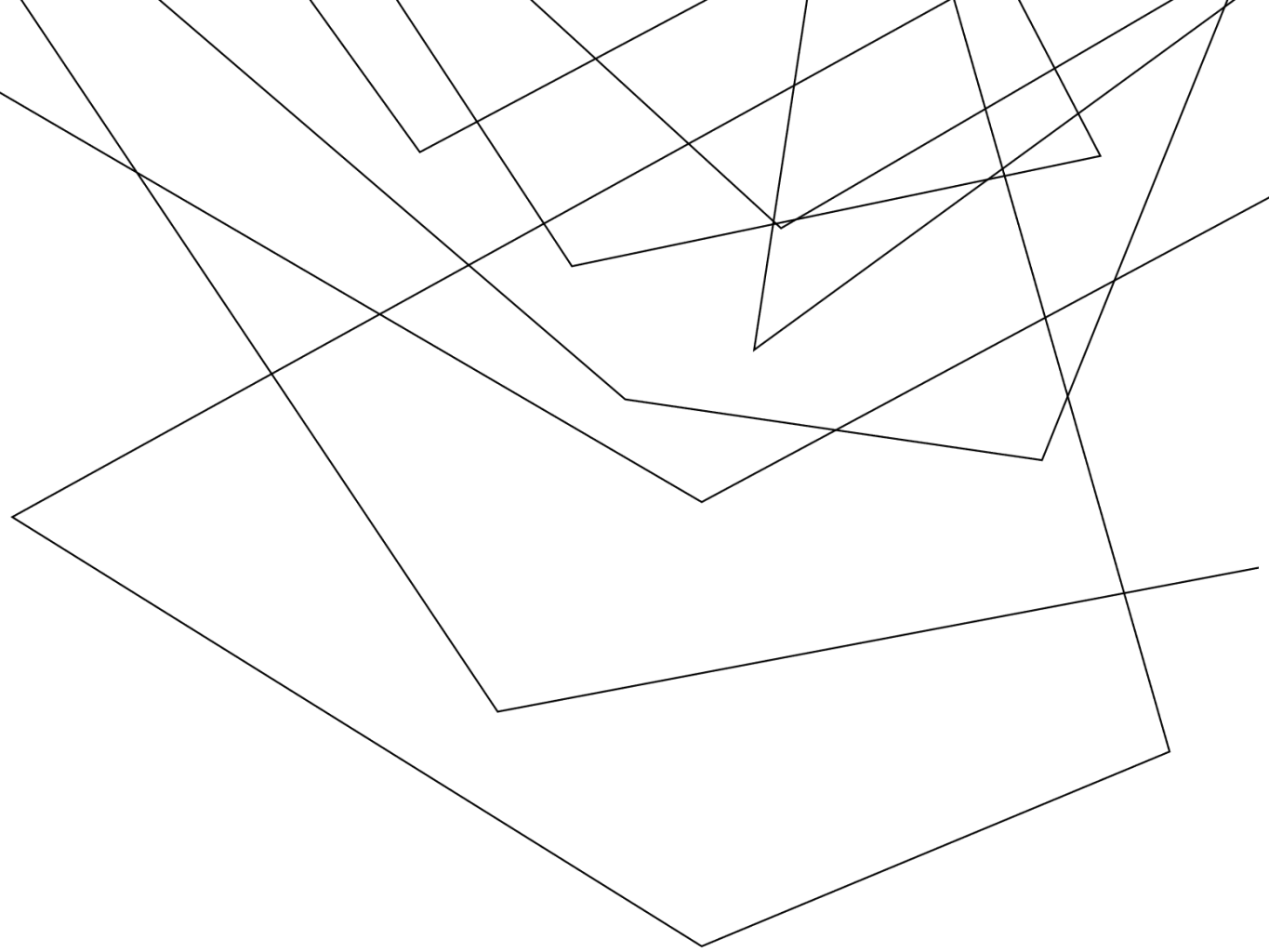
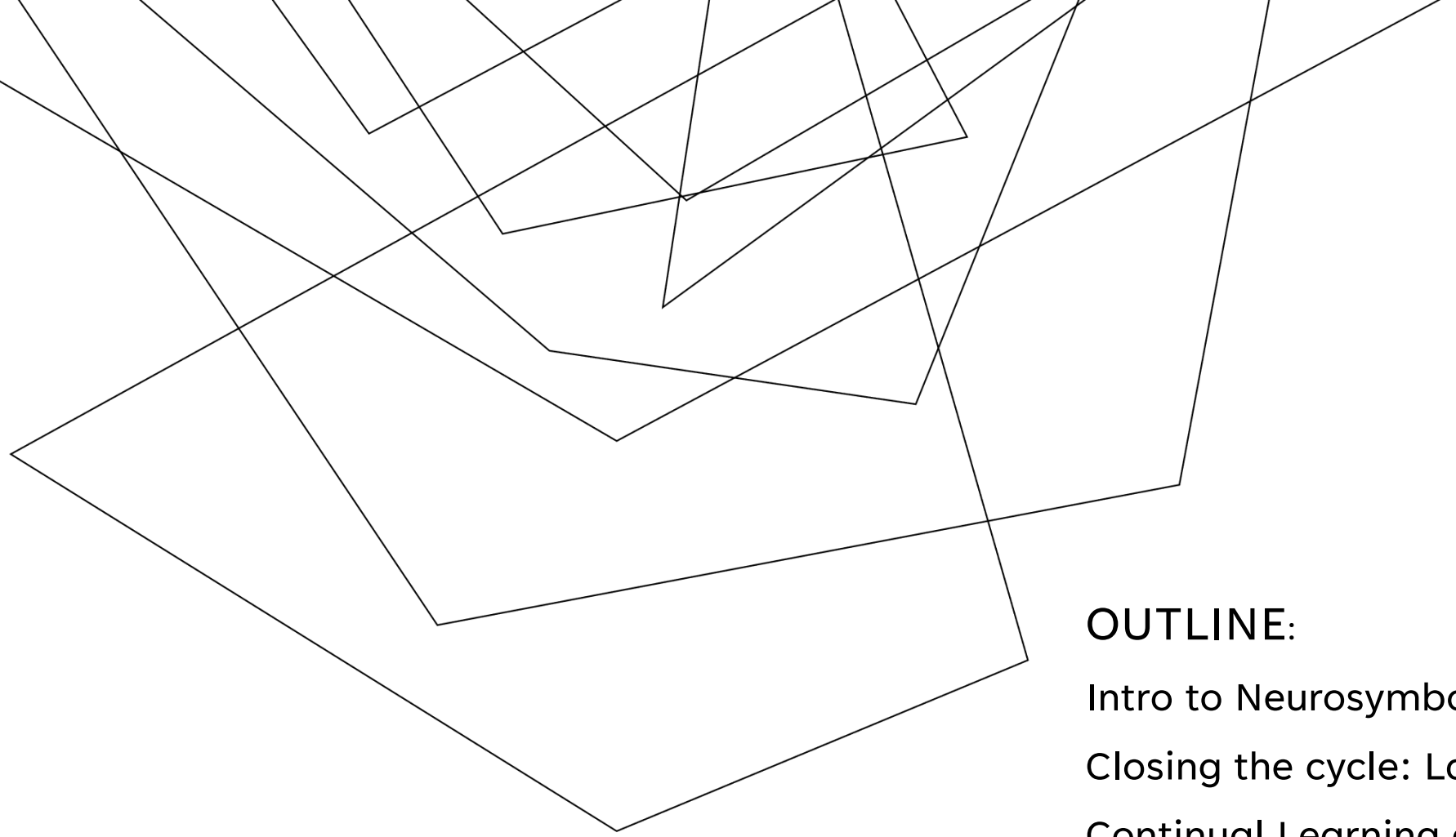


ECML workshop Neuro-symbolic Metalearning and AutoML

18 Sep 2023

NEUROSYMBOLIC AI CONTRIBUTIONS TO METALEARNING

Artur d’Avila Garcez



OUTLINE:

Intro to Neurosymbolic AI cycle

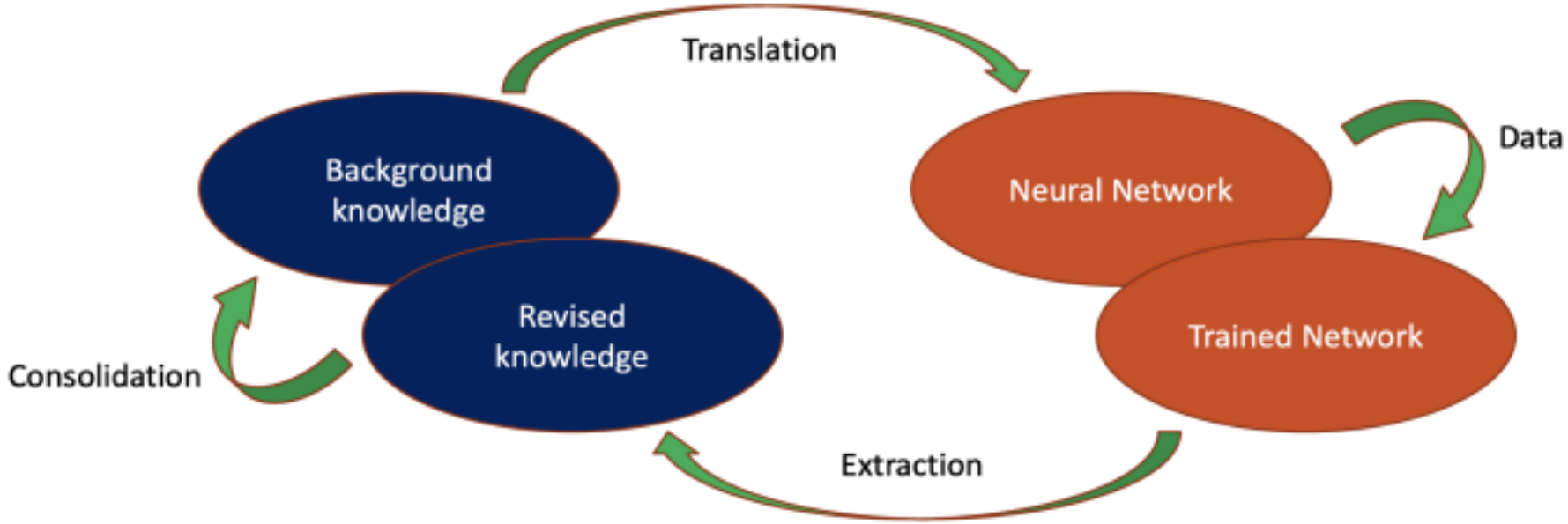
Closing the cycle: Layerwise global XAI

Continual Learning and Reasoning

Coherence and consistency

AI risks and challenges (learning to learn)

Neurosymbolic (NeSy) AI



Bias: Percy et al (2021) Accountability in AI, AI Comm: https://arxiv.org/abs/2110.09232

Fairness: Wagner and Garcez (2021): https://openaccess.city.ac.uk/id/eprint/26151/

Explainability: Ngan et at (2023) Closing the Neural-Symbolic Cycle: Knowledge Extraction, User Intervention and Distillation https://ceur-ws.org/Vol-3432/paper3.pdf

Closing the Neurosymbolic Cycle

Domain expert co-design of AI system; expert can ask what-if questions!

Enlargement of the heart AND obscured diaphragm IMPLY pleural effusion…

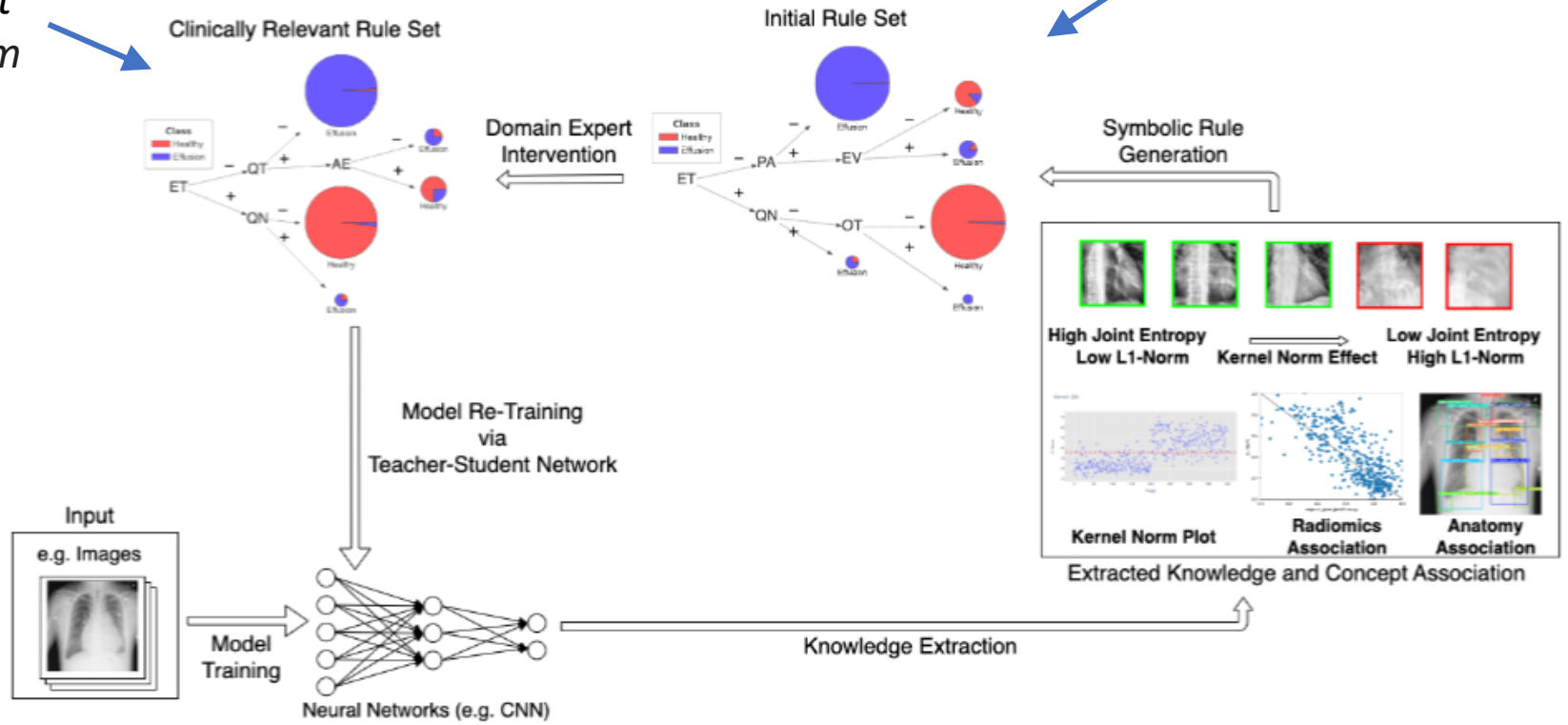
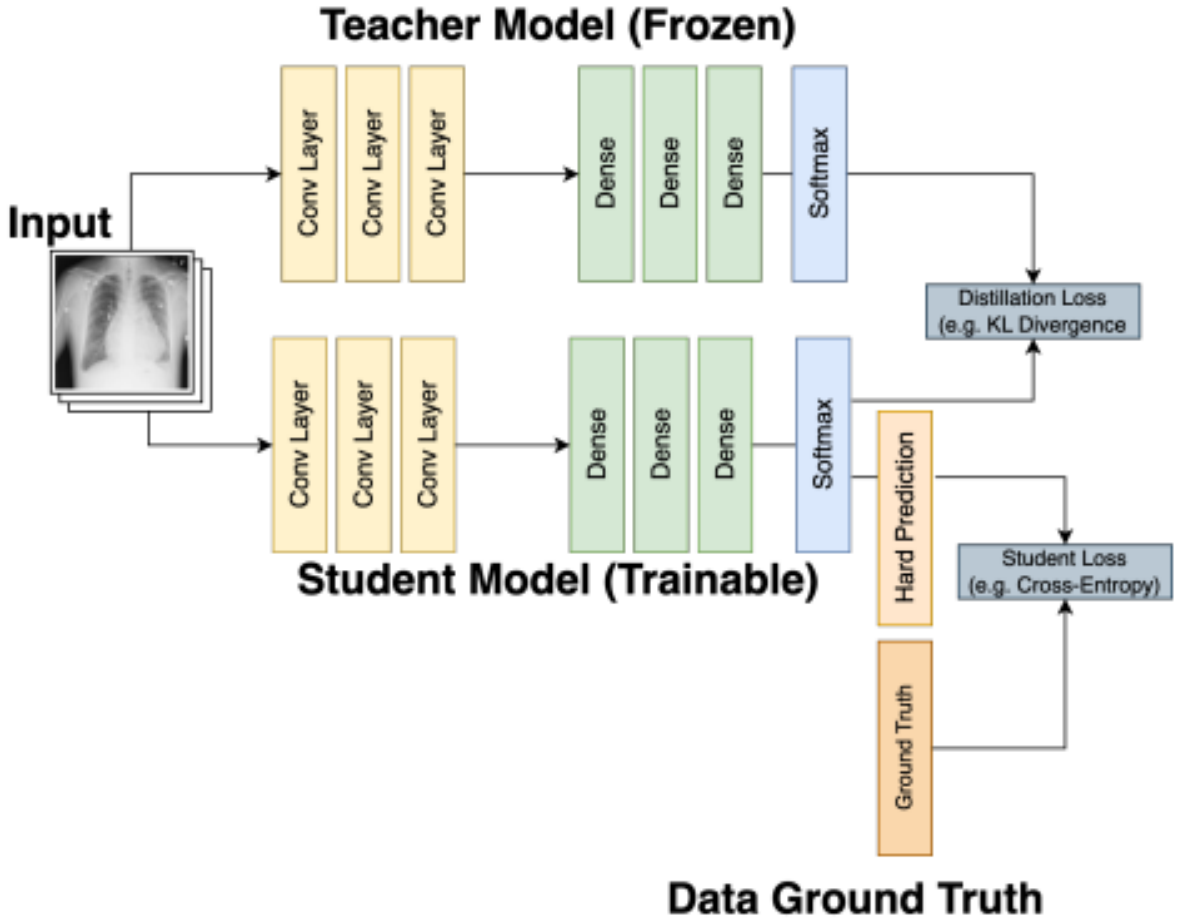


Figure 1: An overview of a neural-symbolic cycle illustrating the process of (a) extracting knowledge from a trained CNN for medical image diagnosis, (b) generating symbolic rules based on the extracted knowledge, (c) expert rule interaction and intervention to produce clinically-relevant knowledge, (d) transferring of relevant knowledge from the rules to a student CNN, closing the neuro-symbolic cycle.

Ngan et at (2023), NeSy2023

NESY CYCLE: EXPLANATION FIRST



Concept learning:

ET = heart enlargement

PA = obscured diaphragm

Concept blending (**multiple teachers**):

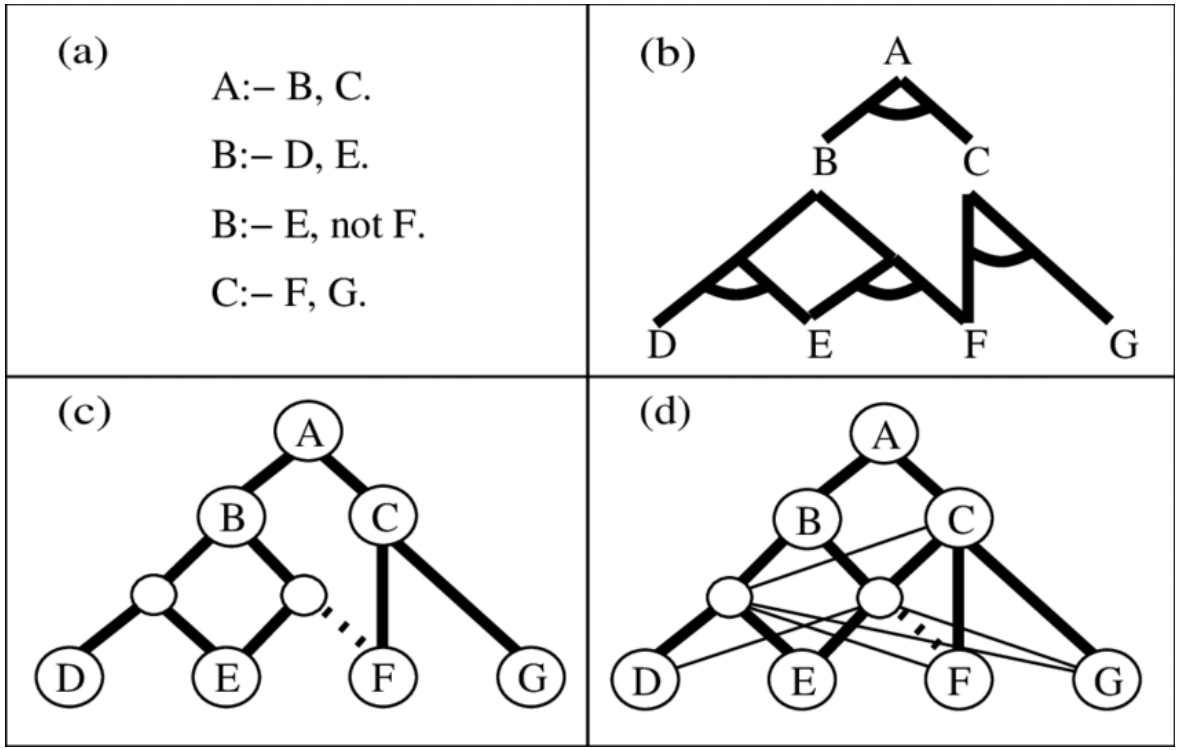
Specialized CNN (image)

Transformer (text)

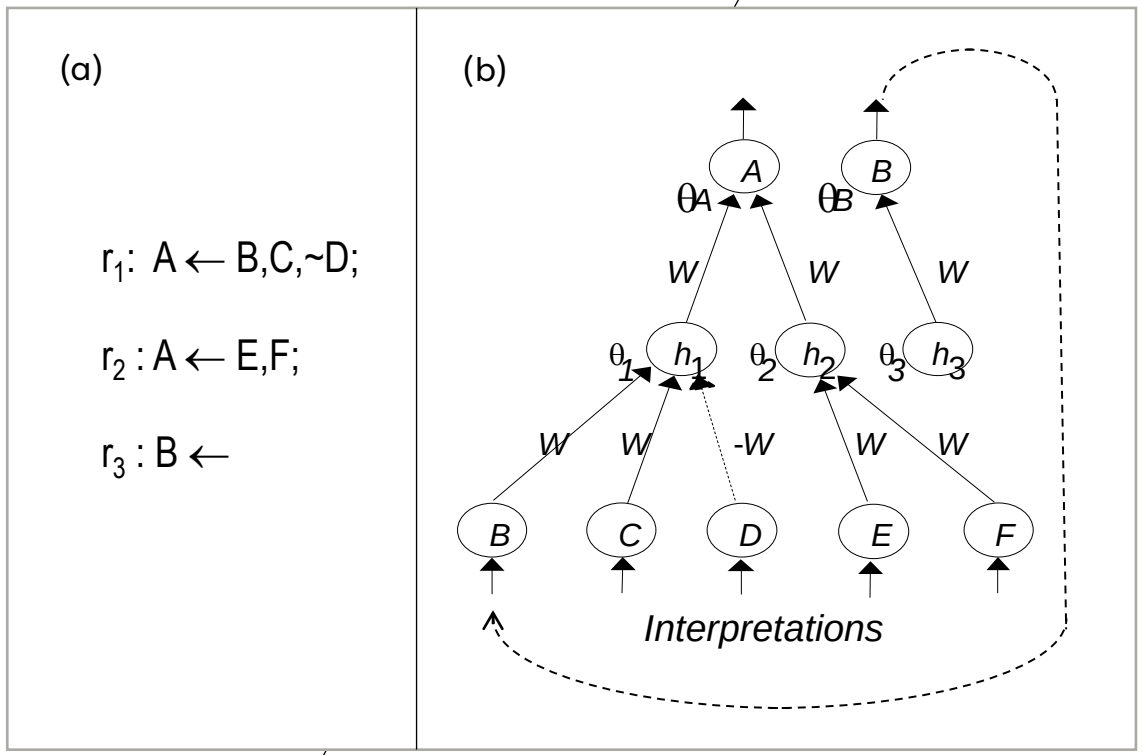
Decision Tree (blood tests)

NESY CYCLE: KNOWLEDGE FIRST

KBANN: LEARNING WITH BACKGROUND KNOWLEDGE



CILP: LEARNING AND REASONING (PROOF OF SOUNDNESS)



SOUNDNESS PRODUCES BETTER LEARNING PERFORMANCE

USE OF NAF MAKES CILP NON-MONOTONIC

A. d'Avila Garcez, G. Zaverucha. The Connectionist Inductive Learning and Logic Programming System. *Applied Intelligence* 11, 59–77 (1999).

CONNECTIONIST MODAL LOGIC

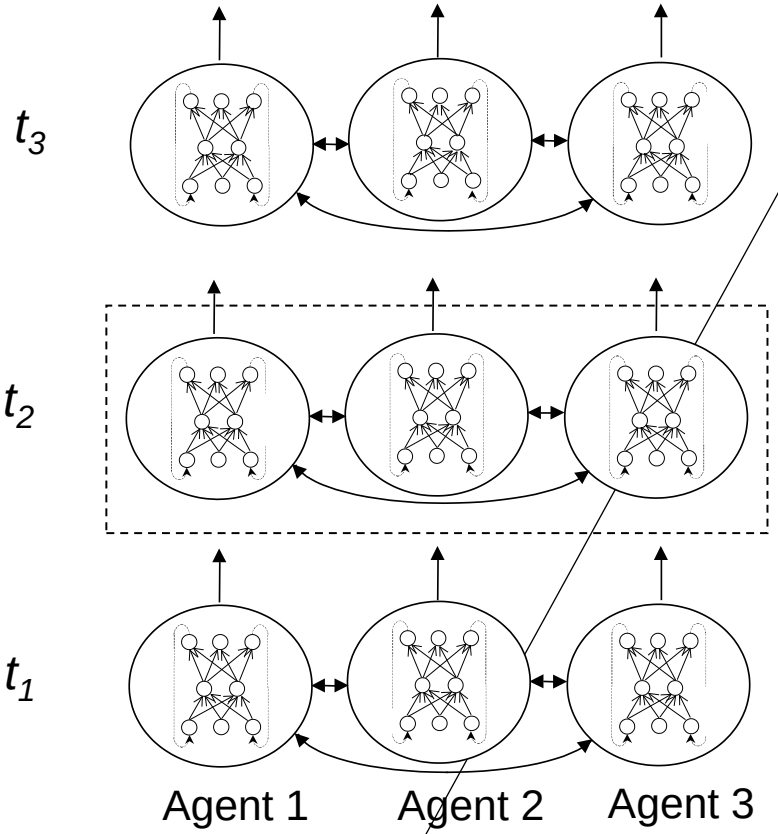
CML = Robustly Decidable (M. Vardi)

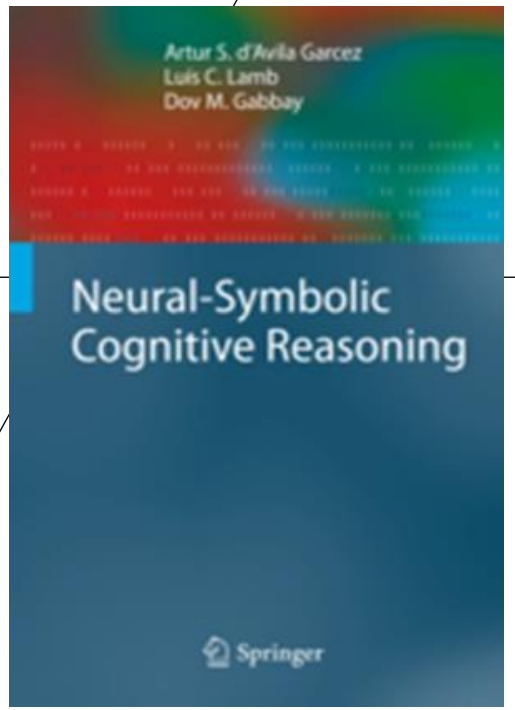
Formal **reasoning** guarantees

How to get to First-order Logic (FOL)?

- Knowledge goes into the Loss Function
- Logic Tensor Networks

SOLUTION TO THE MUDDY CHILDREN PUZZLE:





LOGIC TENSOR NETWORKS

FOL (REAL LOGIC) BY MAPPING TO LOSS FUNCTION

SATISFIABILITY = LEARNING

Reasoning mechanism? We measure **Reasoning Capability** instead

MODULARITY (LTN + XAI)

Proof Theory approaches:

- reduction to logic circuits
- soft-unification
- SAT solving

Example

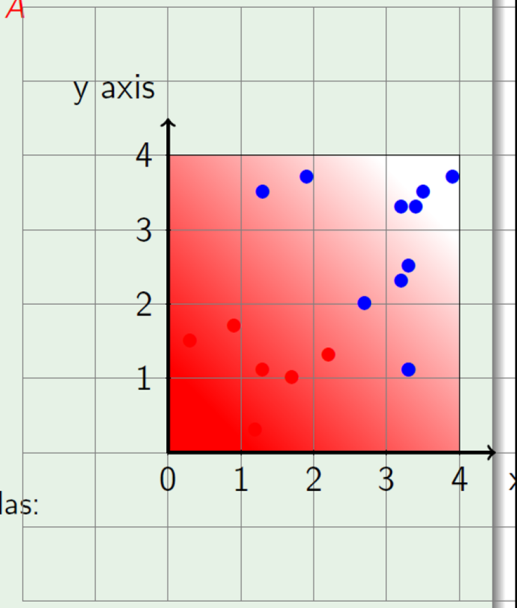
- The domain is the square $[0, 4] \times [0, 4]$;
- We have a set of examples of the class A
- And a set of examples the the class B
- We know that A and B are disjoint
- and let the shape of the membership function of the classes be

$$\sigma(w_1 \cdot x + w_2 \cdot y + w_3)$$

with $\sigma(x)$ the sigmoid function $\frac{1}{1+e^{-x}}$

- We have to find the parameters w_1^A, w_2^A, w_3^A and w_1^B, w_2^B, w_3^B that maximize the satisfiability of the formulas:

$$A(\mathbf{x}) \wedge B(\mathbf{y}) \wedge \forall x : A(x) \rightarrow \neg B(x)$$

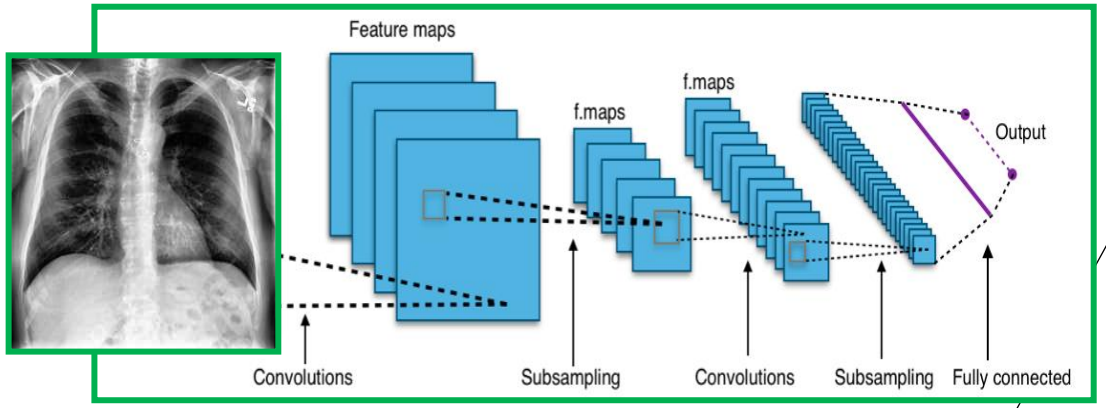


S. Badreddine, A. d'Avila Garcez, L. Serafini, M. Spranger. Logic Tensor Networks, AI Journal 303, Feb 2022

LEVELS OF ABSTRACTION

HARD/SOFT CONSTRAINTS **ON TOP** OF COMPLEX NETWORK

MEDICAL APPLICATION (COLLABORATION WITH FUJITSU RESEARCH):



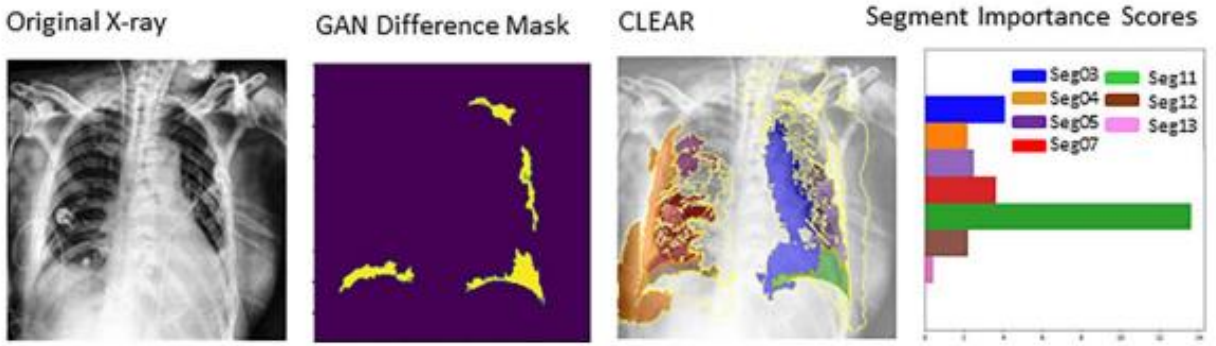
ERIC

LOGIC PROGRAM:

-CL, -PA → pleural_effusion
CL, RW → pleural_effusion
CL, RA, -RW → pleural_effusion
-CL, PA → healthy
CL, -RA, -RW → healthy

K. Ngan, A. d'Avila Garcez, J. Townsend. Extracting Meaningful High-Fidelity Knowledge from Convolutional Neural Networks. IJCNN 2022.

LESSONS FROM XAI



Contrastive Counterfactual Visual Explanations With Overdetermination, MLJ, 2023, https://arxiv.org/pdf/2106.14556.pdf

NETWORKS ARE MUCH MORE COMPLEX NOW (1 TRILLION PARAMENTERS)

GLOBAL XAI IDEAL BUT COMPUTATIONALLY INTRACTABLE

LOCAL XAI CAN HELP BUT LIMITED

MEASURING UTILITY OF EXPLANATION IS EXPENSIVE AND APPLICATION DEPENDENT

MEASURING FIDELITY IS KEY (BOTH OF GLOBAL AND LOCAL XAI METHODS)

PROPOSED SOLUTION: BETTER CONTROL OF MODEL LEARNING (MODULARITY) + LAWERWISE GLOBAL EXPLANATIONS WITH CLEAR FIDELITY MEASURE

BIAS, FAIRNESS AND ALIGNMENT

1. LTN + XAI: Query LTN to obtain SAT levels

(Horses with Stripes are Zebras)

2. Extinct Quagga (not in the data) gets classified as a zebra

3. Add Quagga to the Equidae (horse) family by expressing it symbolically for further training of LTN

4. Repeat

Modularity

Better control

Stopping criteria?

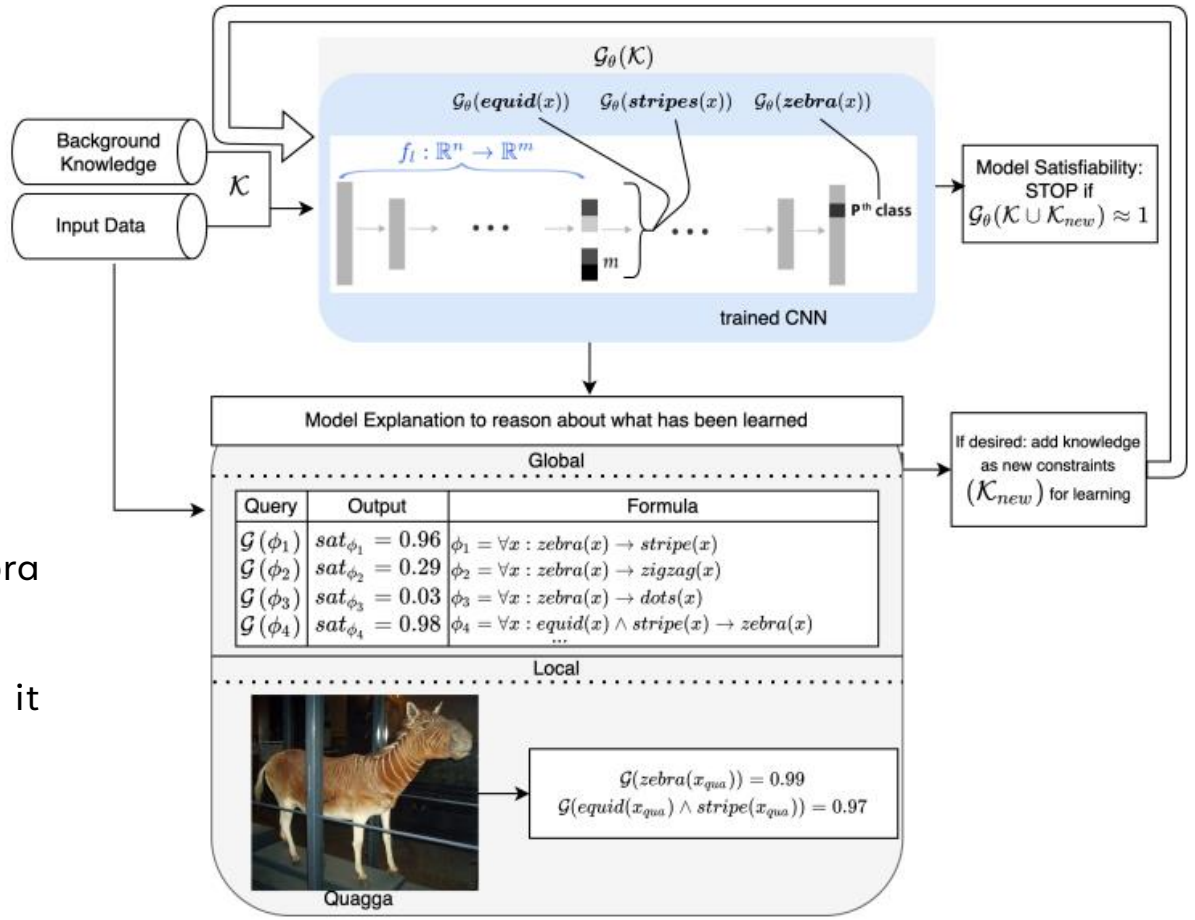
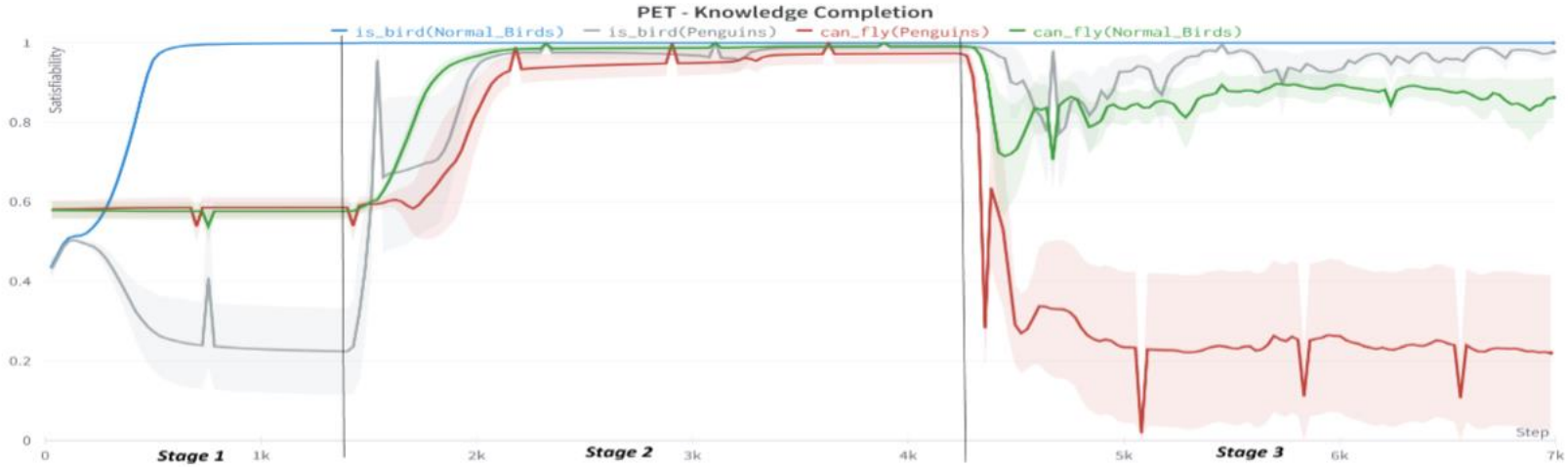


Fig. 3. We produce local explanations (for individual inputs/images) and global explanations (universally-quantified formulas) for the deep learning model by querying it. We then reason about the generality of the explanations given the satisfiability of the queries obtained from the trained network. The figure shows some of the queries associated with groundings in the neural network and their satisfiability (sat) levels. Using linear probes to ground the activation patterns of internal representations into the language of LTN, we are able to utilise abstract concepts as symbols in the logic. Following querying, the neural model can be constrained based on a user selection of logical formulas $\mathcal{K}_{new}$ for further training. This iterative process seeks to align the model with user values in the form of symbolic knowledge $\mathcal{K}$. In the figure, the Quagga is classified as a zebra. A user's desire to change such classification should trigger the addition of knowledge into $\mathcal{K}_{new}$ informed by the queries to be satisfied by the final trained model. Notice that training from data may begin without any knowledge (an empty knowledge-base) which can be revised later by querying user-defined concepts and constraints, such as the fairness constraints from earlier, deemed as necessary for the network to learn.

CONTINUAL REASONING



NeSy 2023 https://arxiv.org/abs/2305.02171

CURRICULUM LEARNING (STAGES 1 TO 3 IN THE GRAPH)

SYMBOLIC KNOWLEDGE ADDED TO LTN TRAINING IN STAGES

QUERYING TO CHECK LTN SAT LEVELS OVER TIME

JUMPING TO CONCLUSIONS:

1. BIRDS FLY IN STAGE 1
2. PENGUINS ARE BIRDS (THEREFORE FLY) IN STAGE 2
3. PENGUINS DON'T FLY IS LEARNED AS EXCEPTION TO THE RULE IN STAGE 3

Continual Reasoning: Non-monotonic Reasoning in Neurosymbolic AI using Continual Learning

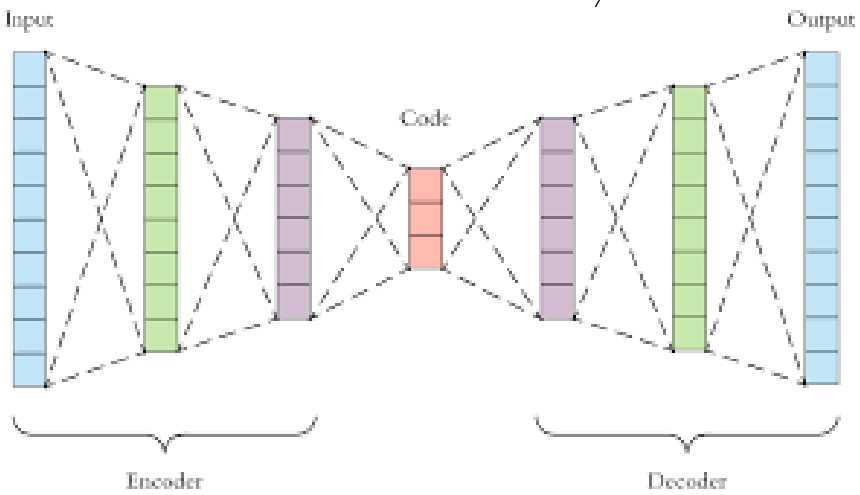
Sofoklis Kyriakopoulos[1,*], Artur S. d'Avila Garcez[1]

[1]*Department of Computer Science, City, University of London, London, UK*

Abstract

Despite the extensive investment and impressive recent progress at reasoning by similarity, deep learning continues to struggle with more complex forms of reasoning such as non-monotonic and commonsense reasoning. Non-monotonicity is a property of non-classical reasoning typically seen in commonsense reasoning, whereby a reasoning system is allowed (differently from classical logic) to *jump to conclusions* which may be retracted later, when new information becomes available. Neural-symbolic systems such as Logic Tensor Networks (LTN) have been shown to be effective at enabling deep neural networks to achieve reasoning capabilities. In this paper, we show that by combining a neural-symbolic system with methods from continual learning, LTN can obtain a higher level of accuracy when addressing non-monotonic reasoning tasks. Continual learning is added to LTNs by adopting a curriculum of learning from knowledge and data with recall. We call this process *Continual Reasoning*, a new methodology for the application of neural-symbolic systems to reasoning tasks. Continual Reasoning is applied to a prototypical non-monotonic reasoning problem as well as other reasoning examples. Experimentation is conducted to compare and analyze the effects that different curriculum choices may have on overall learning and reasoning results. Results indicate significant improvement on the prototypical non-monotonic reasoning problem and a promising outlook for the proposed approach on statistical relational learning examples.

CONSISTENCY AND COHERENCE PRINCIPLES

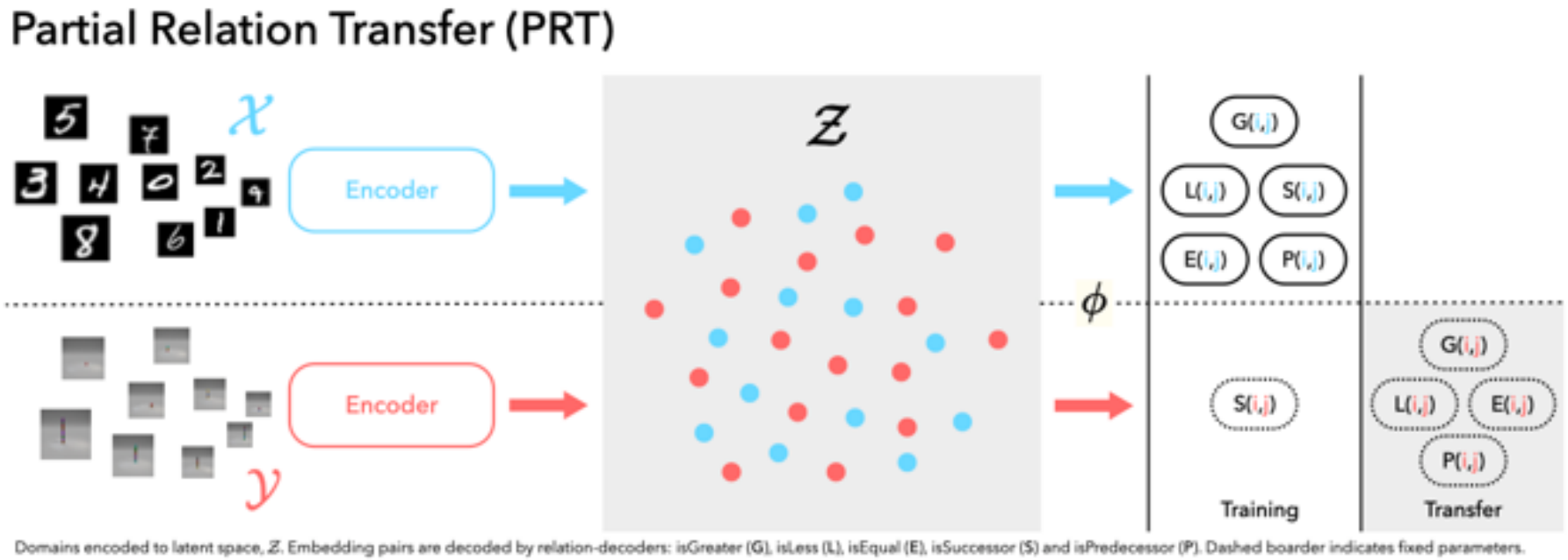


LEARNING CONCEPT THAT A BLOCK STACK IS LARGER THAN ANOTHER

TRANSFERRING TO ABSTRACT DOMAIN e.g. 5 IS LARGER THAN 4

REASONING e.g. TRANSITIVITY

CONSISTENCY ACROSS DOMAINS

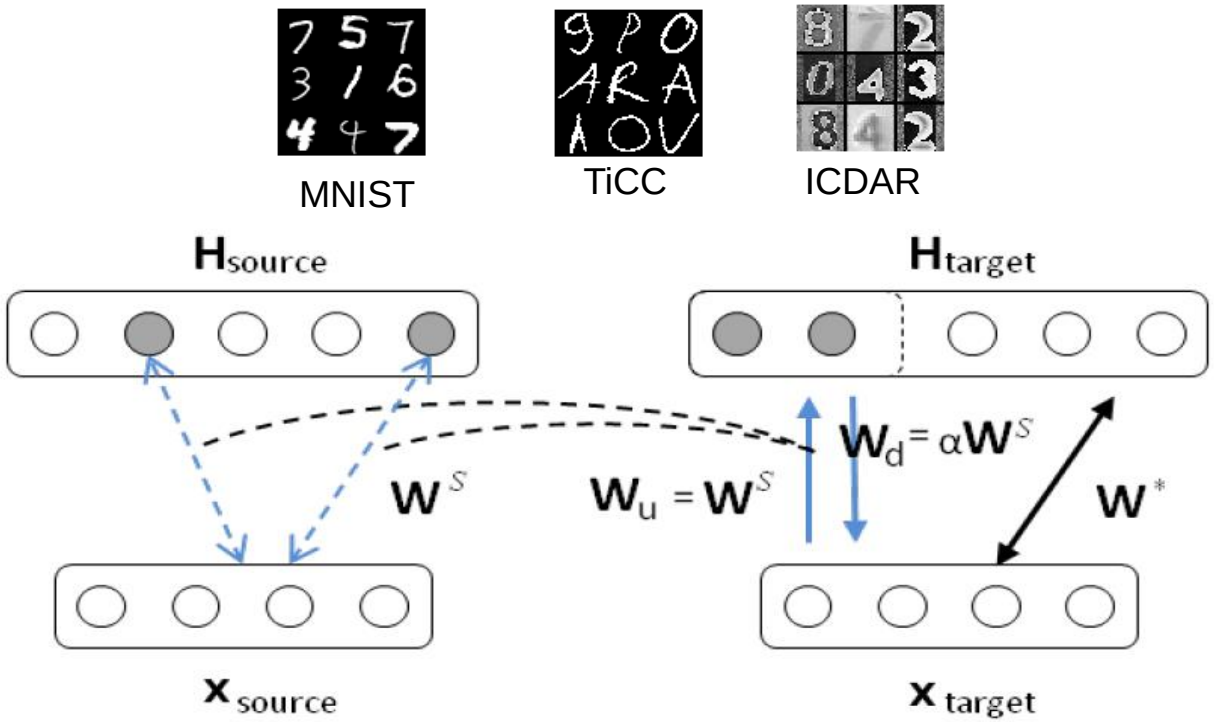


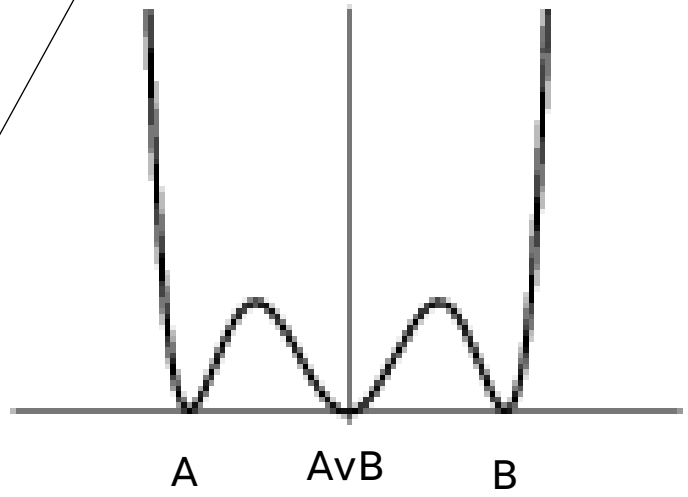
H. Strömfelt, L. Dickens, A. d'Avila Garcez, A. Russo. Formalizing Consistency and Coherence of Representation Learning. NeurIPS 2022

BEYOND BACKPROP

RBM TRAINED WITH CONTRASTIVE DIVERGENCE REPRESENT PROPOSITIONAL LOGIC

TRANSFER LEARNING APPLICATION: PLEURAL EFFUSION TO COVID, DIGITS TO CHARACTERS…





S. Tran, A. d'Avila Garcez. Neurosymbolic Reasoning and Learning with Restricted Boltzmann Machines, AAAI 2023.

CHALLENGES AND NEXT STEPS:

https://sites.google.com/view/nesy2023/home
(see *Danny Silver* and *Tom Mitchell's paper*)

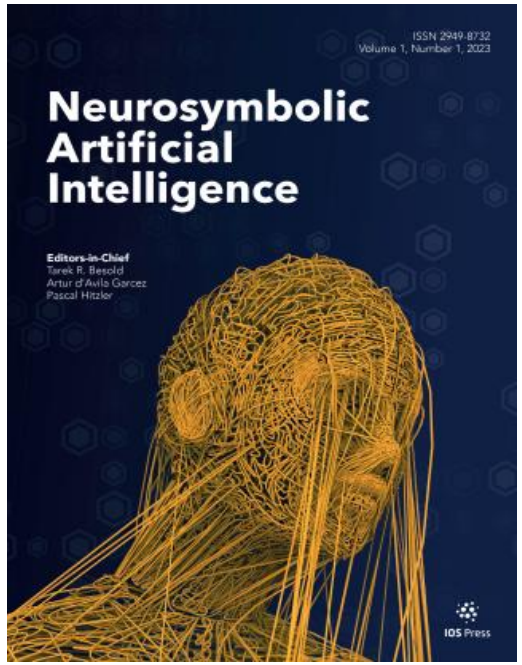
Many new approaches/hybrid systems with different forms of representation / loose formalization of reasoning

Clear definitions of reasoning needed (fixed-point computation/resolution, etc) / soundness results

Need for benchmarks on multimodal coherent learning and reasoning (e.g. consistency of image and text when moving from objects to abstract concepts)

Applications: healthcare (NSI cycle applied at scale), AI accountability, finance?

NeSy killer app: Chatbot with fact checking (not RLHF)



New journal with IOS Press

The (existential) Risk of AI LLM

https://github.com/turing-knowledge-graphs/meet-ups/blob/main/symposium-2023

AI debate 3 (Dec 2022): the main present danger is disinformation at scale, not autonomous weapons (https://www.youtube.com/watch?v=JGiLz_Jx9uI)

23 May 2023: false explosion near the Pentagon caused real dip in the stock market (posted on fake "verified" Bloomberg account)

- Disinformation at scale with risk to democracy
  (ChatGPT deployed to 100 million users worldwide)
- Polluting the internet (with AI-generated content)
- Will GPT4 try to deceive us?
  - AIGC can be wrong but convincing
  - Fact-checking LLMs needed at the right level of abstraction (not easy)
  - Facts exist in logic: instil, distil, repeat

LeCun: open source is the solution!

Hinton: how do you feel about open sourcing nuclear power research?

Bengio signed new statement from The Centre for AI Safety: https://www.safe.ai/statement-on-ai-risk: “Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.”

June 2023: Hinton and Ng agree that they need to agree on what to agree…

LeCun: Meta is using AI to mitigate risk (number of accounts taken down every quarter goes from 400 million to 1.5 billion accounts; worldwide 100+ languages). The bottleneck isn't content production, but social media platforms (and their systems for moderation of content) and being able to capture the public's attention...

The Race is On

- Biden/Sunak meeting (8 June 2023): UK looks forward to hosting the first global summit on AI safety in Nov 2023:

  https://www.gov.uk/government/news/uk-to-host-first-global-summit-on-artificial-intelligence
  (with DeepMind, Anthropic, Palantir, Microsoft, Faculty.ai)
- No involvement of academia or civil society!?

- In the meantime… DeepMind's **Nature** paper (7 June 2023): "Faster sorting algorithms discovered using deep reinforcement learning" https://www.nature.com/articles/s41586-023-06004-9
- **Not a new algorithm, instead a better compiler optimization**
- The hype associated with the AI race is bad for AI…

- **UK government sets out AI Safety Summit ambitions (4 Sep 2023)**
  https://www.gov.uk/government/news/uk-government-sets-out-ai-safety-summit-ambitions

- **The Next Global Superpower Isn't Who You Think | Ian Bremmer | TED**
  https://www.youtube.com/watch?v=uiUPD-z9DTg

A CERN for AI

OPEN SOURCE LARGE-SCALE AI RESEARCH AND ITS SAFETY

Calls on governments of US, UK, EU, AU and Canada to establish an international, open-source supercomputing research facility for AI with oversight by elected institutions.

Petition at:

https://www.openpetition.eu/petition/online/securing-our-digital-future-a-cern-for-open-source-large-scale-ai-research-and-its-safety

SUMMARY

Use symbolic knowledge to control learning (architecture and loss function), add fairness constraints and perform alignment

Use neurosymbolic cycle with human-in-the-loop to promote curriculum learning and better concept learning

Use continual learning and reasoning to control reasoning

Use coherence principles to design learning models with better transfer learning performance

All of the above require a human-in-the-loop approach (not RLHF though)

To Do:

- Evaluation on multi-task benchmarks
- Learning to learn that is fully-automated

Thank you!

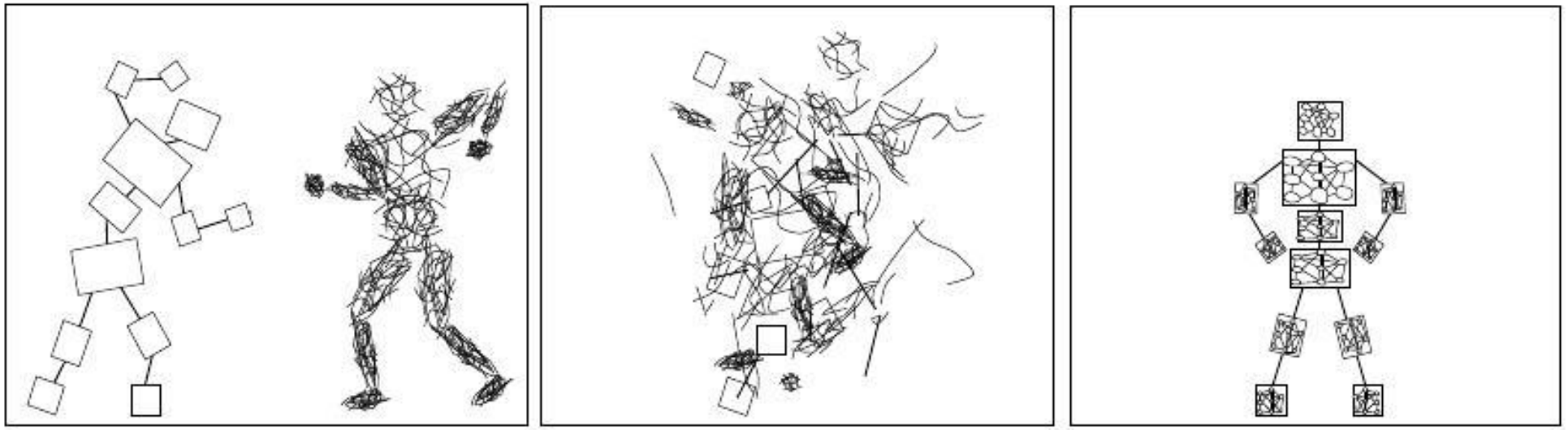


Figure 1. Conflict between theoretical extremes.