explanations consists of the three following steps
Generating counterfactuals
Ensemble of Explainers
Validity Checking
Check if counterfactual actually changes the outcome (i.e., has different class than original example)
Actionability enforcement using constraints
Remove counterfactuals that violate the constraints (i.e., propose a change on immutable attribute)
Filtering dominated solutions
Filter explanations using dominance relation to obtain concise pareto front of non-dominated explanations
Multi-criteria
In multi-criteria analysis, three following quality measures for counterfactuals are considered:
Proximity - distance to the original instance x
Feasibility - distance to the k-neighbours of x
Discriminative Power - distance to k-neighbours of x having the same class as counterfactual
Other criteria skipped, partly because some of them are considerably correlated with the ones we take into account.
Results
For the explainers in the ensemble, we chose a set of 9 methodologically different methods
Explainers were set to generate up to 20 explanations, which were all later combined together to form a large set of alternative counterfactual explanations
To investigate the Actionability, we imposed a set of constraints for certain immutable attributes (e.g., ethnicity, age)
To simulate the black box classifier, we trained a shallow neural network
Dataset All Valid Actionable P-Front
Adult 85.4 ± 2.8 67.1 ± 9.2 60.2 ± 8 7.6 ± 3.7
German 45 ± 3.8 40.3 ± 7 45.4 ± 7.3 11.9 ± 3.9
Our approach produces, on average, over 80 counterfactual explanations per instance. A relatively large number of these explanations fail to meet the validity and actionability requirements, and are subsequently removed from further consideration.
Adult German
Any Val. Act. Front Any Val. Act. Front
Dice 20 20 20 11.7 20 20 20 2.22
Cadex 6.53 5.38 5.38 0.98 11.9 10.67 10.67 1.56
Fimap 6 4.87 4.87 0.37 6 3.38 3.38 0.86
Wach. 9.96 3.40 3.40 1.03 10 2.33 2.33 2.71
CEM 1 0.5 0.5 0.4 1 0.21 0.21 0.41
CFPy 7.06 7.61 2.26 0.14 5.76 5.71 0.81 0.3
GS 20 14.34 14.34 0.18 20 13.55 13.55 0.24
AR 0.29 0.15 0.15 0.1 0.54 0.17 0.17 0.15
FACE 10 9.78 8.28 3.27 10 9.06 0.68 3.74
The multi-criteria analysis:
With the dominance relation, we identify a concise Pareto front — (8 – 12) non-dominated alternatives
The ideal point method allows us to select the best compromise explanation
Table: The results obtained for German dataset. The best results are bolded, the second best are underlined and the third best are in italics.
Acknowledgment
The research has been partially supported by TAILOR, a project funded by EU Horizon 2020 research and innovation programme under GA no. 952215.

# Multi-criteria approach for selecting an explanation from the set of counterfactuals

Ignacy Stępka, Jerzy Stefanowski, Mateusz Lango

Contact : jerzy.stefanowski@cs.put.poznan.pl

Counterfactual explanations are widely used to explain predictions of black box ML models. They can be generated by a variety of methods that optimize different, sometimes conflicting, quality measures and produce quite different solutions. However, choosing the most appropriate explanation method and one of generated counterfactuals is not an easy task. Instead of forcing the user to test many different explanation methods and analysing conflicting solutions, we propose to use an approach that integrates several popular methods of generating counterfactuals to obtain a diverse set of explanations. Then, we introduce an approach that filters out a relatively large set of counterfactuals generated by this set of diverse algorithms through a multi-criteria subset selection problem solved using the dominance relation. Experiments show that using multiple counterfactual generation methods and selecting the subset of alternatives using the dominance relation results in a well-performing and concise set of counterfactual explanations.

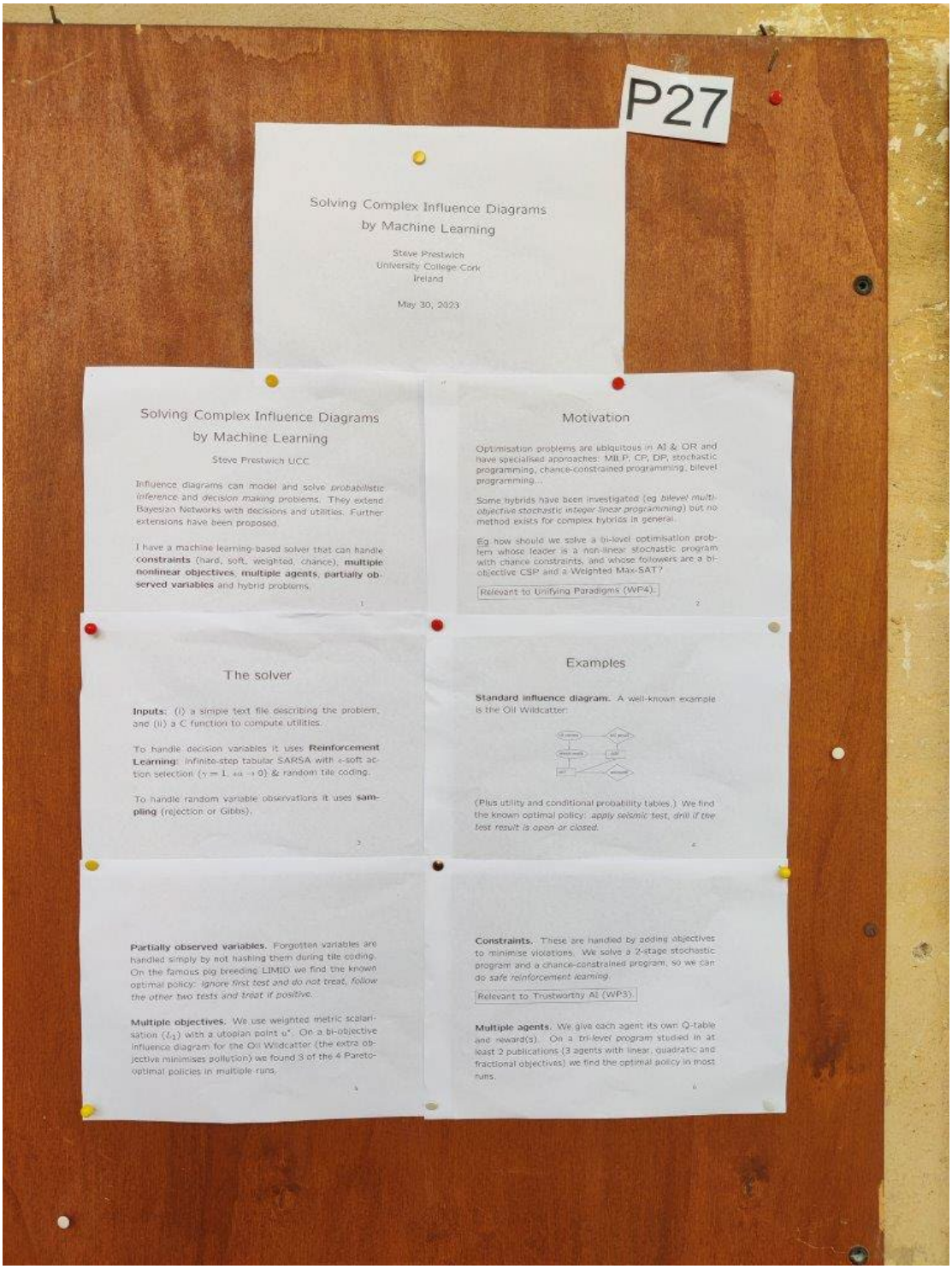

AI-based Task Classification with Pressure Insoles for Occupational Safety in Industry 5.0
Tyndall National Institute
Confirm Smart Manufacturing
P. O'Sullivan, M. Menolotto, A. Visentin, B. O'Flynn, DS. Komaris
Introduction
Pressure insoles measure the plantar pressure between the foot and the shoe, and they allow for the collection of real time pressure data inside and outside a laboratory setting, as such devices are non-intrusive and can be simply integrated into everyday life. Pressure insoles are being used in a range of sectors, from sports performance [1] to occupational health and safety for the monitoring of slips [2] and lower back loading [3]. Activity recognition through machine learning algorithms has been implemented using pressure insoles in a wide variety of domains [4, 5].
Objective
The aim of the research was to use a random forest classifier for the detection of common daily activities and industry-specific tasks using statistical and morphological features extracted from pressure insole (Fig. 1) data.
Fig. 1: Loadsol® pressure insoles contain 3 sensors, one at the heel, midfoot and forefoot.
Methodology
Data Collection
20 participants completed 5 tasks (standing, walking, manual handling, pick and place, assembly) while wearing pressure insoles.
Feature Extraction
Statistical features relating to force, centre of pressure (CoP) and peaks morphology were extracted from the acquired data to create training and testing datasets.
Supervised Machine Learning Classifier
Classifier: Random forest (RF) with 10-fold cross-validation. Feature Analysis: SHAP [6] allowed direct evaluation of feature importance.
Results and Discussion
Accuracy levels of 82% when evaluated over a window of 5 seconds.
From the confusion matrix (Fig. 2), we see that the accuracy decreases where the CoP is more confined, see Fig. 3.
SHAP analysis showed that statistical features relating to force, CoP, spectral entropy and morphology of peaks played a major role in the classification.
This denotes a move away from the 'black box' approach of certain ML classifiers that may lead to a lack of trust. Random forest provide a good balance between accuracy and interpretability.
The dataset was named PID4TC (Pressure Insole Data for Task Classification) and is publicly available at [7].
Fig. 2: Confusion matrix for the 5 second window.
Fig. 3: Centre of mass path during the tasks.
Fig. 5: SHAP analysis of feature importance for the classification of all tasks (top) and manual handling (bottom).
Conclusions
Our results show that pressure insoles can be used for real-time task detection in industry settings. The non-invasiveness of pressure insoles mean they can be integrated into the personal protective equipment of smart industry workers for task detection and fatigue prediction and management.
Acknowledgements
References
European Union European Regional Development Fund
UCC

# AI-based Task Classification with Pressure Insoles for Occupational Safety

## Patricia OSullivan, Matteo Menolotto, Andrea Visentin, Brendan OFlynn, Dimitrios- Sokratis Komaris

Contact : andrea.visentin@ucc.ie

Pressure insoles allow for the collection of real-time pressure data inside and outside a laboratory setting as they are non-intrusive and can be simply integrated into everyday life. Occupational health and safety are relevant to the present research, where pressure insoles have been used for manual handling monitoring, detecting loss of balance events and construction activity recognition. Activity detection is important for the safety and well-being of workers, and the present study aims to 1) describe existing pressure insole and machine learning research and 2) use pressure insoles to detect the type of industry-related task an individual is performing by using an artificial intelligence-based classification technique, random forest. Twenty subjects wore loadsol pressure insoles and performed five tasks: standing, walking, pick and place, assembly, and manual handling. For each activity, statistical and morphological features were extracted to create a training dataset. The classifier performed with an accuracy of over 82%, using ten-fold cross-validation for a time window of 5 seconds, presenting the potential for task classification in edge-AI applications in smart manufacturing environments. The features extracted and the validation approach, in combination with the use of an explainable classifier, offer a notable contribution to the field of noninvasive human-centric data extraction for AI-based analysis in occupational health and safety since they increase transparency and, thereby, trust in the classifier decisions.

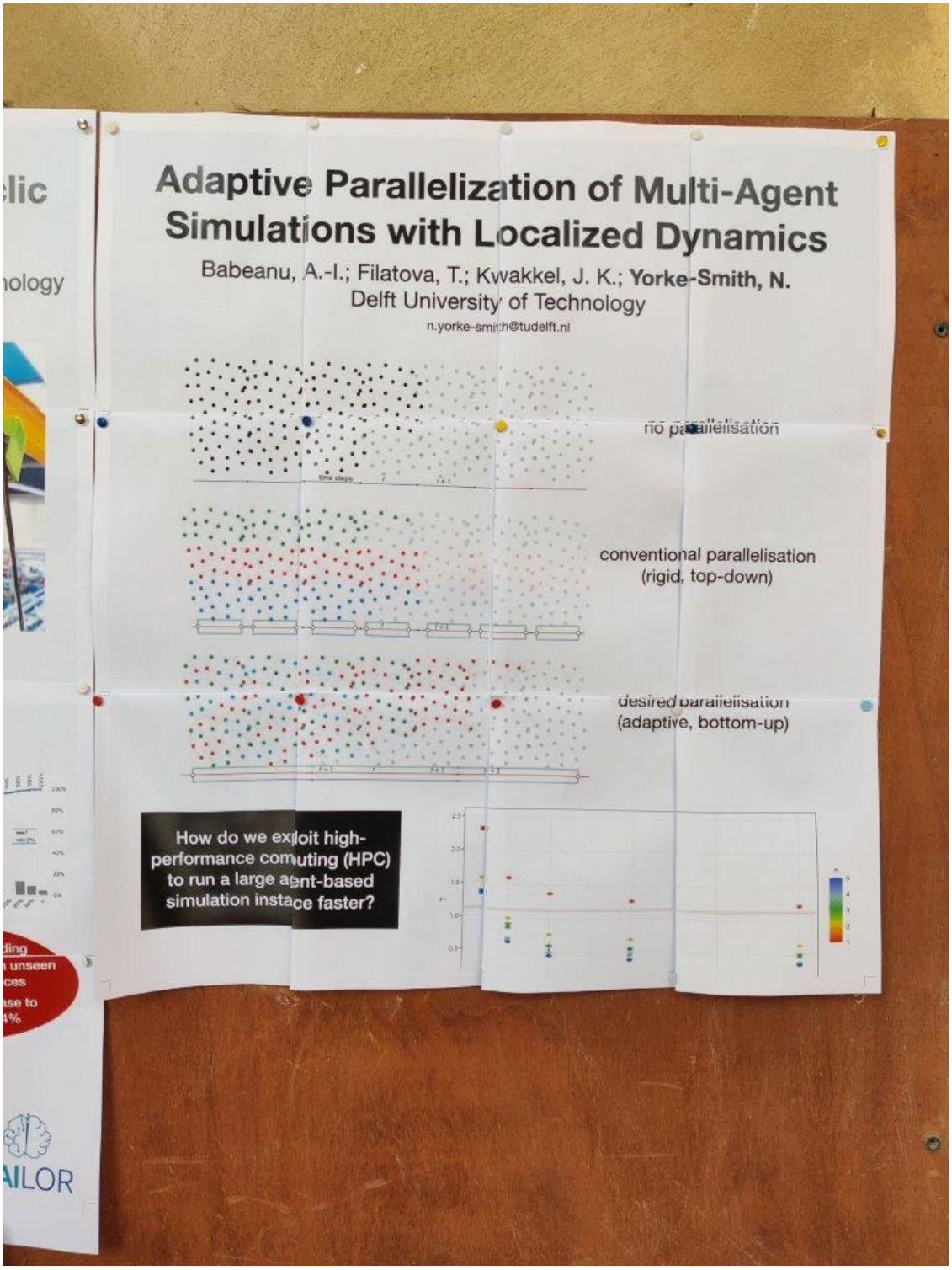

P27
Solving Complex Influence Diagrams
by Machine Learning
Steve Prestwich
University College Cork
Ireland
May 30, 2023
Solving Complex Influence Diagrams
by Machine Learning
Steve Prestwich UCC
Influence diagrams can model and solve probabilistic inference and decision making problems. They extend Bayesian Networks with decisions and utilities. Further extensions have been proposed.
I have a machine learning-based solver that can handle constraints (hard, soft, weighted, chance), multiple nonlinear objectives, multiple agents, partially observed variables and hybrid problems.
Motivation
Optimisation problems are ubiquitous in AI & OR and have specialised approaches: MILP, CP, DP, stochastic programming, chance-constrained programming, bilevel programming...
Some hybrids have been investigated (eg bilevel multi-objective stochastic integer linear programming) but no method exists for complex hybrids in general.
Eg how should we solve a bi-level optimisation problem whose leader is a non-linear stochastic program with chance constraints, and whose followers are a bi-objective CSP and a Weighted Max-SAT?
Relevant to Unifying Paradigms (WP4).
The solver
Inputs: (i) a simple text file describing the problem, and (ii) a C function to compute utilities.
To handle decision variables it uses Reinforcement Learning: infinite-step tabular SARSA with ε-soft action selection ($\gamma = 1$, $\alpha_t \to 0$) & random tile coding.
To handle random variable observations it uses sampling (rejection or Gibbs).
Examples
Standard influence diagram. A well-known example is the Oil Wildcatter:
(Plus utility and conditional probability tables.) We find the known optimal policy: apply seismic test, drill if the test result is open or closed.
Partially observed variables. Forgotten variables are handled simply by not hashing them during tile coding. On the famous pig breeding LIMID we find the known optimal policy: ignore first test and do not treat, follow the other two tests and treat if positive.
Multiple objectives. We use weighted metric scalarisation ($L_3$) with a utopian point $u^*$. On a bi-objective influence diagram for the Oil Wildcatter (the extra objective minimises pollution) we found 3 of the 4 Pareto-optimal policies in multiple runs.
Constraints. These are handled by adding objectives to minimise violations. We solve a 2-stage stochastic program and a chance-constrained program, so we can do safe reinforcement learning.
Relevant to Trustworthy AI (WP3).
Multiple agents. We give each agent its own Q-table and reward(s). On a bi-level program studied in at least 2 publications (3 agents with linear, quadratic and fractional objectives) we find the optimal policy in most runs.

# Solving Complex Optimisation Problems by Machine Learning

## Steve Prestwich

Contact : s.prestwich@cs.ucc.ie

Influence diagrams can model and solve probabilistic inference and decision making problems. They extend Bayesian Networks with decisions and utilities, and further extensions have been proposed. We describe a machine learning-based solver that can handle several extensions: constraints (hard, soft, weighted, chance), multiple nonlinear objectives, multiple decision makers and partially observed variables. We apply it to a 2-stage stochastic program, a chance-constrained program, a tri-level program, a limited-memory and a multi-objective influence diagram. This work relates to Unifying Paradigms (WP4) as it uses machine learning for a wide range of optimisation problems. It also relates to Trustworthy AI (WP3) as it can learn policies that satisfy constraints, to mitigate risk or guarantee a service level.

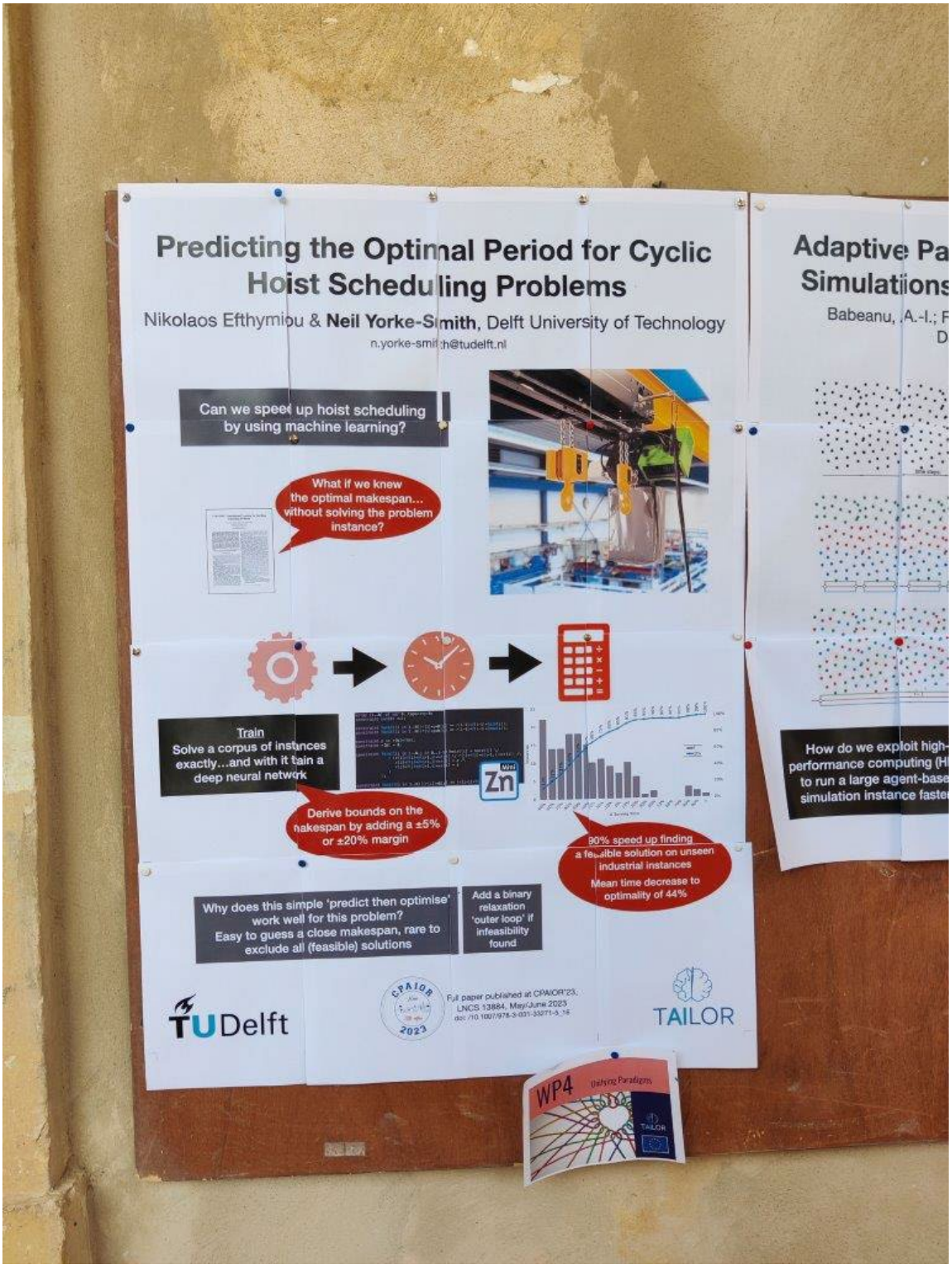

Adaptive Parallelization of Multi-Agent Simulations with Localized Dynamics
Babeanu, A.-I.; Filatova, T.; Kwakkel, J. K.; Yorke-Smith, N.
Delft University of Technology
n.yorke-smith@tudelft.nl
no parallelisation
conventional parallelisation (rigid, top-down)
desired parallelisation (adaptive, bottom-up)
How do we exploit high-performance computing (HPC) to run a large agent-based simulation instance faster?

# Adaptive Parallelization of Multi-Agent Simulations with Localized Dynamics

**Babeanu, A.-I.; Filatova, T.; Kwakkel, J. K.; Yorke-Smith, N.**

Contact : n.yorke-smith@tudelft.nl

Agent-based modelling constitutes a versatile approach to representing and simulating complex systems. Studying large-scale systems is challenging because of the computational time required for the simulation runs: scaling is at least linear in system size (number of agents). Given the inherently modular nature of MABSs, parallel computing is a natural approach to overcoming this challenge. However, because of the shared information and communication between agents, parellelization is not simple. We present a protocol for shared-memory, parallel execution of MABSs. This approach is useful for models that can be formulated in terms of sequential computations, and that involve updates that are localized, in the sense of involving small numbers of agents. The protocol has a bottom-up and asynchronous nature, allowing it to deal with heterogeneous computation in an adaptive, yet graceful manner. We illustrate the potential performance gains on exemplar cultural dynamics and disease spreading MABSs.

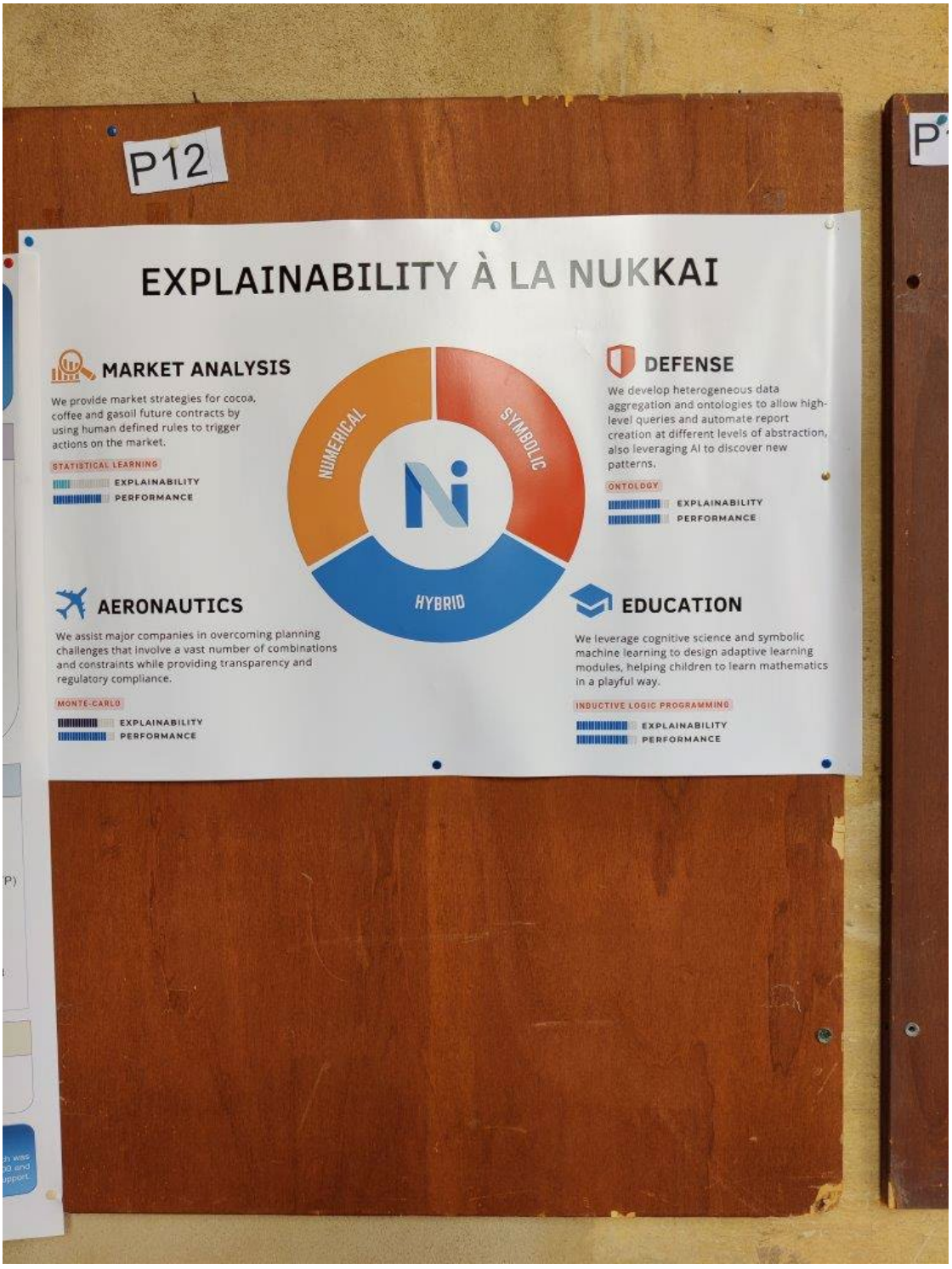

Predicting the Optimal Period for Cyclic Hoist Scheduling Problems
Nikolaos Efthymiou & Neil Yorke-Smith, Delft University of Technology
Can we speed up hoist scheduling by using machine learning?
What if we knew the optimal makespan... without solving the problem instance?
Train
Solve a corpus of instances exactly...and with it tain a deep neural network
Derive bounds on the makespan by adding a ±5% or ±20% margin
80% speed up finding a feasible solution on unseen industrial instances
Mean time decrease to optimality of 44%
Why does this simple 'predict then optimise' work well for this problem?
Easy to guess a close makespan, rare to exclude all (feasible) solutions
Add a binary relaxation 'outer loop' if infeasibility found
TUDelft
TAILOR
WP4
Adaptive Pa
Simulations
How do we exploit high-performance computing (H
to run a large agent-base
simulation instance faste

# Predicting the Optimal Period for Cyclic Hoist Scheduling Problems

## Efthymiou, N.; Yorke-Smith, N.

Contact : n.yorke-smith@tudelft.nl

Since combinatorial scheduling problems are usually NP-hard, this paper investigates whether machine learning (ML) can accelerate exact solving of a problem instance. We adopt supervised learning on a corpus of problem instances, to acquire a function that predicts the optimal makespan for a given instance. The learned predictor is invariant to the instance size as it uses statistics of instance attributes. We provide this prediction to a solving algorithm in the form of bounds on the objective function. Specifically, this approach is applied to the well-studied Cyclic Hoist Scheduling Problem (CHSP). The goal for a CHSP instance is to find a feasible schedule for a hoist which moves objects between tanks with minimal cyclic period. Taking an existing Constraint Programming (CP) model for this problem, and an exact CP-SAT solver, we implement a Deep Neural Network, a Random Forest and a Gradient Boosting Tree in order to predict the optimal period p. Experimental results find that, first, ML models (in particular DNNs), can be good predictors of the optimal p; and, second, providing tight bounds for p around the predicted value to an exact solver significantly reduces the solving time without compromising the optimality of the solutions.

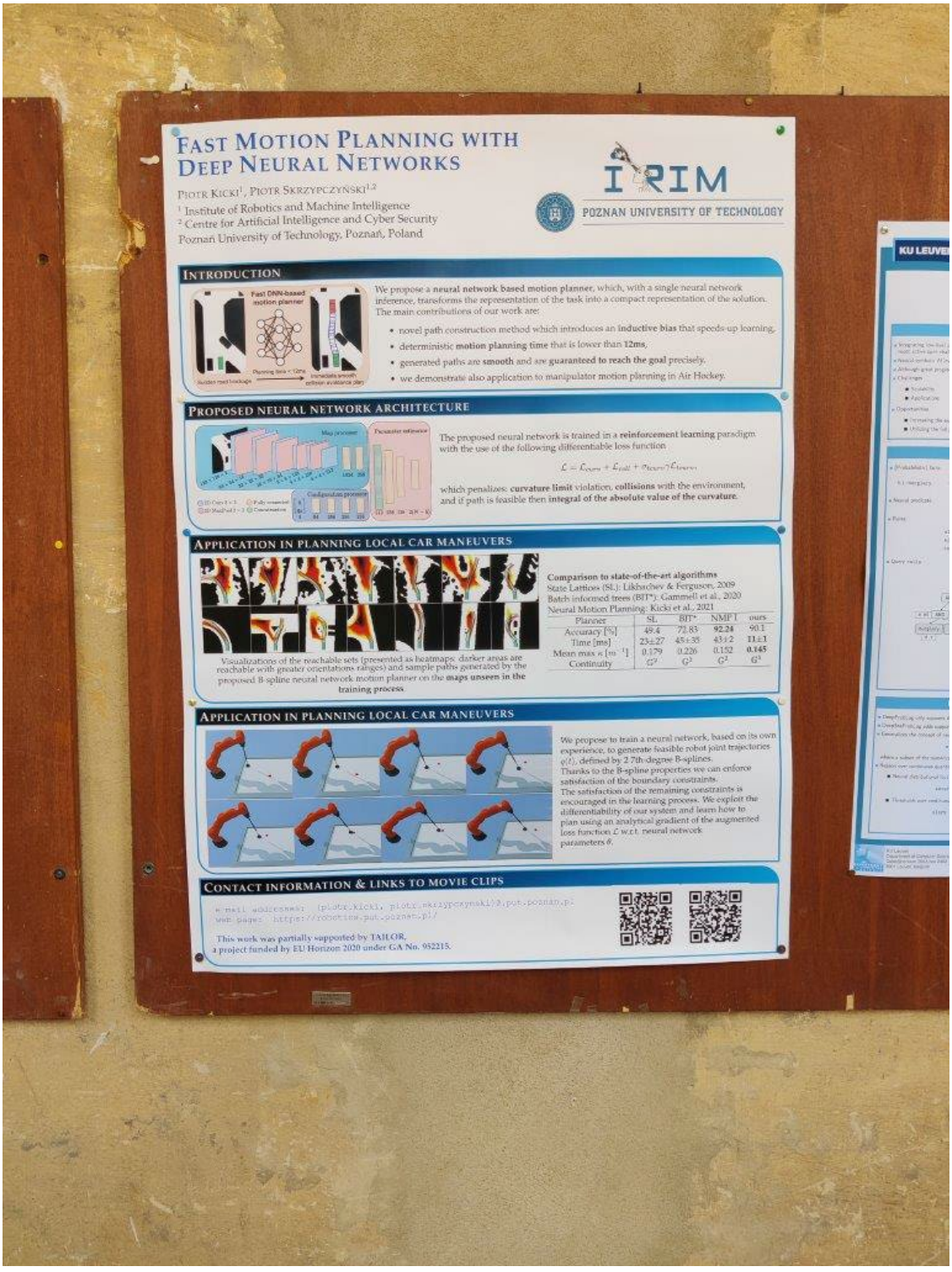

P12
EXPLAINABILITY À LA NUKKAI
MARKET ANALYSIS
We provide market strategies for cocoa, coffee and gasoil future contracts by using human defined rules to trigger actions on the market.
STATISTICAL LEARNING
EXPLAINABILITY
PERFORMANCE
NUMERICAL
SYMBOLIC
HYBRID
DEFENSE
We develop heterogeneous data aggregation and ontologies to allow high-level queries and automate report creation at different levels of abstraction, also leveraging AI to discover new patterns.
ONTOLOGY
EXPLAINABILITY
PERFORMANCE
AERONAUTICS
We assist major companies in overcoming planning challenges that involve a vast number of combinations and constraints while providing transparency and regulatory compliance.
MONTE-CARLO
EXPLAINABILITY
PERFORMANCE
EDUCATION
We leverage cognitive science and symbolic machine learning to design adaptive learning modules, helping children to learn mathematics in a playful way.
INDUCTIVE LOGIC PROGRAMMING
EXPLAINABILITY
PERFORMANCE

# Explainability à la NukkAI

## Veronique Ventos

Contact : vventos@nukk.ai

One year after the success of our Bridge robot Nook, which is capable of explaining its gameplay decision, we are here presenting the products developed at NukkAI that combine different paradigms to build high-performing and trustworthy AIs. NukkAI has developed its products for various industries such as transport logistics, finance, education, and defense, thanks to a close collaboration with experts from each field, thereby making the human factor the common denominator of our AIs. In terms of both their design and how they benefit the users, each product of NukkAI is thoughtfully crafted to meet the demands of the AI of tomorrow.

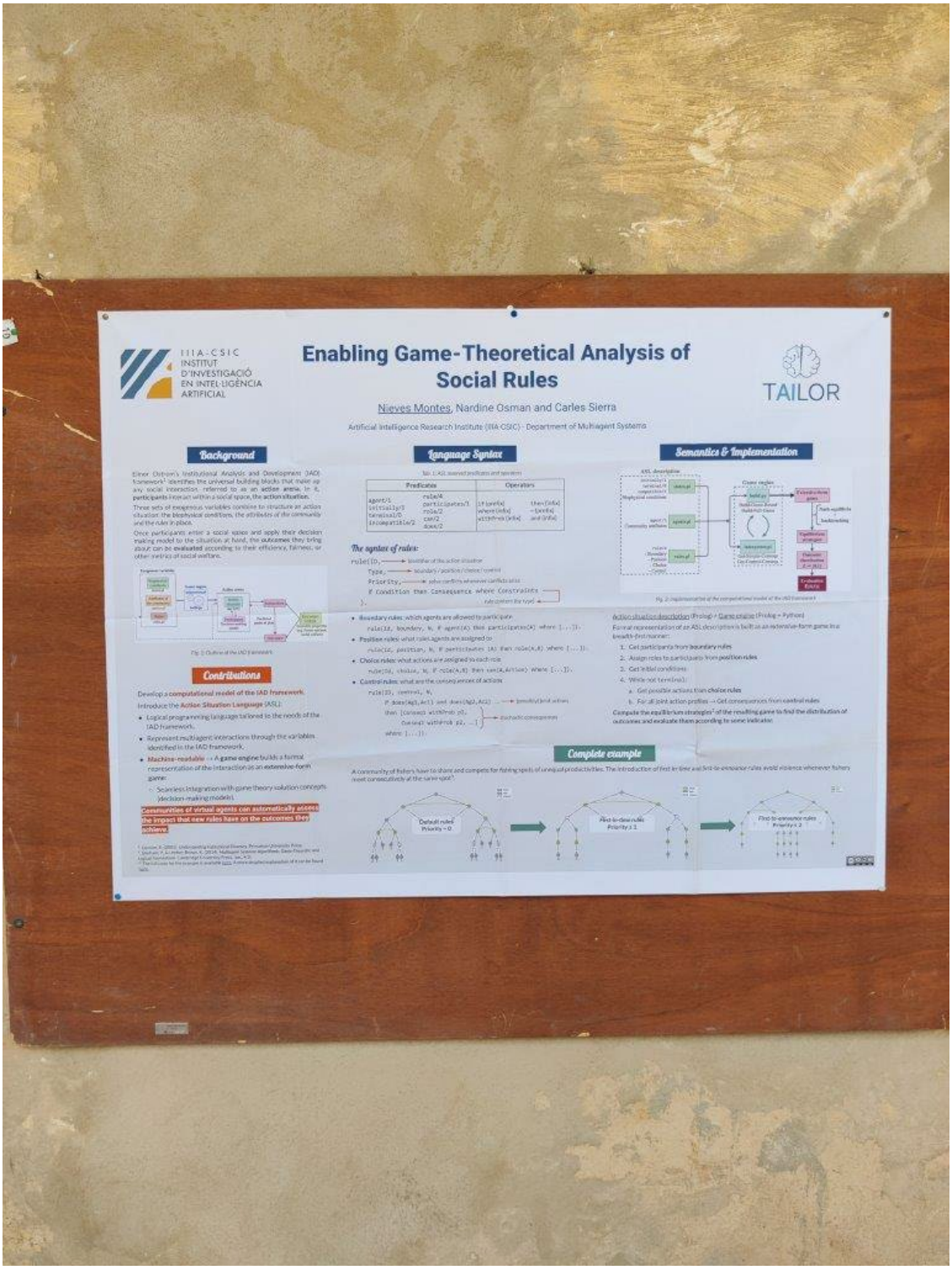

FAST MOTION PLANNING WITH DEEP NEURAL NETWORKS
PIOTR KICKI[1], PIOTR SKRZYPCZYŃSKI[1,2]
[1] Institute of Robotics and Machine Intelligence
[2] Centre for Artificial Intelligence and Cyber Security
Poznań University of Technology, Poznań, Poland
IRIM
POZNAN UNIVERSITY OF TECHNOLOGY
INTRODUCTION
We propose a neural network based motion planner, which, with a single neural network inference, transforms the representation of the task into a compact representation of the solution. The main contributions of our work are:
• novel path construction method which introduces an inductive bias that speeds-up learning,
• deterministic motion planning time that is lower than 12ms,
• generated paths are smooth and are guaranteed to reach the goal precisely,
• we demonstrate also application to manipulator motion planning in Air Hockey.
PROPOSED NEURAL NETWORK ARCHITECTURE
The proposed neural network is trained in a reinforcement learning paradigm with the use of the following differentiable loss function
which penalizes: curvature limit violation, collisions with the environment, and if path is feasible then integral of the absolute value of the curvature.
APPLICATION IN PLANNING LOCAL CAR MANEUVERS
Visualizations of the reachable sets (presented as heatmaps: darker areas are reachable with greater orientations ranges) and sample paths generated by the proposed B-spline neural network motion planner on the maps unseen in the training process.
Comparison to state-of-the-art algorithms
State Lattices (SL): Likhachev & Ferguson, 2009
Batch informed trees (BIT*): Gammell et al., 2020
Neural Motion Planning: Kicki et al., 2021
Planner | SL | BIT* | NMPT | ours
Accuracy [%] | 48.4 | 72.83 | 92.24 | 90.1
Time [ms] | 23±27 | 45±35 | 43±2 | 11±1
Mean max κ [m⁻¹] | 0.179 | 0.226 | 0.152 | 0.145
Continuity | G¹ | G² | G² | G³
APPLICATION IN PLANNING LOCAL CAR MANEUVERS
We propose to train a neural network, based on its own experience, to generate feasible robot joint trajectories φ(t), defined by 2 7th-degree B-splines.
Thanks to the B-spline properties we can enforce satisfaction of the boundary constraints.
The satisfaction of the remaining constraints is encouraged in the learning process. We exploit the differentiability of our system and learn how to plan using an analytical gradient of the augmented loss function L w.r.t. neural network parameters θ.
CONTACT INFORMATION & LINKS TO MOVIE CLIPS
e-mail addresses: {piotr.kicki, piotr.skrzypczynski}@put.poznan.pl
web pages: http://robotics.put.poznan.pl/
This work was partially supported by TAILOR, a project funded by EU Horizon 2020 under GA No. 952215.

# Fast Motion Planning with Deep Neural Networks

## Piotr Skrzypczynski, Piotr Kicki

Contact : piotr.skrzypczynski@put.poznan.pl

We present a new approach to very fast motion planning using a neural network trained on example tasks using a variant of reinforcement learning with structural priors in the form of parametrized trajectories. This approach is exemplified on two different tasks: planning local manuvers of a self-driving car and kinodynamic planning for a robotic arm playing air hockey.

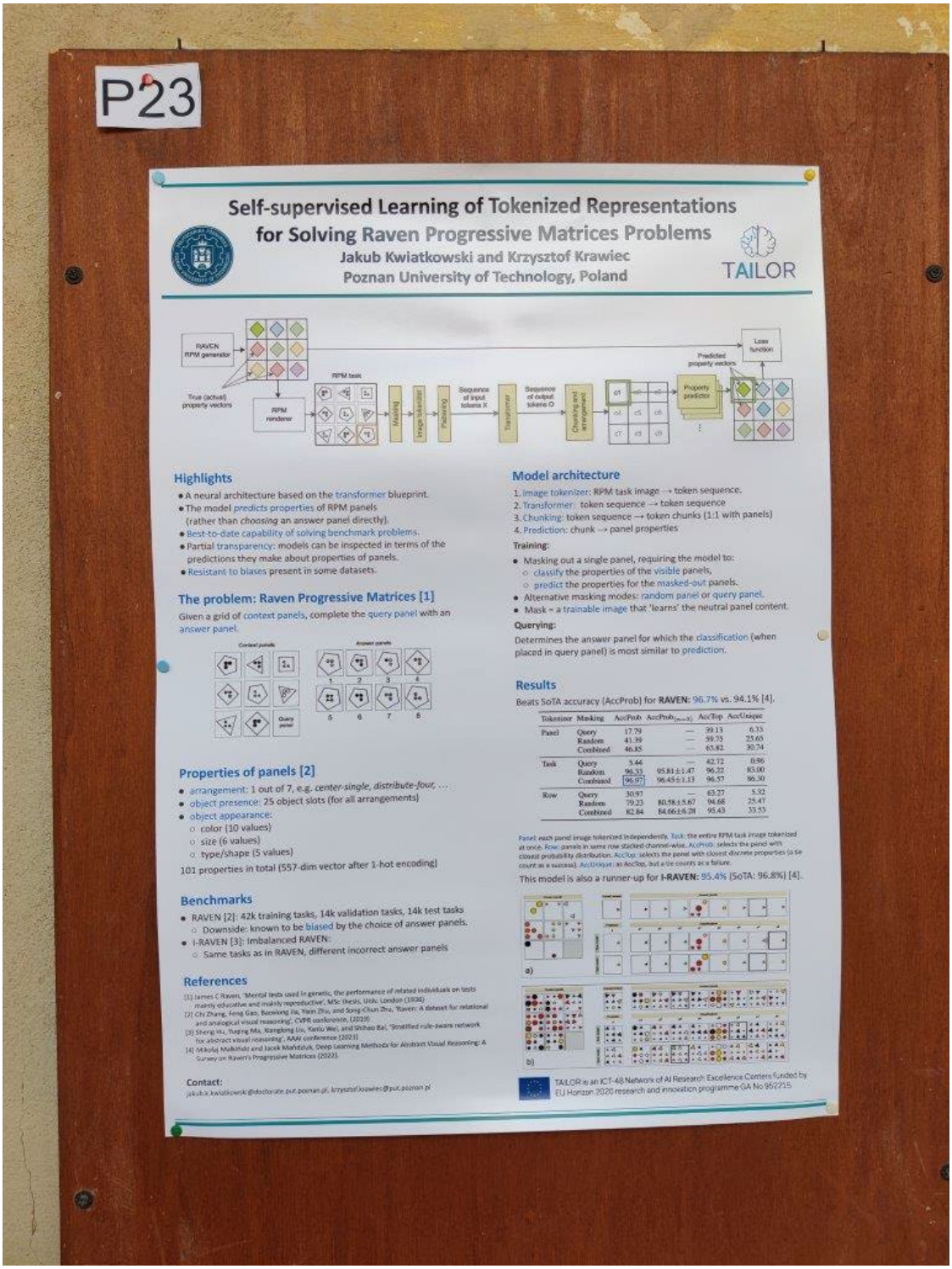

IIIA-CSIC
INSTITUT D'INVESTIGACIÓ EN INTEL·LIGÈNCIA ARTIFICIAL
Enabling Game-Theoretical Analysis of Social Rules
Nieves Montes, Nardine Osman and Carles Sierra
Artificial Intelligence Research Institute (IIIA-CSIC) · Department of Multiagent Systems
TAILOR
Background
Language Syntax
Semantics & Implementation
Contributions
Complete example

# Enabling Game-Theoretical Analysis of Social Rules

## Nieves Montes, Nardine Osman and Carles Sierra

Contact : sierra@iiia.csic.es

The Institutional Analysis and Development (IAD) framework developed by Elinor Ostrom and colleagues provides great conceptual clarity on the immensely varied topic of social interactions. In this work, we propose a computational model to examine the impact that any of the variables outlined in the IAD framework has on the resulting social interactions. Of particular interest are the rules adopted by a community of agents, as they are the variables most susceptible to change in the short term. To provide systematic descriptions of social interactions, we define the Action Situation Language (ASL) and provide a game engine capable of automatically generating formal game-theoretical models out of ASL descriptions. Then, by incorporating any agent decision-making models, the connection from a rule configuration description to the outcomes encouraged by it is complete. Overall, our model enables any community of agents to perform what-if analysis, where they can foresee and examine the impact that a set of regulations will have on the social interaction they are engaging in. Hence, they can decide whether their implementation is desirable.

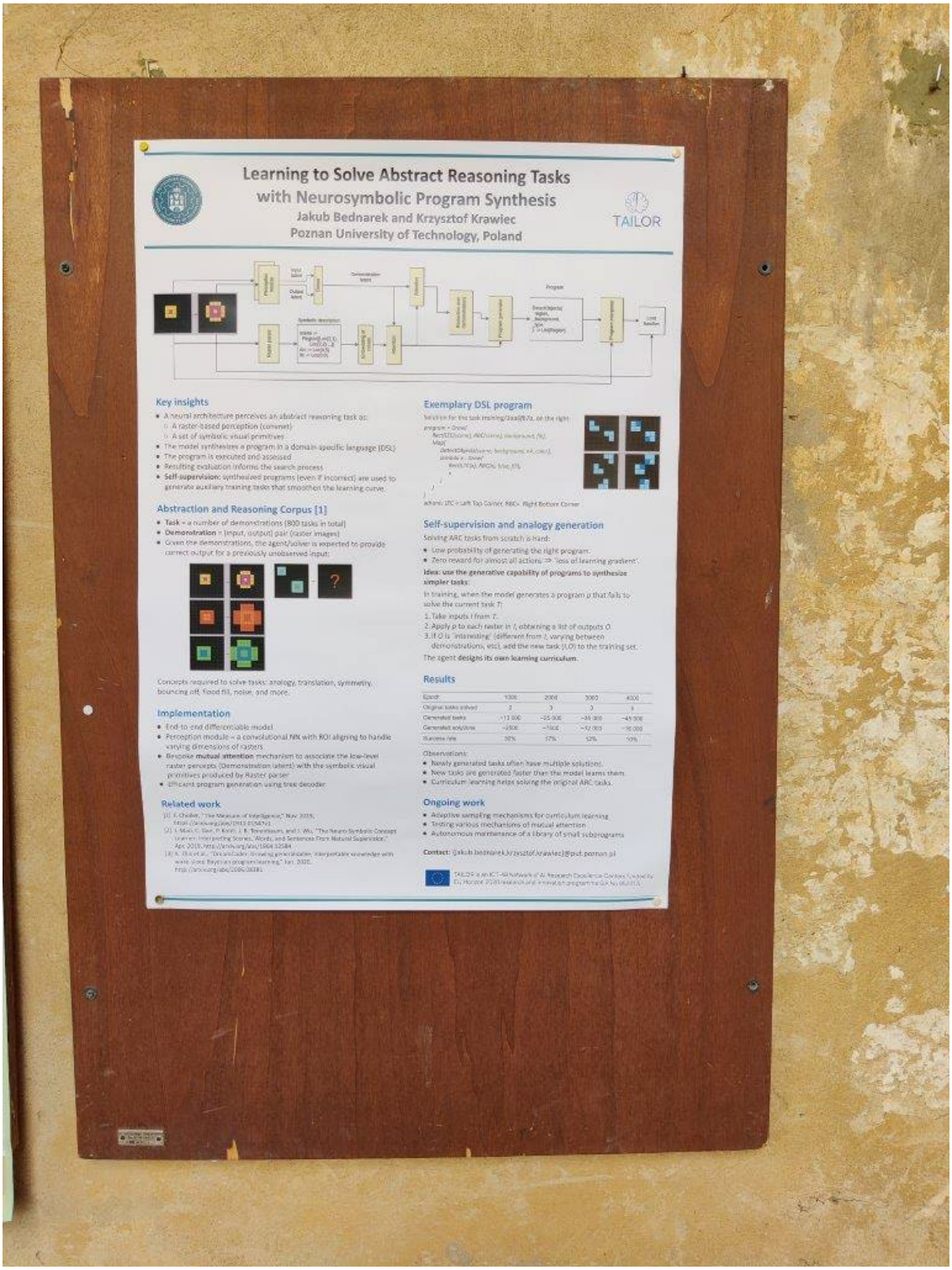

P23
Self-supervised Learning of Tokenized Representations for Solving Raven Progressive Matrices Problems
Jakub Kwiatkowski and Krzysztof Krawiec
Poznan University of Technology, Poland
TAILOR
Highlights
A neural architecture based on the transformer blueprint.
The model predicts properties of RPM panels (rather than choosing an answer panel directly).
Best-to-date capability of solving benchmark problems.
Partial transparency: models can be inspected in terms of the predictions they make about properties of panels.
Resistant to biases present in some datasets.
The problem: Raven Progressive Matrices [1]
Given a grid of context panels, complete the query panel with an answer panel.
Properties of panels [2]
arrangement: 1 out of 7, e.g. center-single, distribute-four, ...
object presence: 25 object slots (for all arrangements)
object appearance:
color (10 values)
size (6 values)
type/shape (5 values)
101 properties in total (557-dim vector after 1-hot encoding)
Benchmarks
RAVEN [2]: 42k training tasks, 14k validation tasks, 14k test tasks
Downside: known to be biased by the choice of answer panels.
I-RAVEN [3]: Imbalanced RAVEN:
Same tasks as in RAVEN, different incorrect answer panels
References
Contact:
Model architecture
1. Image tokenizer: RPM task image → token sequence.
2. Transformer: token sequence → token sequence
3. Chunking: token sequence → token chunks (1:1 with panels)
4. Prediction: chunk → panel properties
Training:
Masking out a single panel, requiring the model to:
classify the properties of the visible panels,
predict the properties for the masked-out panels.
Alternative masking modes: random panel or query panel.
Mask = a trainable image that 'learns' the neutral panel content.
Querying:
Determines the answer panel for which the classification (when placed in query panel) is most similar to prediction.
Results
Beats SoTA accuracy (AccProb) for RAVEN: 96.7% vs. 94.1% [4].
This model is also a runner-up for I-RAVEN: 95.4% (SoTA: 96.8%) [4].
TAILOR is an ICT-48 Network of AI Research Excellence Centers funded by EU Horizon 2020 research and innovation programme GA No 952215

# Self-supervised Learning of Tokenized Representations for Solving Raven Progressive Matrices Problems

## Jakub Kwiatkowski and Krzysztof Krawiec

Contact : krawiec@cs.put.poznan.pl

One of the challenges in learning to perform abstract reasoning is that problems are often posed as monolithic tasks, with no intermediate subgoals. In Raven Progressive Matrices (RPM), the task is to choose one of the available answers given a context, where both contexts and answers are composite images featuring multiple objects in various spatial arrangements. As this high-level goal is the only guidance available, learning is challenging and most contemporary solvers tend to be opaque. In this study, we propose a deep learning architecture based on the transformer blueprint which, rather than directly making the above choice, predicts the visual properties of individual objects and their arrangements. The multidimensional predictions obtained in this way are then directly juxtaposed to choose the answer. We consider a few ways in which the model parses the visual input into tokens and several regimes of masking parts of the input in self-supervised training. In experimental assessment, the models not only outperform state-of-the-art methods but also provide interesting insights and partial explanations about the inference. The design of the method also makes it immune to biases that are known to exist in some RPM benchmarks.

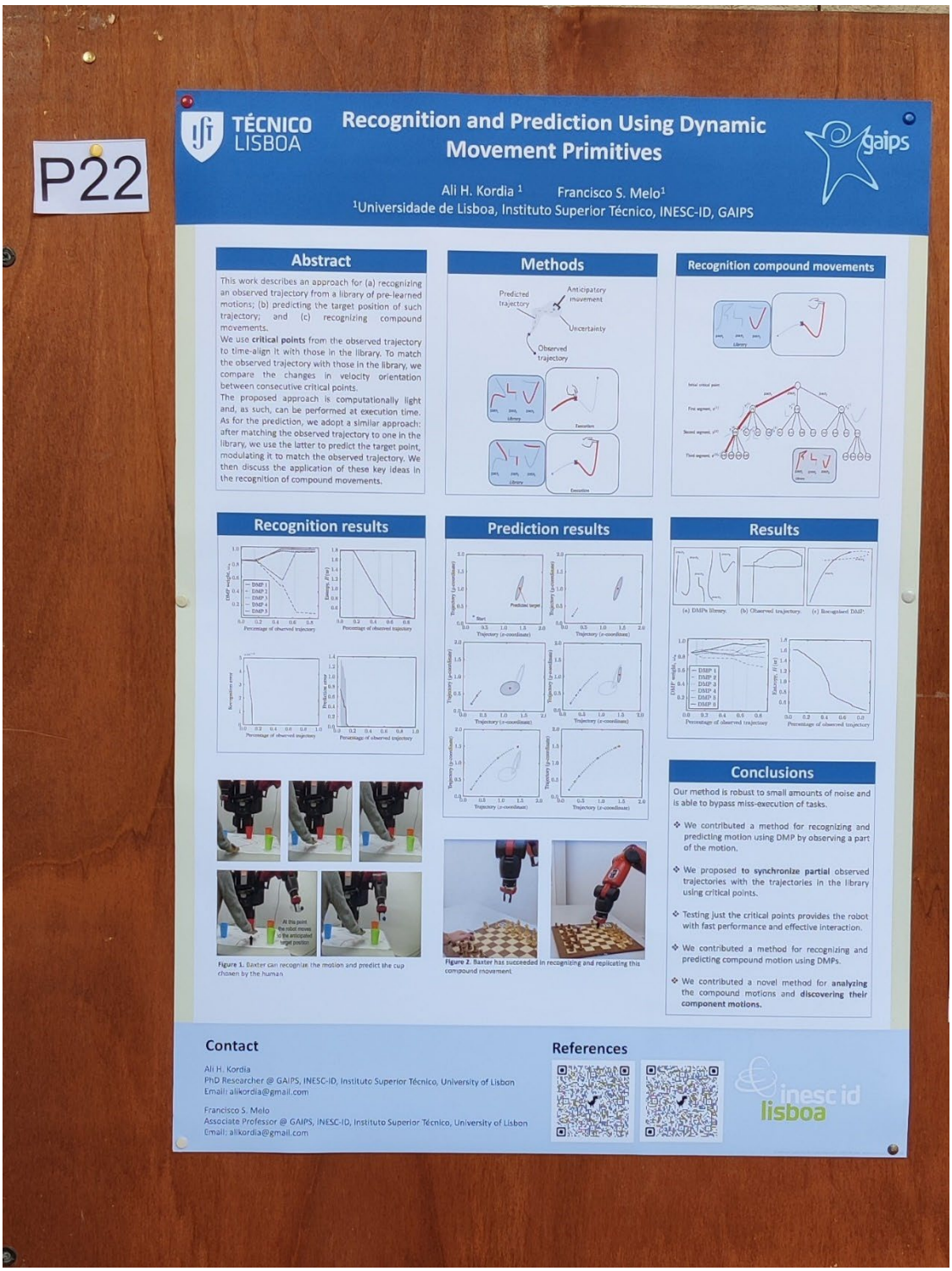

Learning to Solve Abstract Reasoning Tasks
with Neurosymbolic Program Synthesis
Jakub Bednarek and Krzysztof Krawiec
Poznan University of Technology, Poland
TAILOR
Key insights
A neural architecture perceives an abstract reasoning task as:
A raster-based perception (convnet)
A set of symbolic visual primitives
The model synthesizes a program in a domain-specific language (DSL)
The program is executed and assessed
Resulting evaluation informs the search process
Self-supervision: synthesized programs (even if incorrect) are used to generate auxiliary training tasks that smoothen the learning curve.
Abstraction and Reasoning Corpus [1]
Task = a number of demonstrations (800 tasks in total)
Demonstration = [input, output] pair (raster images)
Given the demonstrations, the agent/solver is expected to provide correct output for a previously unobserved input:
Concepts required to solve tasks: analogy, translation, symmetry, bouncing off, flood fill, noise, and more.
Implementation
End-to-end differentiable model
Perception module – a convolutional NN with ROI aligning to handle varying dimensions of rasters
Bespoke mutual attention mechanism to associate the low-level raster percepts (Demonstration latent) with the symbolic visual primitives produced by Raster parser
Efficient program generation using trie decoder
Related work
Exemplary DSL program
Self-supervision and analogy generation
Solving ARC tasks from scratch is hard:
Low probability of generating the right program.
Zero reward for almost all actions ⇒ 'loss of learning gradient'.
Idea: use the generative capability of programs to synthesize simpler tasks:
In training, when the model generates a program p that fails to solve the current task T:
1. Take inputs I from T.
2. Apply p to each raster in I, obtaining a list of outputs O.
3. If O is 'interesting' (different from I, varying between demonstrations, etc), add the new task (I,O) to the training set.
The agent designs its own learning curriculum.
Results
Epoch
Original tasks solved
Generated tasks
Generated solutions
Success rate
Observations:
Newly generated tasks often have multiple solutions.
New tasks are generated faster than the model learns them.
Curriculum learning helps solving the original ARC tasks.
Ongoing work
Adaptive sampling mechanisms for curriculum learning
Testing various mechanisms of mutual attention
Autonomous maintenance of a library of small subprograms
Contact: {jakub.bednarek,krzysztof.krawiec}@put.poznan.pl

# Learning to Solve Abstract Reasoning Tasks with Neurosymbolic Program Synthesis

## Jakub Bednarek and Krzysztof Krawiec

Contact : krawiec@cs.put.poznan.pl

We approach the problem of solving tasks from the Abstraction and Reasoning Corpus/Challenge (ARC), which require distilling a transformation pattern from a handful of visual input-output demonstrations and applying that transformation to a new query raster. The proposed approach is a neurosymbolic system that, given the demonstrations, synthesizes a program in a domain-specific language that captures a wide range of elementary concepts, features, and manipulations. The synthesized program is then executed on the query raster and the result of execution submitted to the oracle that verifies its validity. The system perceives the problem as raster images as well as a low-level symbolic representation automatically parsed from the rasters. Training involves self-supervision: the synthesized programs generate new tasks that smoothen the learning curve.


P22
TÉCNICO LISBOA
Recognition and Prediction Using Dynamic Movement Primitives
gaips
Ali H. Kordia [1] Francisco S. Melo[1]
[1]Universidade de Lisboa, Instituto Superior Técnico, INESC-ID, GAIPS
Abstract
This work describes an approach for (a) recognizing an observed trajectory from a library of pre-learned motions; (b) predicting the target position of such trajectory; and (c) recognizing compound movements.
We use critical points from the observed trajectory to time-align it with those in the library. To match the observed trajectory with those in the library, we compare the changes in velocity orientation between consecutive critical points.
The proposed approach is computationally light and, as such, can be performed at execution time. As for the prediction, we adopt a similar approach: after matching the observed trajectory to one in the library, we use the latter to predict the target point, modulating it to match the observed trajectory. We then discuss the application of these key ideas in the recognition of compound movements.
Methods
Predicted trajectory
Anticipatory movement
Uncertainty
Observed trajectory
Recognition compound movements
Recognition results
Prediction results
Results
Conclusions
Our method is robust to small amounts of noise and is able to bypass miss-execution of tasks.
We contributed a method for recognizing and predicting motion using DMP by observing a part of the motion.
We proposed to synchronize partial observed trajectories with the trajectories in the library using critical points.
Testing just the critical points provides the robot with fast performance and effective interaction.
We contributed a method for recognizing and predicting compound motion using DMPs.
We contributed a novel method for analyzing the compound motions and discovering their component motions.
Figure 1. Baxter can recognize the motion and predict the cup chosen by the human
Figure 2. Baxter has succeeded in recognizing and replicating this compound movement
Contact
Ali H. Kordia
PhD Researcher @ GAIPS, INESC-ID, Instituto Superior Técnico, University of Lisbon
Email: alikordia@gmail.com
Francisco S. Melo
Associate Professor @ GAIPS, INESC-ID, Instituto Superior Técnico, University of Lisbon
Email: alikordia@gmail.com
References
inesc id lisboa

# Recognition and Prediction Using Dynamic Movement Primitives

## Ali H. Kordia & Francisco S. Melo

Contact : alikordia@gmail.com

This work describes an approach for (a) recognizing an observed trajectory from a library of pre-learned motions; (b) predicting the target position of such trajectory; and (c) recognizing compound movements.

We use critical points from the observed trajectory to time-align it with those in the library. To match the observed trajectory with those in the library, we compare the changes in velocity orientation between consecutive critical points.

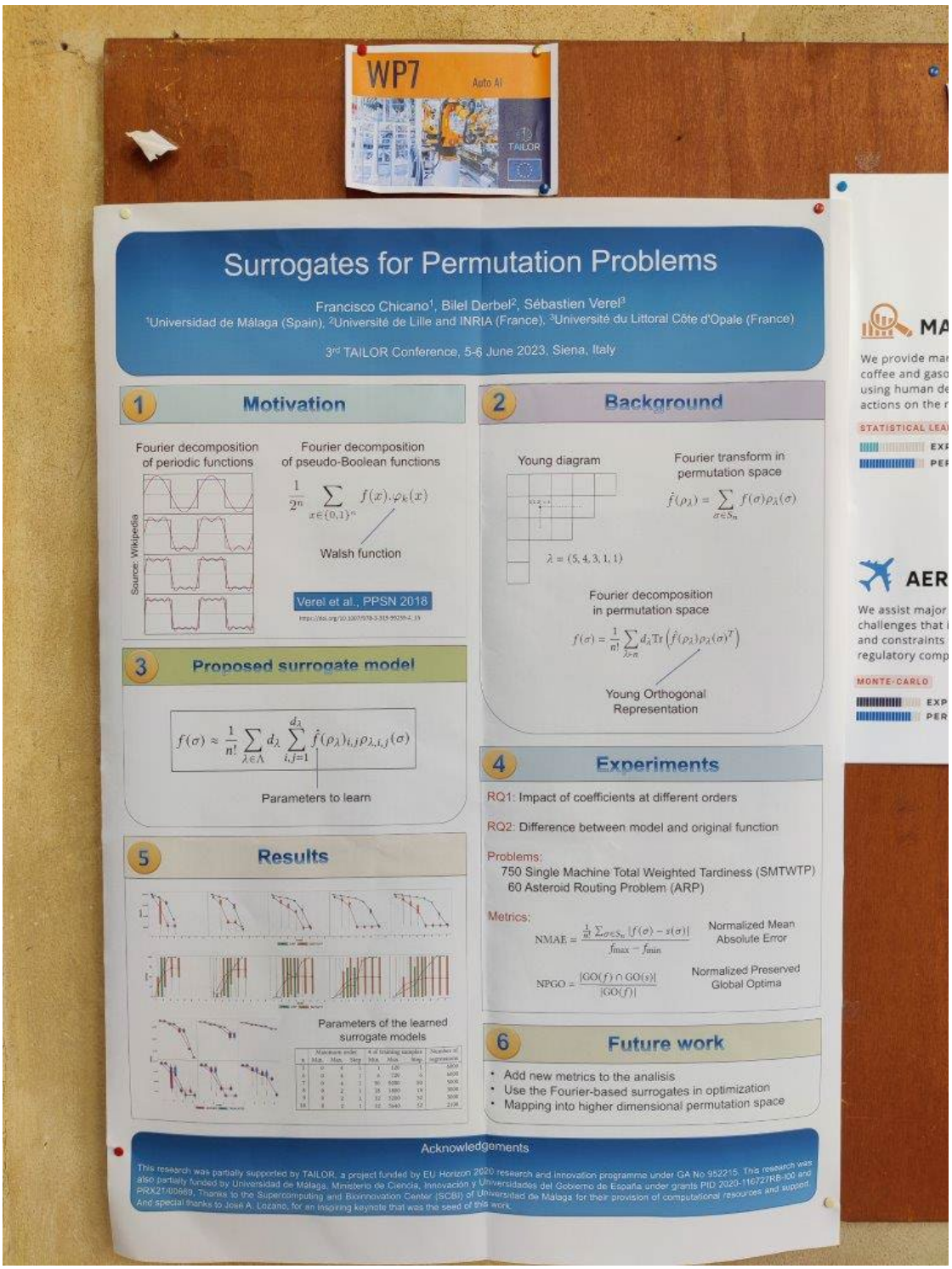
The proposed approach is computationally light and, as such, can be performed at execution time. As for the prediction, we adopt a similar approach: after matching the observed trajectory to one in the library, we use the latter to predict the target point, modulating it to match the observed trajectory. We then discuss the application of these key ideas in the recognition of compound movements.

# Time Series Forecasting for Parking Occupancy

## Jose Angel Morell, Zazaria Abdelmoiz Dahi, Francisco Chicano, Gabriel Luque and Enrique Alba

Contact : chicano@uma.es

The smart city concept refers principally to employing technology to deal with different problems surrounding the city and the citizens. Urban mobility is one of the most challenging aspects considering the logistical complexity as well as the ecological relapses. More specifically, parking is a daily tedious task that citizens confront especially considering the large number of vehicles compared to the limited parking, the rush hours peaks, etc. Forecasting parking occupancy might allow citizens to plan their parking better and therefore enhance their mobility. Time-series forecasting methods have proved their efficiency for such tasks, and this work goes in the same line by exploring how to provide more accurate parking occupancy forecasting. Concretely, its contributions stand in a complete pipeline, including i) the automatic extra-transform-load data module and ii) the time-series forecasting methods themselves, where four have been studied: one additive regression model (Prophet), the Seasonal Auto-Regressive Integrated Moving Average (SARIMAX), and two deep learning models, the Long Short Term memory neural networks (LSTM), and Neural Prophet. Experiments have been performed on data of 3 and 28 parking from the city of Malaga (Spain) and Birmingham (England) using data recorded through 6 months (June-November 2022) and two and a half months (October-December, 2016), for Malaga and Birmingham, respectively. The results showed that Prophet provided very competitive results compared to the literature.

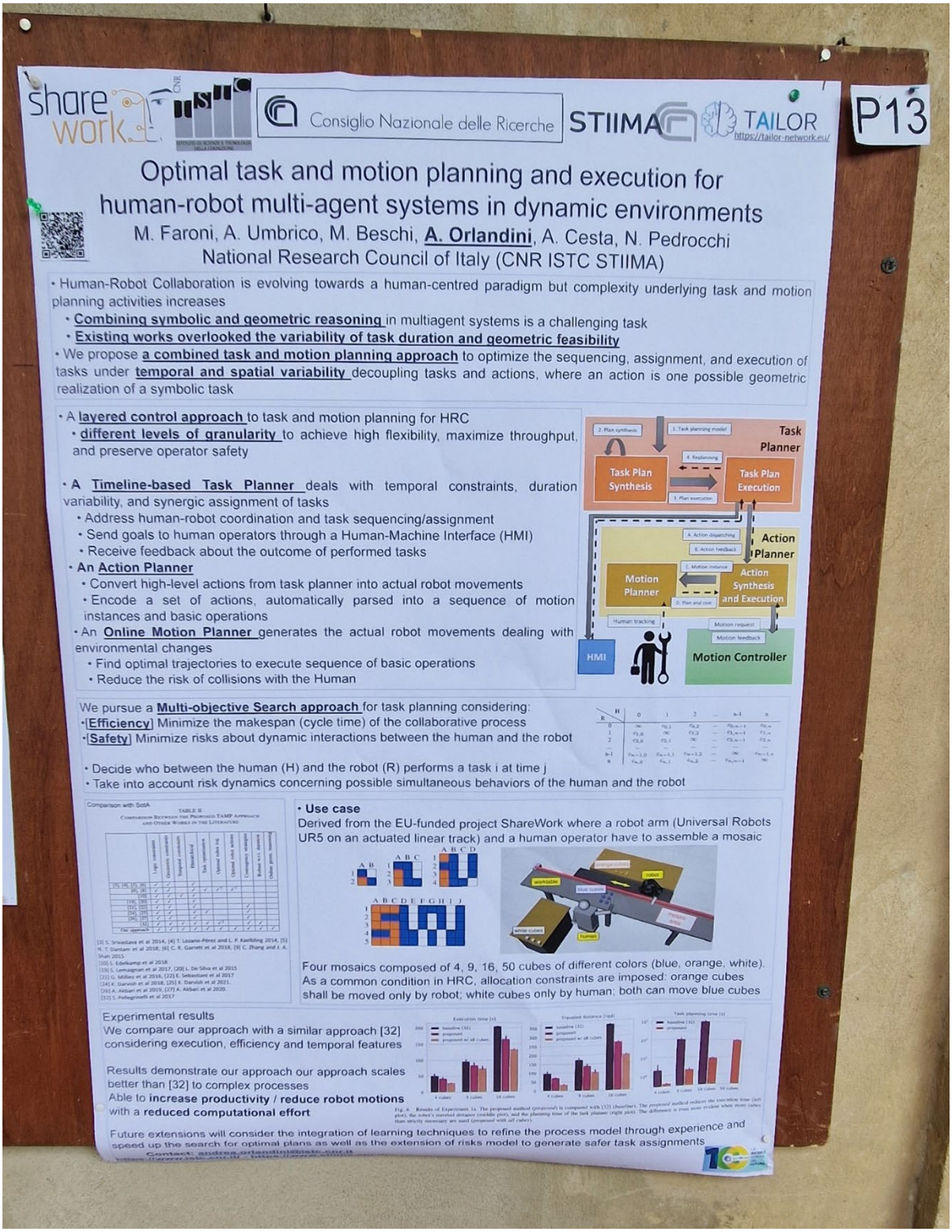

WP7
Auto AI
TAILOR
Surrogates for Permutation Problems
Francisco Chicano[1], Bilel Derbel[2], Sébastien Verel[3]
[1]Universidad de Málaga (Spain), [2]Université de Lille and INRIA (France), [3]Université du Littoral Côte d'Opale (France)
3rd TAILOR Conference, 5-6 June 2023, Siena, Italy
1 Motivation
Fourier decomposition of periodic functions
Source: Wikipedia
Fourier decomposition of pseudo-Boolean functions
$\frac{1}{2^n}\sum_{x\in\{0,1\}^n} f(x)\cdot\varphi_k(x)$
Walsh function
Verel et al., PPSN 2018
2 Background
Young diagram
$\lambda = (5,4,3,1,1)$
Fourier transform in permutation space
$\hat{f}(\rho_\lambda) = \sum_{\sigma\in S_n} f(\sigma)\rho_\lambda(\sigma)$
Fourier decomposition in permutation space
$f(\sigma) = \frac{1}{n!}\sum_{\lambda\vdash n} d_\lambda \mathrm{Tr}\left(\hat{f}(\rho_\lambda)\rho_\lambda(\sigma)^T\right)$
Young Orthogonal Representation
3 Proposed surrogate model
$f(\sigma) \approx \frac{1}{n!}\sum_{\lambda\in\Lambda} d_\lambda \sum_{i,j=1}^{d_\lambda} \hat{f}(\rho_\lambda)_{i,j}\rho_{\lambda,i,j}(\sigma)$
Parameters to learn
4 Experiments
RQ1: Impact of coefficients at different orders
RQ2: Difference between model and original function
Problems:
750 Single Machine Total Weighted Tardiness (SMTWTP)
60 Asteroid Routing Problem (ARP)
Metrics:
$NMAE = \frac{\frac{1}{n!}\sum_{\sigma\in S_n}|f(\sigma)-s(\sigma)|}{f_{max}-f_{min}}$ Normalized Mean Absolute Error
$NPGO = \frac{|GO(f)\cap GO(s)|}{|GO(f)|}$ Normalized Preserved Global Optima
5 Results
Parameters of the learned surrogate models
6 Future work
Add new metrics to the analisis
Use the Fourier-based surrogates in optimization
Mapping into higher dimensional permutation space
Acknowledgements
This research was partially supported by TAILOR, a project funded by EU Horizon 2020 research and innovation programme under GA No 952215. This research was also partially funded by Universidad de Málaga, Ministerio de Ciencia, Innovación y Universidades del Gobierno de España under grants PID 2020-116727RB-I00 and PRX21/00669. Thanks to the Supercomputing and Bioinnovation Center (SCBI) of Universidad de Málaga for their provision of computational resources and support. And special thanks to José A. Lozano, for an inspiring keynote that was the seed of this work.

# Surrogates for Permutation Problems

## Francisco Chicano, Bilel Derbel and Sébastien Verel

Contact : chicano@uma.es

We propose using surrogate functions based on Fourier transforms in the permutation space. We evaluate the similarity between the proposed surrogate models and the original objective function, and employ regression techniques to develop surrogate models using Fourier transform. The experimental analysis involves two permutation problems, the Asteroid Routing Problem and the Single Machine Total Weighted Tardiness, where the exact Fourier transform is unknown based on problem parameters.

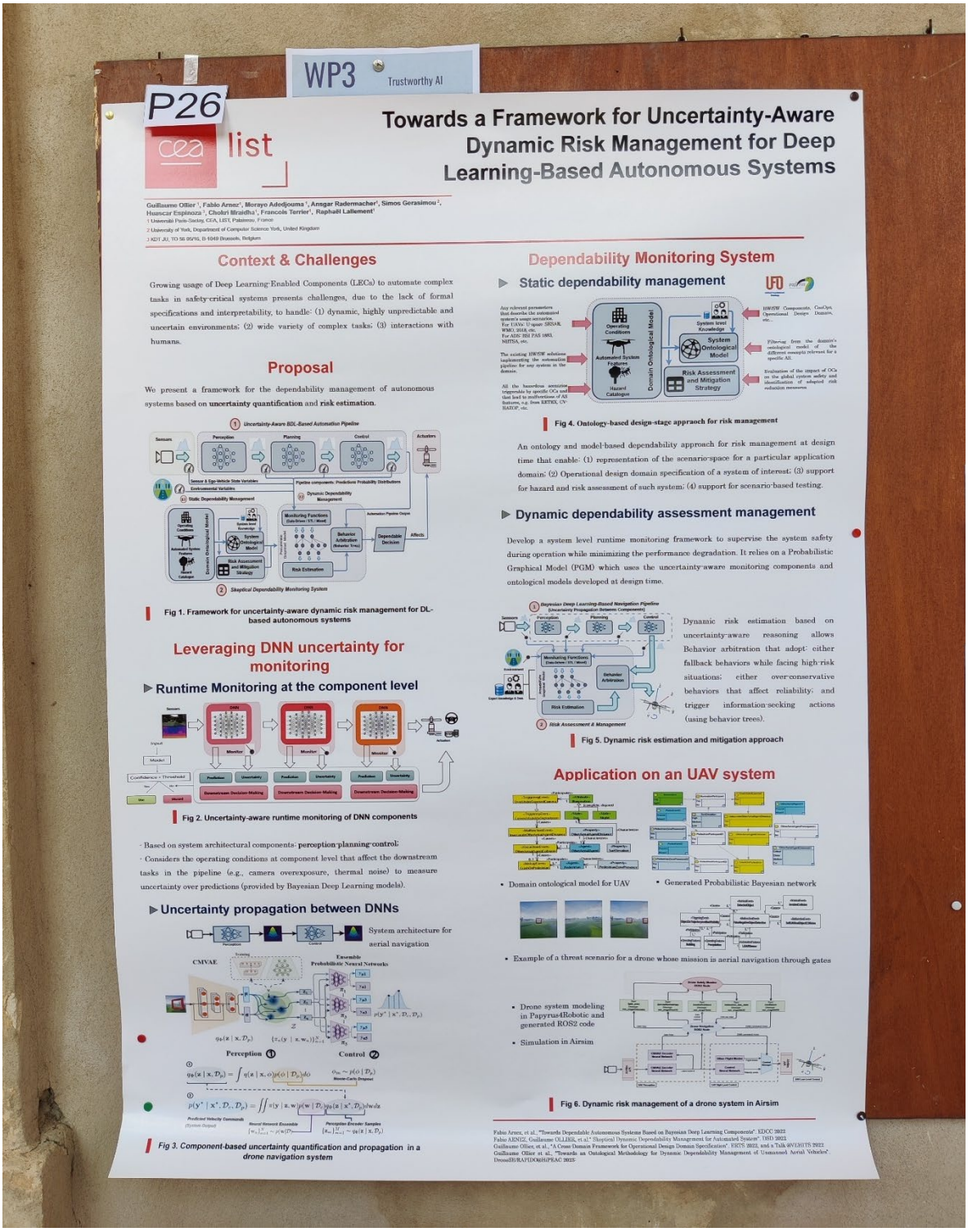

P13
Consiglio Nazionale delle Ricerche
STIIMA
TAILOR
https://tailor-network.eu/
Optimal task and motion planning and execution for human-robot multi-agent systems in dynamic environments
M. Faroni, A. Umbrico, M. Beschi, A. Orlandini, A. Cesta, N. Pedrocchi
National Research Council of Italy (CNR ISTC STIIMA)
• Human-Robot Collaboration is evolving towards a human-centred paradigm but complexity underlying task and motion planning activities increases
• Combining symbolic and geometric reasoning in multiagent systems is a challenging task
• Existing works overlooked the variability of task duration and geometric feasibility
• We propose a combined task and motion planning approach to optimize the sequencing, assignment, and execution of tasks under temporal and spatial variability decoupling tasks and actions, where an action is one possible geometric realization of a symbolic task
• A layered control approach to task and motion planning for HRC
• different levels of granularity to achieve high flexibility, maximize throughput, and preserve operator safety
• A Timeline-based Task Planner deals with temporal constraints, duration variability, and synergic assignment of tasks
• Address human-robot coordination and task sequencing/assignment
• Send goals to human operators through a Human-Machine Interface (HMI)
• Receive feedback about the outcome of performed tasks
• An Action Planner
• Convert high-level actions from task planner into actual robot movements
• Encode a set of actions, automatically parsed into a sequence of motion instances and basic operations
• An Online Motion Planner generates the actual robot movements dealing with environmental changes
• Find optimal trajectories to execute sequence of basic operations
• Reduce the risk of collisions with the Human
Task Planner
Task Plan Synthesis
Task Plan Execution
Action Planner
Motion Planner
Action Synthesis and Execution
HMI
Motion Controller
We pursue a Multi-objective Search approach for task planning considering:
•[Efficiency] Minimize the makespan (cycle time) of the collaborative process
•[Safety] Minimize risks about dynamic interactions between the human and the robot
• Decide who between the human (H) and the robot (R) performs a task i at time j
• Take into account risk dynamics concerning possible simultaneous behaviors of the human and the robot
• Use case
Derived from the EU-funded project ShareWork where a robot arm (Universal Robots UR5 on an actuated linear track) and a human operator have to assemble a mosaic
Four mosaics composed of 4, 9, 16, 50 cubes of different colors (blue, orange, white). As a common condition in HRC, allocation constraints are imposed: orange cubes shall be moved only by robot; white cubes only by human; both can move blue cubes
Experimental results
We compare our approach with a similar approach [32] considering execution, efficiency and temporal features
Results demonstrate our approach our approach scales better than [32] to complex processes
Able to increase productivity / reduce robot motions with a reduced computational effort
Future extensions will consider the integration of learning techniques to refine the process model through experience and speed up the search for optimal plans as well as the extension of risks model to generate safer task assignments

# Optimal task and motion planning and execution for human-robot multi-agent systems in dynamic environments

Marco Faroni, Alessandro Umbrico, Manuel Beschi, Andrea Orlandini, Amedeo Cesta, Nicola Pedrocchi

Contact : andrea.orlandini@istc.cnr.it

Combining symbolic and geometric reasoning in multi-agent systems is a challenging task that involves planning, scheduling, and synchronization problems. Existing works overlooked the variability of task duration and geometric feasibility that is intrinsic to these systems because of the interaction between agents and the environment. We propose a combined task and motion planning approach to optimize sequencing, assignment, and execution of tasks under temporal and spatial variability. The framework relies on decoupling tasks and actions, where an action is one possible geometric realization of a symbolic task. At the task level, timeline-based planning deals with temporal constraints, duration variability, and synergic assignment of tasks. At the action level, online motion planning plans for the actual movements dealing with environmental changes. We demonstrate the approach effectiveness in a collaborative manufacturing scenario, in which a robotic arm and a human worker shall assemble a mosaic in the shortest time possible. Compared with existing works, our approach applies to a broader range of applications and reduces the execution time of the process.


WP3
Trustworthy AI
P26
list
Towards a Framework for Uncertainty-Aware Dynamic Risk Management for Deep Learning-Based Autonomous Systems
Guillaume Ollier [1], Fabio Arnez [1], Morayo Adedjouma [1], Anegar Radermacher [1], Simos Gerasimou [2], Huascar Espinoza [1], Chokri Mraidha [1], Francois Terrier [1], Raphaël Lallement [3]
Context & Challenges
Growing usage of Deep Learning-Enabled Components (LECs) to automate complex tasks in safety-critical systems presents challenges, due to the lack of formal specifications and interpretability, to handle: (1) dynamic, highly unpredictable and uncertain environments; (2) wide variety of complex tasks; (3) interactions with humans.
Proposal
We present a framework for the dependability management of autonomous systems based on uncertainty quantification and risk estimation.
Fig 1. Framework for uncertainty-aware dynamic risk management for DL-based autonomous systems
Leveraging DNN uncertainty for monitoring
Runtime Monitoring at the component level
Fig 2. Uncertainty-aware runtime monitoring of DNN components
- Based on system architectural components: perception-planning-control.
- Considers the operating conditions at component level that affect the downstream tasks in the pipeline (e.g., camera overexposure, thermal noise) to measure uncertainty over predictions (provided by Bayesian Deep Learning models).
Uncertainty propagation between DNNs
System architecture for aerial navigation
CMVAE
Ensemble Probabilistic Neural Networks
Perception
Control
Fig 3. Component-based uncertainty quantification and propagation in a drone navigation system
Dependability Monitoring System
Static dependability management
Fig 4. Ontology-based design-stage appraoch for risk management
An ontology and model-based dependability approach for risk management at design time that enable: (1) representation of the scenario-space for a particular application domain; (2) Operational design domain specification of a system of interest; (3) support for hazard and risk assessment of such system; (4) support for scenario-based testing.
Dynamic dependability assessment management
Develop a system level runtime monitoring framework to supervise the system safety during operation while minimizing the performance degradation. It relies on a Probabilistic Graphical Model (PGM) which uses the uncertainty-aware monitoring components and ontological models developed at design time.
Dynamic risk estimation based on uncertainty-aware reasoning allows Behavior arbitration that adopt: either fallback behaviors while facing high-risk situations; either over-conservative behaviors that affect reliability; and trigger information-seeking actions (using behavior trees).
Fig 5. Dynamic risk estimation and mitigation approach
Application on an UAV system
- Domain ontological model for UAV
- Generated Probabilistic Bayesian network
- Example of a threat scenario for a drone whose mission is aerial navigation through gates
- Drone system modeling in Papyrus4Robotic and generated ROS2 code
- Simulation in Airsim
Fig 6. Dynamic risk management of a drone system in Airsim

# Towards a Framework for Uncertainty-Aware Dynamic Risk Management for Deep Learning-Based Autonomous Systems

Guillaume Ollier , Fabio Arnez, Morayo Adedjouma , Ansgar Radermacher1, Simos Gerasimou 2, Huascar Espinoza 3, Chokri Mraidha, Francois Terrier, Raphaël Lallement

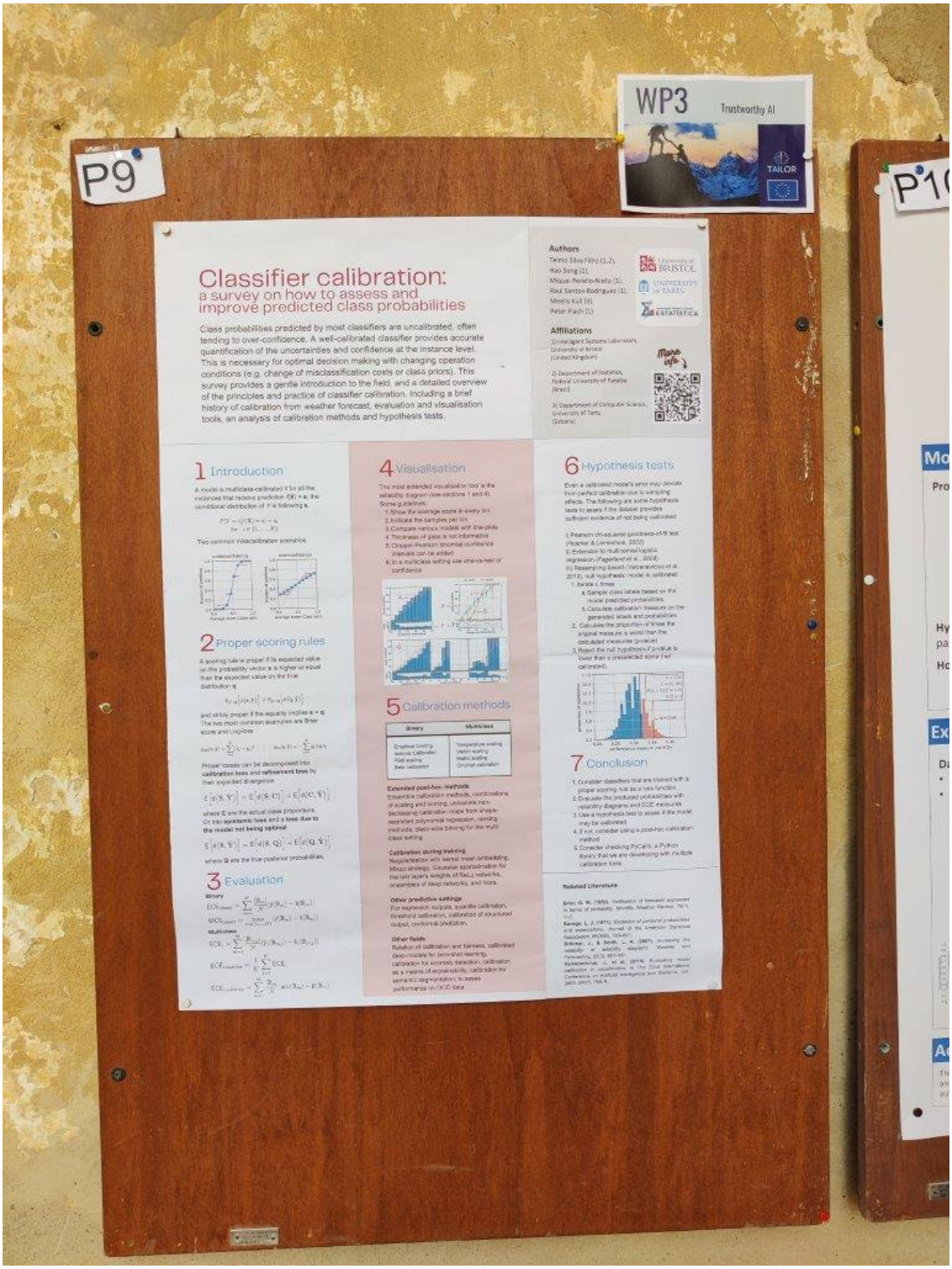
Contact : chokri.mraidha@cea.fr

Dynamic Dependability Management (DDM) is a promising approach to guarantee and monitor the ability of safety-critical Automated Systems ( AS) to deliver the intended service with an acceptable risk level. This poster presents our current work around building a DDM framework for AS using Learning-Enabled Components (LEC).

The framework relies on model-based engineering practices and ontological models to support the development process of such AS.

Starting from the characterization of the AS's operational domain and functional architecture, the framework enables the development of a runtime monitoring system on top of the AS design to supervise the system's safety during operation while minimizing performance degradation.

It relies on a Probabilistic Graphical Model (PGM) inferred from the Uncertainty-aware runtime monitoring components, ontology-based operational domain models, and risk assessment developed at design time. The approach is illustrated in a use case involving Unmanned Aerial Vehicles (UAVs).


P9
WP3
Trustworthy AI
TAILOR
Classifier calibration:
a survey on how to assess and improve predicted class probabilities
Class probabilities predicted by most classifiers are uncalibrated, often tending to over-confidence. A well-calibrated classifier provides accurate quantification of the uncertainties and confidence at the instance level. This is necessary for optimal decision making with changing operation conditions (e.g. change of misclassification costs or class priors). This survey provides a gentle introduction to the field, and a detailed overview of the principles and practice of classifier calibration. Including a brief history of calibration from weather forecast, evaluation and visualisation tools, an analysis of calibration methods and hypothesis tests.
Authors
Affiliations
1 Introduction
2 Proper scoring rules
3 Evaluation
4 Visualisation
5 Calibration methods
6 Hypothesis tests
7 Conclusion

# Classifier Calibration: A survey on how to assess and improve predicted class probabilities

**Telmo Silva Filho, Hao Song, Miquel Perello-Nieto, Raul Santos-Rodriguez, Meelis Kull, Peter Flach**

Contact : perello.nieto@gmail.com

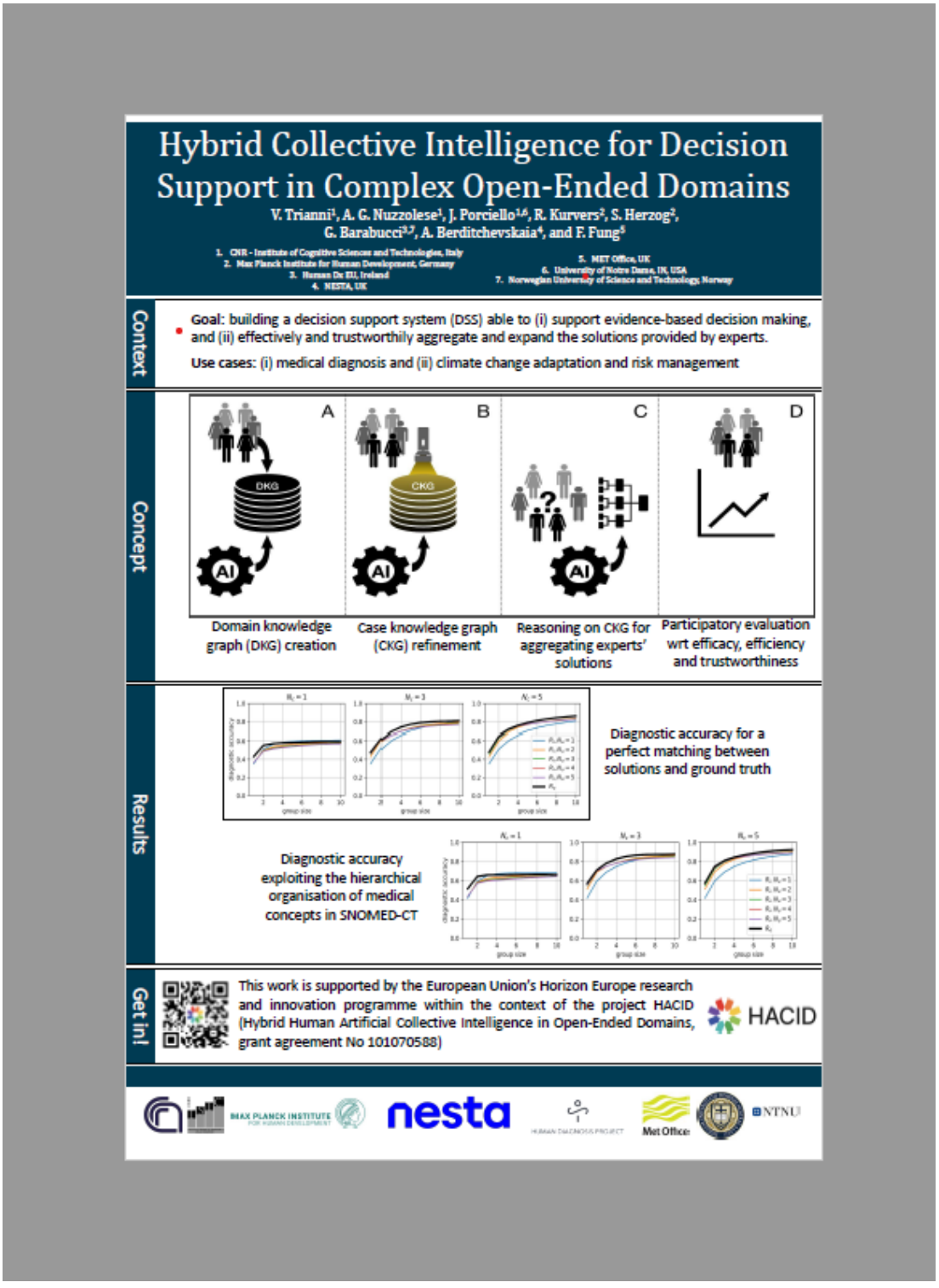
This paper provides both an introduction to and a detailed overview of the principles and practice of classifier calibration. A well-calibrated classifier correctly quantifies the level of uncertainty or confidence associated with its instance-wise predictions. This is essential for critical applications, optimal decision making, cost-sensitive classification, and for some types of context change. Calibration research has a rich history which predates the birth of machine learning as an academic field by decades. However, a recent increase in the interest on calibration has led to new methods and the extension from binary to the multiclass setting. The space of options and issues to consider is large, and navigating it requires the right set of concepts and tools. We provide both introductory material and up-to-date technical details of the main concepts and methods, including proper scoring rules and other evaluation metrics, visualisation approaches, a comprehensive account of post-hoc calibration methods for binary and multiclass classification, and several advanced topics.

# Neuro Symbolic Continual Learning: Knowledge, Reasoning Shortcuts and Concept Rehearsal

## Emanuele Marconato, Gianpaolo Bontempo, Elisa Ficarra, Simone Calderara, Andrea Passerini, Stefano Teso

Contact : stefano.teso@unitn.it

We introduce Neuro-Symbolic Continual Learning, where a model has to solve a sequence of neuro-symbolic tasks, that is, it has to map sub-symbolic inputs to high-level concepts and compute predictions by reasoning consistently with prior knowledge. Our key observation is that neuro-symbolic tasks, although different, often share concepts whose semantics remains stable over time. Traditional approaches fall short: existing continual strategies ignore knowledge altogether, while stock neuro-symbolic architectures suffer from catastrophic forgetting. We show that leveraging prior knowledge by combining neuro-symbolic architectures with continual strategies does help avoid catastrophic forgetting, but also that doing so can yield models affected by reasoning shortcuts. These undermine the semantics of the acquired concepts, even when detailed prior knowledge is provided upfront and inference is exact, and in turn continual performance. To overcome these issues, we introduce COOL, a COncept-level cOntinual Learning strategy tailored for neuro-symbolic continual problems that acquires high-quality concepts and remembers them over time. Our experiments on three novel benchmarks highlights how COOL attains sustained high performance on neuro-symbolic continual learning tasks in which other strategies fail.


Hybrid Collective Intelligence for Decision Support in Complex Open-Ended Domains
V. Trianni[1], A. G. Nuzzolese[1], J. Porciello[1,6], R. Kurvers[2], S. Herzog[2], G. Barabucci[3,7], A. Berditchevskaia[4], and F. Fung[5]
1. CNR - Institute of Cognitive Sciences and Technologies, Italy
2. Max Planck Institute for Human Development, Germany
3. Human Dx ELI, Ireland
4. NESTA, UK
5. MET Office, UK
6. University of Notre Dame, IN, USA
7. Norwegian University of Science and Technology, Norway
Context
Goal: building a decision support system (DSS) able to (i) support evidence-based decision making, and (ii) effectively and trustworthily aggregate and expand the solutions provided by experts.
Use cases: (i) medical diagnosis and (ii) climate change adaptation and risk management
Concept
A
B
C
D
DKG
CKG
AI
Domain knowledge graph (DKG) creation
Case knowledge graph (CKG) refinement
Reasoning on CKG for aggregating experts' solutions
Participatory evaluation wrt efficacy, efficiency and trustworthiness
Results
Diagnostic accuracy for a perfect matching between solutions and ground truth
Diagnostic accuracy exploiting the hierarchical organisation of medical concepts in SNOMED-CT
Get in!
This work is supported by the European Union's Horizon Europe research and innovation programme within the context of the project HACID (Hybrid Human Artificial Collective Intelligence in Open-Ended Domains, grant agreement No 101070588)
HACID
MAX PLANCK INSTITUTE
nesta
Met Office
NTNU

# Hybrid Collective Intelligence for Decision Support in Complex Open-Ended Domains

Vito TRIANNI, Andrea Giovanni NUZZOLESE, Jaron PORCIELLO, Ralf H.J.M. KURVERS, Stefan M. HERZOG, Gioele BARABUCCI, Aleksandra BERDITCHEVSKAIA and Fai FUNG

Contact : vito.trianni@istc.cnr.it

Human knowledge is growing exponentially, providing huge and sometimes contrasting evidence to support decision making in the realm of complex problems. To fight knowledge fragmentation, collective intelligence leverages groups of experts (possibly from diverse domains) that jointly provide solutions. However, to promote beneficial outcomes and avoid herding, it is necessary to (i) elicit diverse responses and (ii) suitably aggregate them in a collective solution. To this end, AI can help with dealing with large knowledge bases, as well as with reasoning on expert-provided knowledge to support decision-making. A hybrid human-artificial collective intelligence can leverage the complementarity of expert knowledge and machine processing to deal with complex problems. We discuss how such a hybrid human-artificial collective intelligence can be deployed to support decision processes, and we present case studies in two different domains: general medical diagnostics and climate change adaptation management.


P24
WP3
Trustworthy AI
WP4
Explainable Malware Detection
Activity Report of TAILOR WP3.1 Coordinated Action
The EMD Coordinated Action
Goals
Methods
Datasets
PE Malware Ontology
Roadmap
Results
Experiment 1: EMBER
Experiment 2: Fixes
Experiment 3: SOREL-20M
Experiment 4: fDLL
Ongoing & Future Work
Baseline experiments
Conc. Learners Comparison
Malware Families
Dynamic Analysis
KB Embedding
TAILOR
Join Our Coordinated Action
COMENIUS UNIVERSITY BRATISLAVA
Martin Homola, Ján Kľuka, Alexander Šimko, Peter Anthony, Daniel Trizna
Peter Švec, Štefan Balogh
Umberto Straccia*, Franca Debole*, Franco Alberto Cardillo*
Claudia d'Amato
(and others)
Learning Symbo
Trivial glob
Neuro-symbolic learning re
strategies:
pre-training
additional supervision
The "weak" a
Symbolic y
Neural
Problem of representa
Propo
Unifying neuro-symbolic
learning (bottom-up) into a

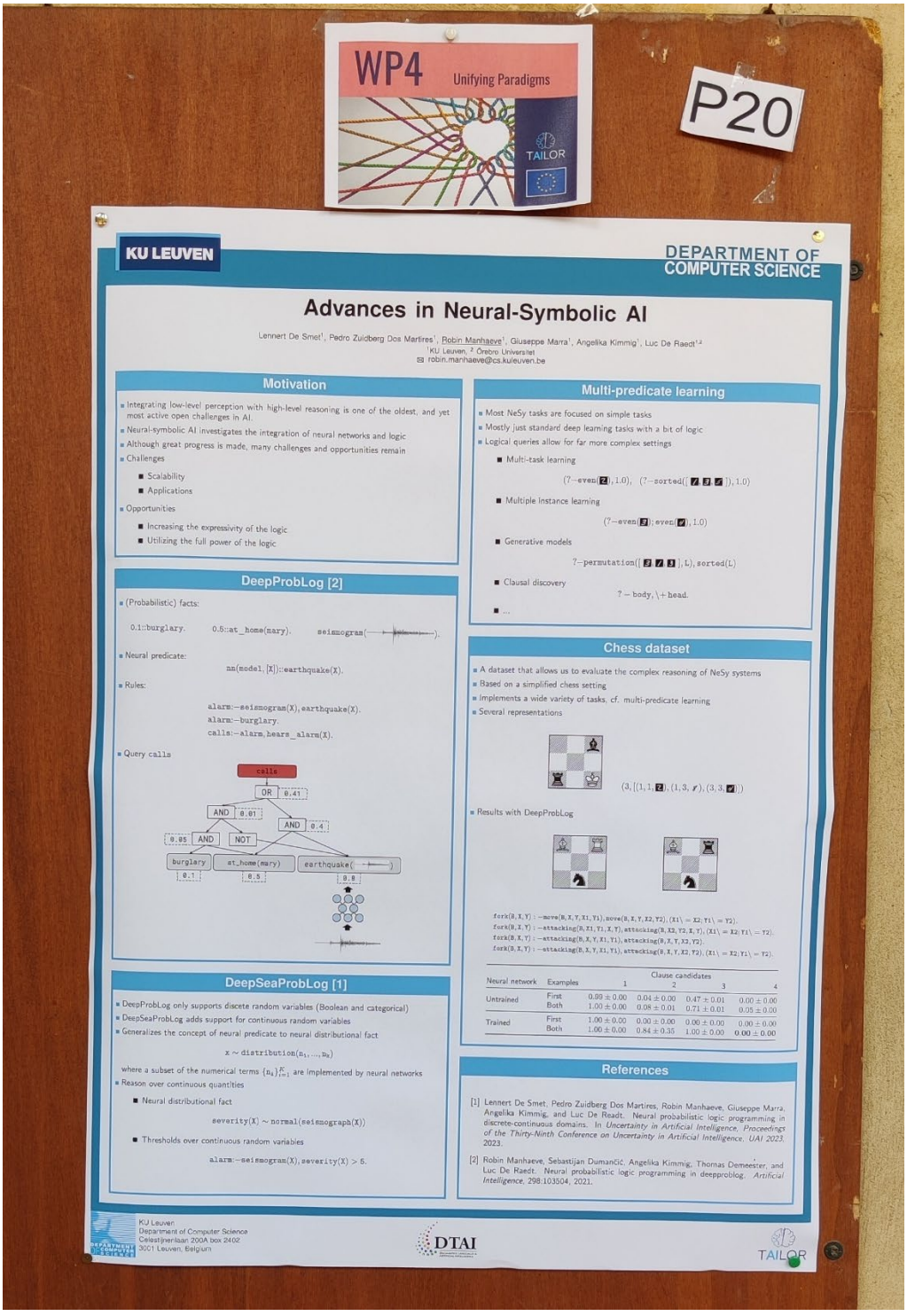
# Explainable Malware Detection

## Martin Homola, Jan Kluka, Alexander Simko, Peter Abnthony, Daniel Trizna, Peter Svec, Stefa Balogh, Umberto Straccia, Franca Debole, Franco Albert Cardillo, Claudia d'Amato

Contact: umberto.straccia@isti.cnr.it

This is an activity report of TAILOR WP3.1 Coordinated action "Explaiable Malware Detection". The goal of this coordinated action is to:

- Provide explainability on top of MD results, and answer why the sample is classified as malware
- Evaluate XAI methods on this use-case
- (Possibly)Extend classic MD toolkit towards explainability (e.g. via hybridization) and generalize of findings to other application domains

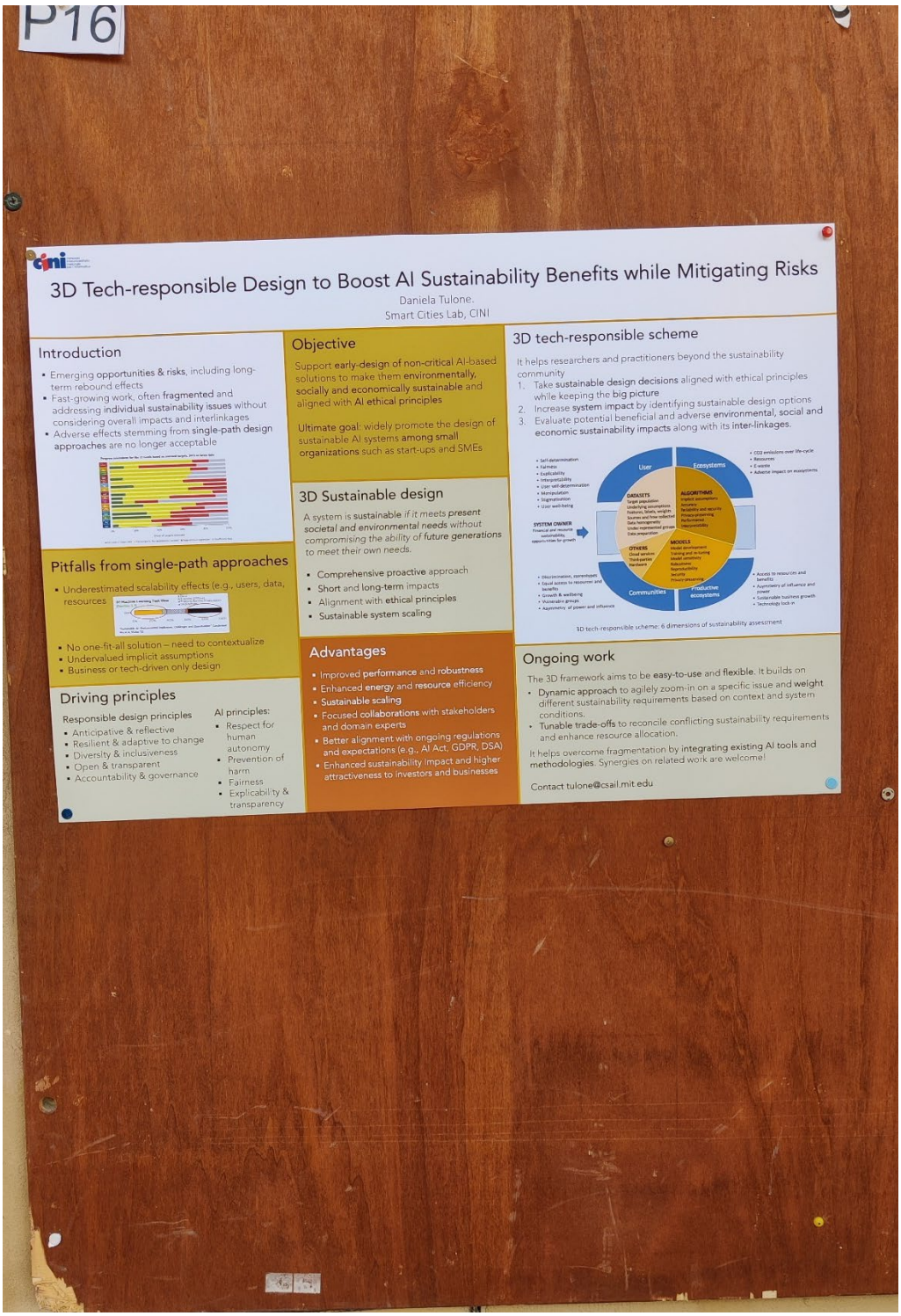

WP4
Unifying Paradigms
TAILOR
P20
KU LEUVEN
DEPARTMENT OF COMPUTER SCIENCE
Advances in Neural-Symbolic AI
Lennert De Smet[1], Pedro Zuidberg Dos Martires[1], Robin Manhaeve[1], Giuseppe Marra[1], Angelika Kimmig[1], Luc De Raedt[1,2]
[1]KU Leuven, [2] Örebro Universitet
robin.manhaeve@cs.kuleuven.be
Motivation
Integrating low-level perception with high-level reasoning is one of the oldest, and yet most active open challenges in AI.
Neural-symbolic AI investigates the integration of neural networks and logic
Although great progress is made, many challenges and opportunities remain
Challenges
Scalability
Applications
Opportunities
Increasing the expressivity of the logic
Utilizing the full power of the logic
Multi-predicate learning
Most NeSy tasks are focused on simple tasks
Mostly just standard deep learning tasks with a bit of logic
Logical queries allow for far more complex settings
Multi-task learning
Multiple instance learning
Generative models
Clausal discovery
DeepProbLog [2]
(Probabilistic) facts:
Neural predicate:
Rules:
Query calls
Chess dataset
A dataset that allows us to evaluate the complex reasoning of NeSy systems
Based on a simplified chess setting
Implements a wide variety of tasks, cf. multi-predicate learning
Several representations
Results with DeepProbLog
DeepSeaProbLog [1]
DeepProbLog only supports discete random variables (Boolean and categorical)
DeepSeaProbLog adds support for continuous random variables
Generalizes the concept of neural predicate to neural distributional fact
Reason over continuous quantities
Neural distributional fact
Thresholds over continuous random variables
References
[1] Lennert De Smet, Pedro Zuidberg Dos Martires, Robin Manhaeve, Giuseppe Marra, Angelika Kimmig, and Luc De Readt. Neural probabilistic logic programming in discrete-continuous domains. In Uncertainty in Artificial Intelligence, Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence, UAI 2023, 2023.
[2] Robin Manhaeve, Sebastijan Dumančić, Angelika Kimmig, Thomas Demeester, and Luc De Raedt. Neural probabilistic logic programming in deepproblog. Artificial Intelligence, 298:103504, 2021.
KU Leuven
Department of Computer Science
Celestijnenlaan 200A box 2402
3001 Leuven, Belgium
DTAI
TAILOR

# Advances I Neural-Symbolic AI

Lennert De Smet, Pedro Zuidberg Dos Martires, Robin Manhaeve, Guiseppe Marra, Angelika Kimmig, Luc de Raedt

Contact : robin.manhaeve@cs.kuleuven.be

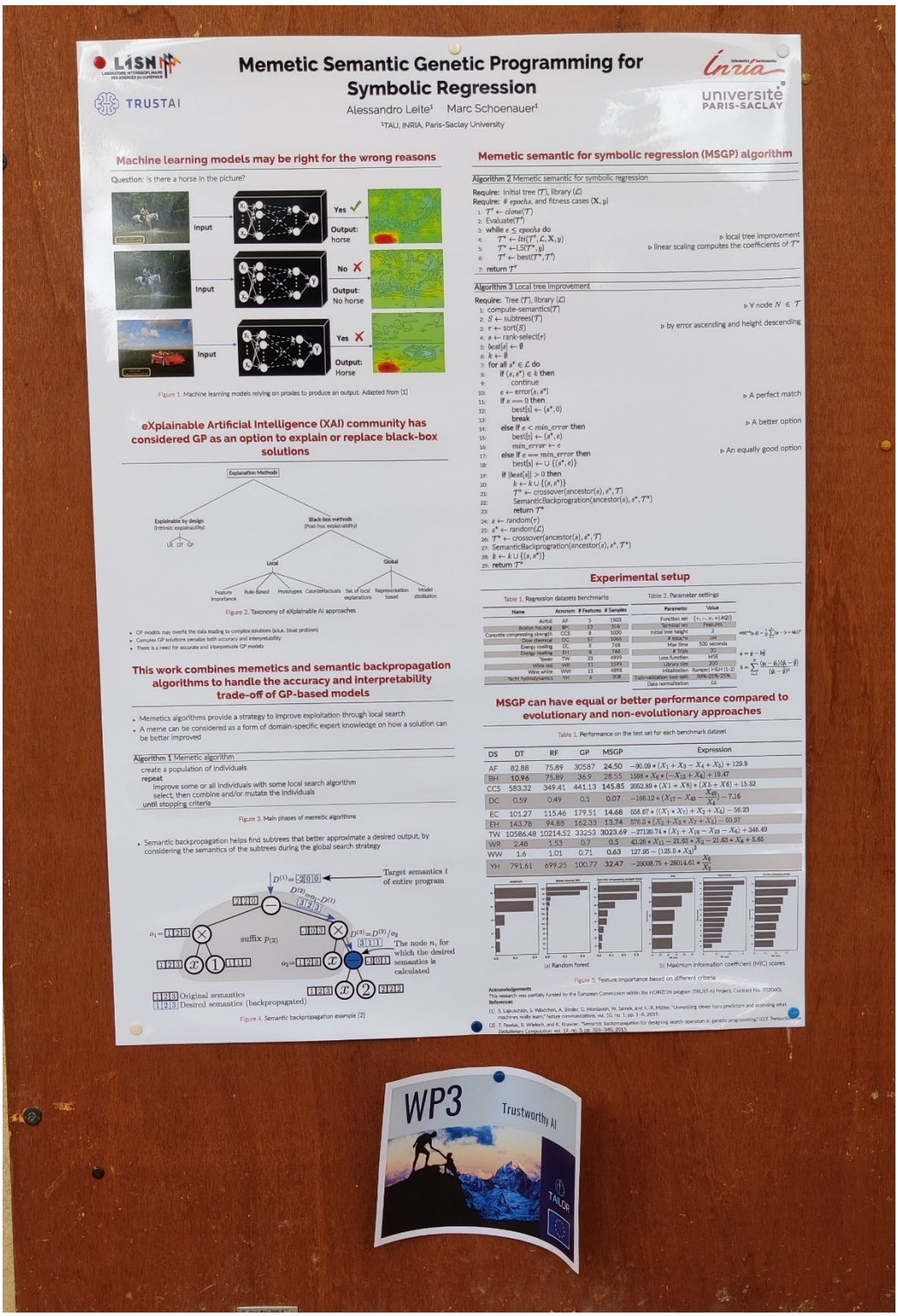
A


P16
3D Tech-responsible Design to Boost AI Sustainability Benefits while Mitigating Risks
Daniela Tulone.
Smart Cities Lab, CINI
Introduction
Emerging opportunities & risks, including long-term rebound effects
Fast-growing work, often fragmented and addressing individual sustainability issues without considering overall impacts and interlinkages
Adverse effects stemming from single-path design approaches are no longer acceptable
Objective
Support early-design of non-critical AI-based solutions to make them environmentally, socially and economically sustainable and aligned with AI ethical principles
Ultimate goal: widely promote the design of sustainable AI systems among small organizations such as start-ups and SMEs
3D tech-responsible scheme
It helps researchers and practitioners beyond the sustainability community
1. Take sustainable design decisions aligned with ethical principles while keeping the big picture
2. Increase system impact by identifying sustainable design options
3. Evaluate potential beneficial and adverse environmental, social and economic sustainability impacts along with its inter-linkages.
3D Sustainable design
A system is sustainable if it meets present societal and environmental needs without compromising the ability of future generations to meet their own needs.
Comprehensive proactive approach
Short and long-term impacts
Alignment with ethical principles
Sustainable system scaling
Pitfalls from single-path approaches
Underestimated scalability effects (e.g., users, data, resources
No one-fit-all solution – need to contextualize
Undervalued implicit assumptions
Business or tech-driven only design
Advantages
Improved performance and robustness
Enhanced energy and resource efficiency
Sustainable scaling
Focused collaborations with stakeholders and domain experts
Better alignment with ongoing regulations and expectations (e.g., AI Act, GDPR, DSA)
Enhanced sustainability impact and higher attractiveness to investors and businesses
Driving principles
Responsible design principles
Anticipative & reflective
Resilient & adaptive to change
Diversity & inclusiveness
Open & transparent
Accountability & governance
AI principles:
Respect for human autonomy
Prevention of harm
Fairness
Explicability & transparency
Ongoing work
The 3D framework aims to be easy-to-use and flexible. It builds on
Dynamic approach to agilely zoom-in on a specific issue and weight different sustainability requirements based on context and system conditions.
Tunable trade-offs to reconcile conflicting sustainability requirements and enhance resource allocation.
It helps overcome fragmentation by integrating existing AI tools and methodologies. Synergies on related work are welcome!
Contact tulone@csail.mit.edu

# #D Tech-responsible Design to Boost AI Sustainability Benefits while Mitigating Risks

**Daniela Tulone**

Contact : tulone@csail.mit.edu


Memetic Semantic Genetic Programming for Symbolic Regression
Alessandro Leite[1] Marc Schoenauer[1]
[1]TAU, INRIA, Paris-Saclay University
Machine learning models may be right for the wrong reasons
Memetic semantic for symbolic regression (MSGP) algorithm
eXplainable Artificial Intelligence (XAI) community has considered GP as an option to explain or replace black-box solutions
This work combines memetics and semantic backpropagation algorithms to handle the accuracy and interpretability trade-off of GP-based models
Experimental setup
MSGP can have equal or better performance compared to evolutionary and non-evolutionary approaches
WP3
Trustworthy AI

# Memetic Semantic Genetic Programming for Symbolic Regression

Alessandro Lelte and Marc Schoenauer

# Time Series Forecasting for Parking Occupancy

José Ángel Morell, Zakaria Abdelmoiz Dahi, Francisco Chicano,
Gabriel Luque and Enrique Alba

ITIS Software, University of Malaga, Spain

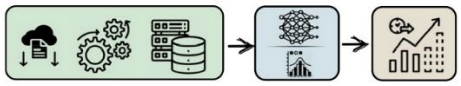
## Motivation

**Problem**: How can AI enhance parking tasks?

**Hypothesis**: Given citizens a step ahead of the parking availability

**How**: Forecasting parking occupancy

## Proposal

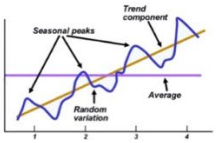


Modules: ETL Forecasting Display

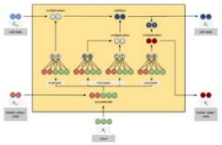
**Data ETL**

- **Extraction:** every 15 minutes (city hall APIs)
- **Transform:** CSV of same structure
- **Normalize:** levels of occupancy to [0,1]

**Forecasting**




- **SARIMAX**
  - S = 48 data points
  - Auto arima
- **LSTM**
  - Epochs = 5000
  - Optimizer = Adam, lr=0.001
  - Batchsize = 32, look_back = 48
- **Prophet**
  - Interval width = 0.80
  - Seasonality = 1 day (48 data points)
  - Change points = auto
- **Neural Prophet**
  - The same as above and
  - Epochs = 5000



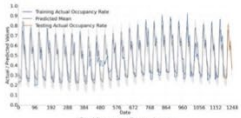
## Experiments and results

**Datasets**

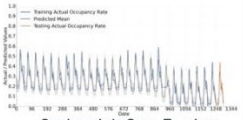
- **Málaga:** 3 parking lots, 6 months of data
- **Birmingham:** 28 parking lots, 2.5 months of data

**Málaga**

| | Per | Technique | Mean | Median | STD |
|---|---|---|---|---|---|
| MAE | Week | Prophet | 0.085 | **0.081** | 0.017 |
| | | Neural Prophet | 0.085 | 0.084 | 0.017 |
| | | LSTM | 0.140 | 0.109 | 0.031 |
| | | SARIMAX | 0.092 | **0.081** | 0.023 |
| | Day | Prophet | 0.080 | 0.078 | 0.006 |
| | | Neural Prophet | 0.075 | **0.074** | 0.002 |
| | | LSTM | 0.083 | 0.082 | 0.020 |
| | | SARIMAX | 0.097 | 0.109 | 0.026 |



Salitre - Sundays



San Juan de la Cruz - Tuesdays

## Conclusions and future work

- Prophet was the best in our datasets
- Predict on-street parking with public cameras
- Optimization of parking search

## Acknowledgements

# Time Series Forecasting for Parking Occupancy

José Angel Morell, Zakaria Abdelmoiz Dahi, Francisco Chicano, Gabriel Luque and Enrique Alba