Playground Global
26 March 2019

Neural-Symbolic Computing

Artur d'Avila Garcez
City, University of London
a.garcez@city.ac.uk

The AI revolution...

The promise of AI:

- Education (adaptive learning)
- Finance (time series prediction)
- Security (image and speech recognition)
- Health (monitoring - IoT, drug design)
- Telecom (infrastructure data analysis)
- Games (interactive learning)
- Transport (logistics optimization)
- etc, etc.

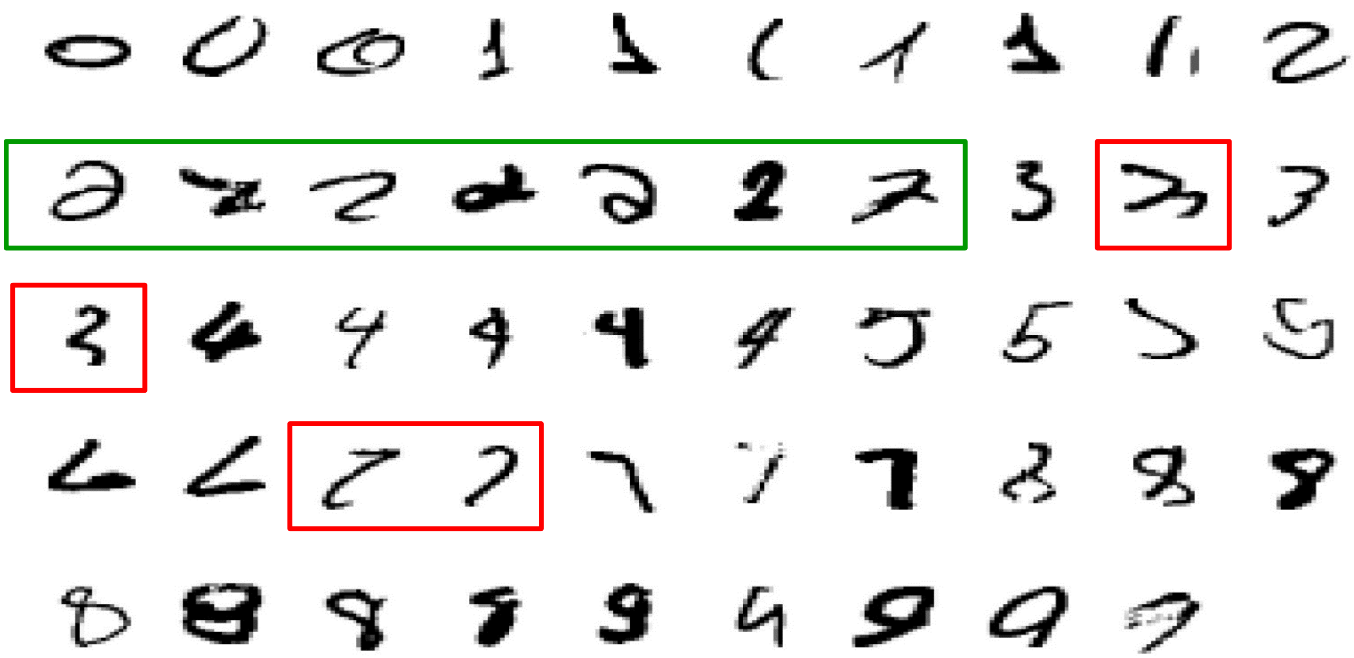
Brain/Mind dichotomy

- Symbolic AI: a symbol system has all that is needed for general intelligence
- Sub-symbolic AI: intelligence emerges from the brain (neural networks)
- What neural nets can and can't do? Association, reasoning (causal inference), planning...

Machine Learning (ML)

Systems that improve their own performance from experience

ML good for: stuff that is hard to specify but with lots of examples



ML (continued)

Systems that, in addition, enable humans to improve their performance

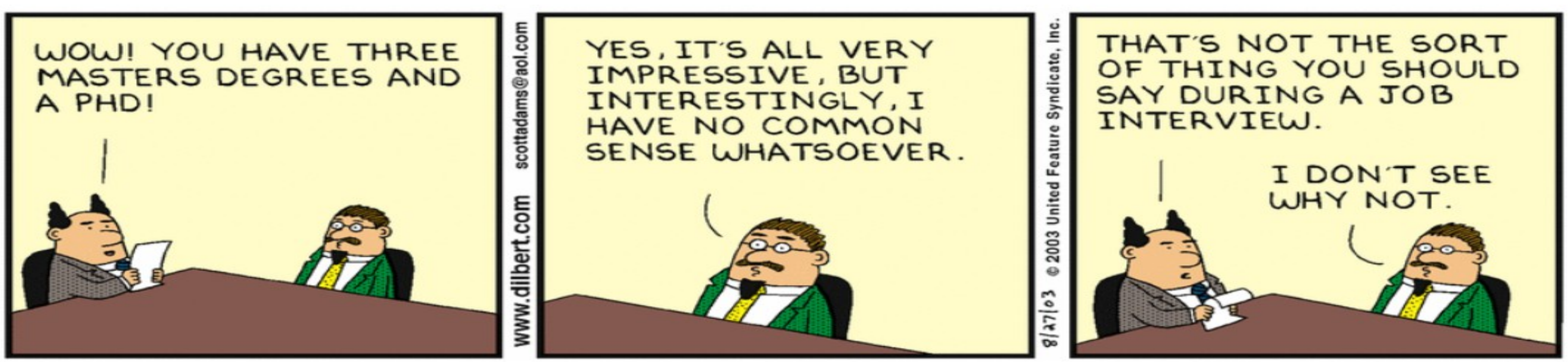
Did AlphaGo help players improve their performance?

Explainable AI: trust, performance improvement and transfer learning

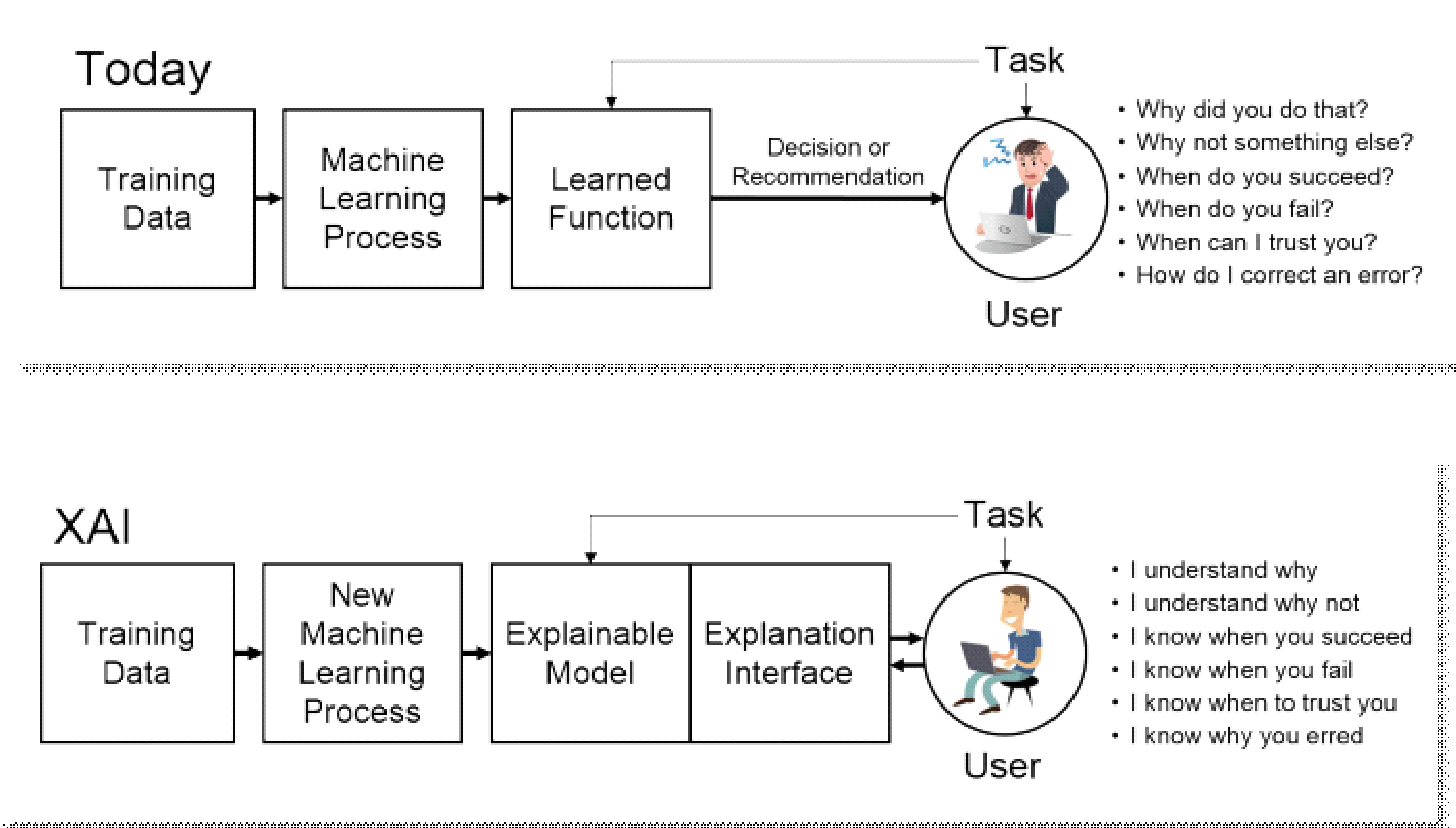
Deep Learning

Given a lot of data (fact of life), "black box" deep neural networks tend to work much better than symbolic ML (decision trees, linear regression) or expert systems...

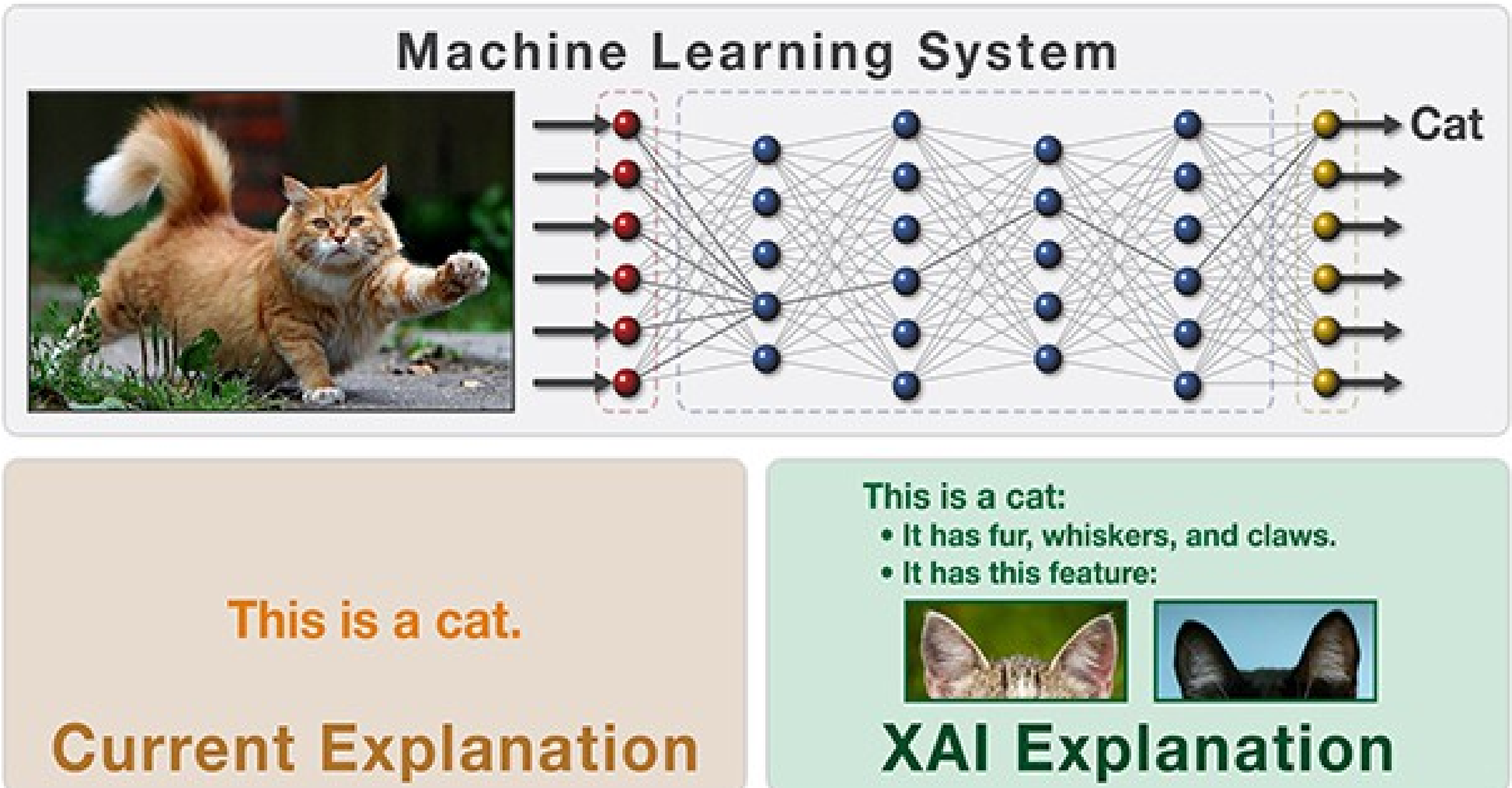
Billions of parameters trained from scratch every time (no prior knowledge); where is commonsense?



DARPA's XAI programme



Explainable AI or Interpretable ML?



XAI = Interpretable ML (local)

Explanation = knowledge extraction (global)

Forms of explaination

- Local vs Global (explain instance or model?)
- Is there need for a measure of fidelity (safety)?
- System maintenance, improvement or transfer learning?
- Ethical (bias in the data, c.f. The Digital Poorhouse
- Regulatory (GDPR's right to an explanation)
- How is it best communicated? Levels of abstraction (drill down)

Types of explanation

Domain specific:

medical diagnosis (high accuracy) vs credit application vs self-driving car (moral judgment)

- No explanation desirable (privacy, fairness)
- No explanation needed (but desirable)
- Explain how decision was made (normative)
- Explain why decision was made (reasoning)
- Explain what would need to change for outcome to be different (what if...)
- Safety-critical systems (fidelity measures, verification/assurance cases needed)

Outcomes of explanation

- You don't believe the decision was right and you want to find out (trust, process)
- You want to know how to get a different outcome the next time (e.g. play the system)
- You want the system or your understanding of it to improve (ML engineer)
- A combination of the above or another reason you may not even be aware of...

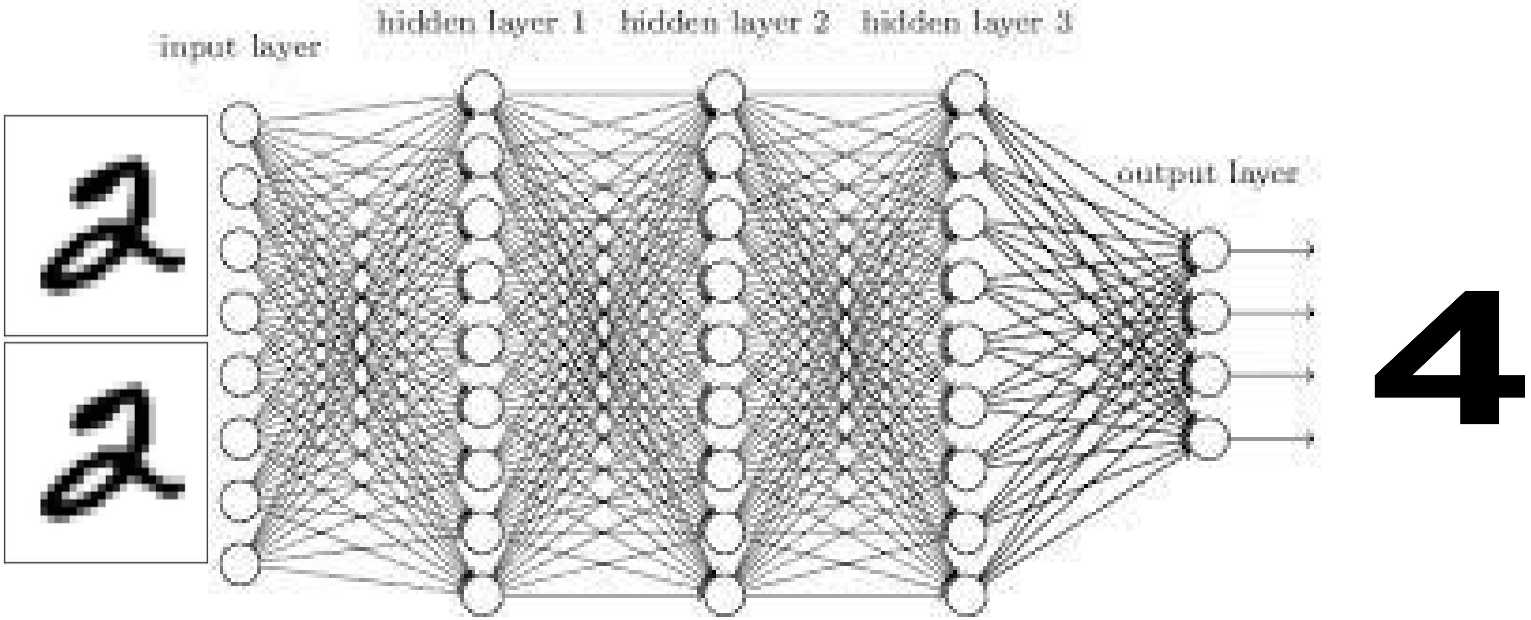
The need for knowledge extraction

- “We need a language for describing the alternative algorithms that a network of neurons may be implementing” L. Valiant

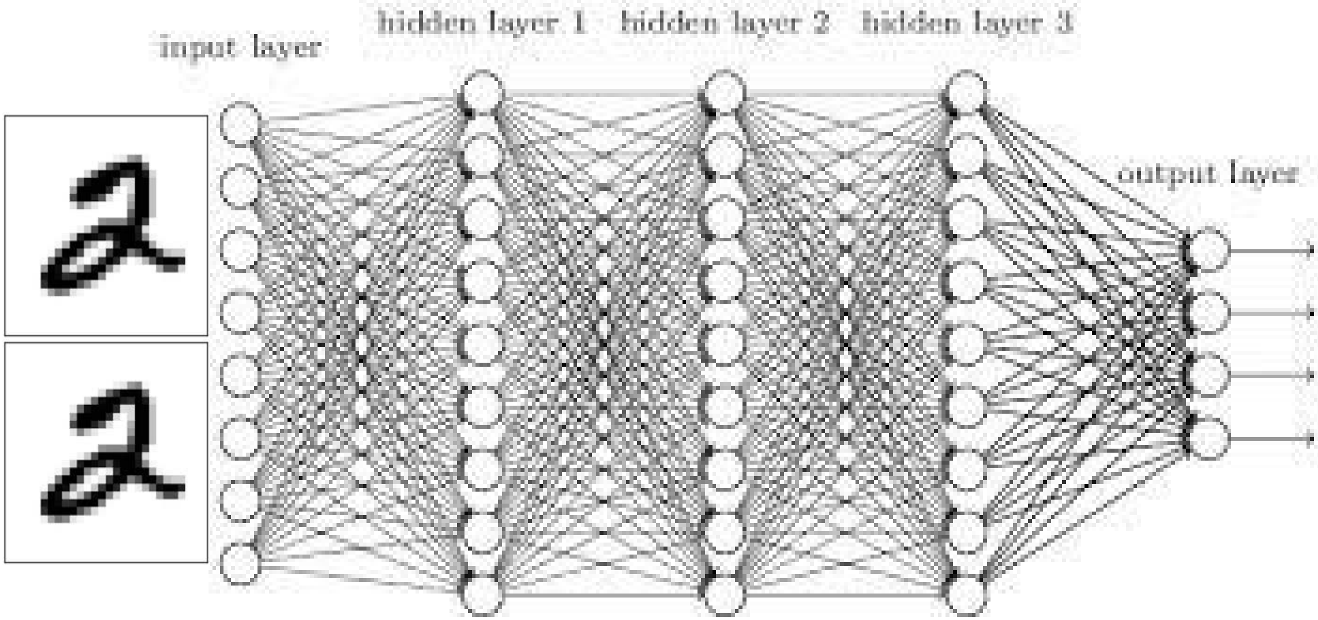
(New) Logic + Neural Computation

Goal of Neural-Symbolic Computing: Learning from experience and reasoning about what has been learned in an uncertain environment in a computationally efficient way

Knowledge Extraction from Deep Nets



Knowledge Extraction from Deep Nets



4

2+2=4

Neural-Symbolic Methodology

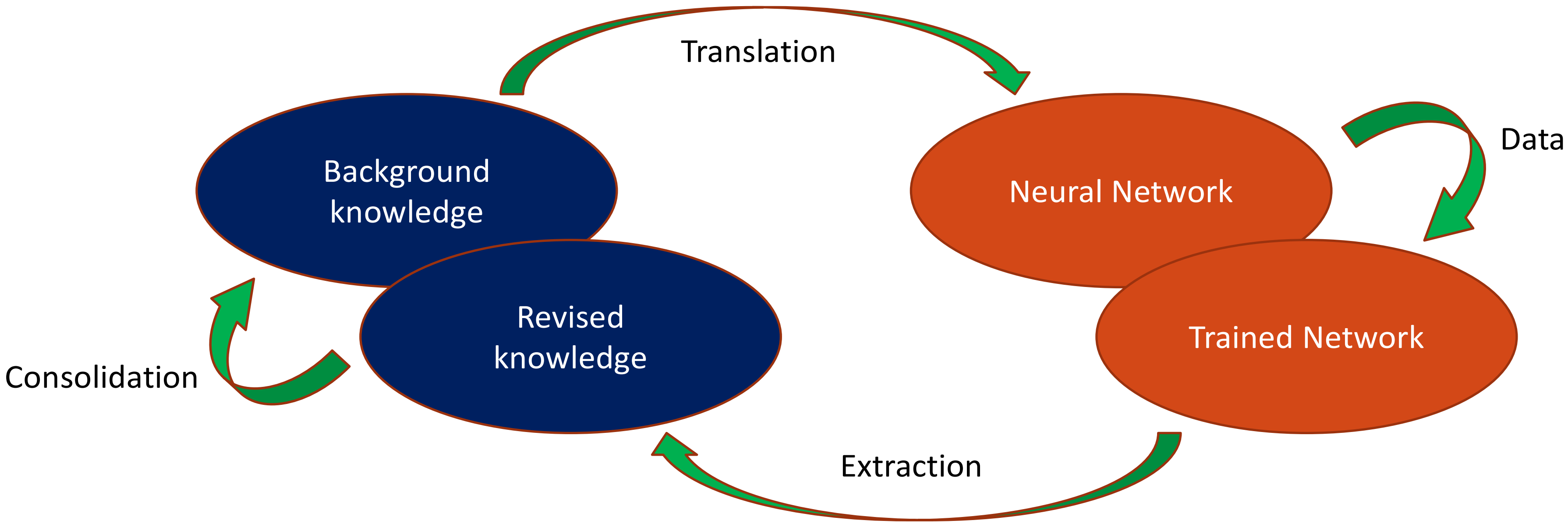
high-level symbolic representations
(abstraction, recursion, relations, modalities)

translations

low level, efficient neural structures
(with the same, simple architecture throughout)

Analogy: low-level implementation (machine code) of high-level representations (e.g. java, requirement specs)

Neural-Symbolic Learning Cycle



Connectionist Inductive Logic Programming (CILP System)

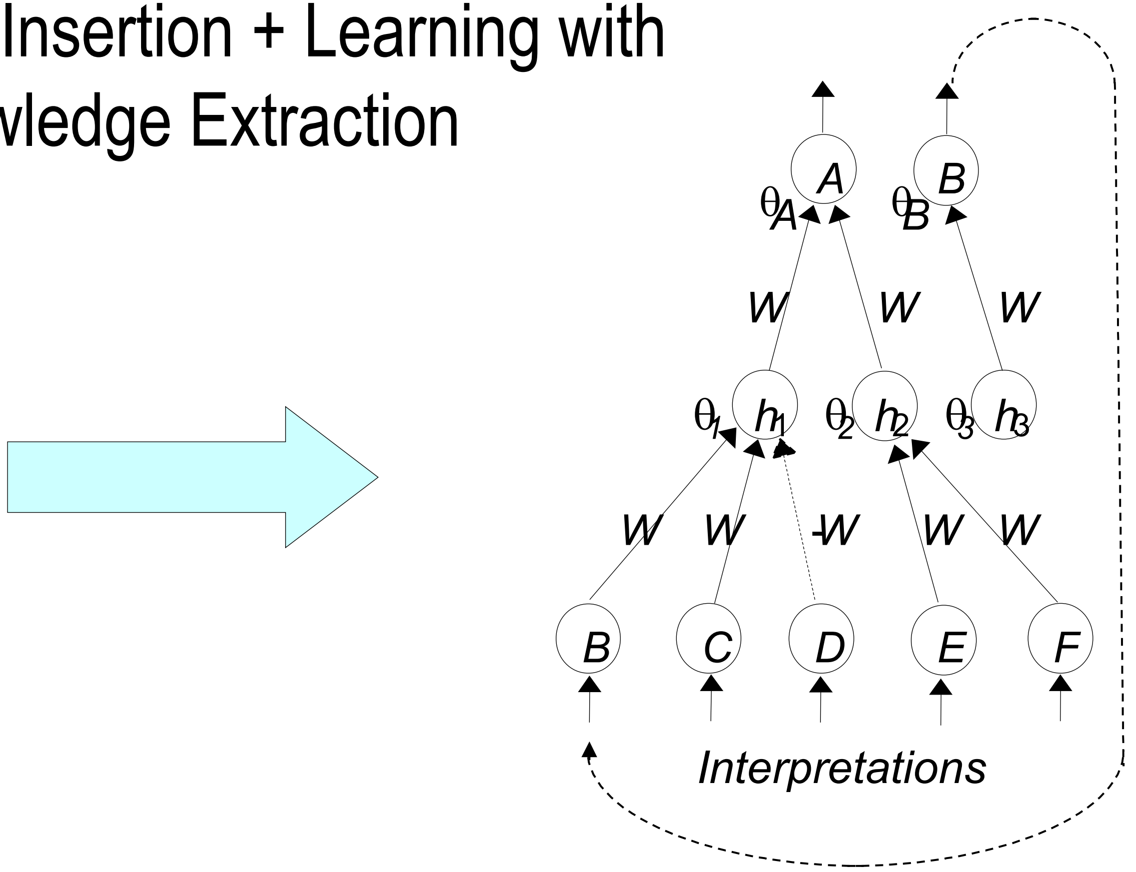
A Neural-Symbolic System for Integrated Reasoning and Learning (neural nets + logic programming)

Background Knowledge Insertion + Learning with Backpropagation + Knowledge Extraction

r_1: $A \leftarrow B, C, \sim D;$

r_2 : $A \leftarrow E, F;$

r_3 : $B \leftarrow$



CILP Rule Extraction

- Knowledge is extracted by querying/sampling the trained network;
- A partial ordering helps guide the search, reducing complexity on the average case;
- A proof of soundness guarantees that the rules approximate the behaviour of the network;
- Rule simplification and visualization in the form of a state transition diagram help experts validate the rules

Example: Pump System

A pump system controls the levels of water in a mine to avoid the risk of overflow; an initial, partial system description is available.

State variables: *CrMeth* (level of methane is critical)
HiWat (level of water is high)
PumpOn (pump is turned on)

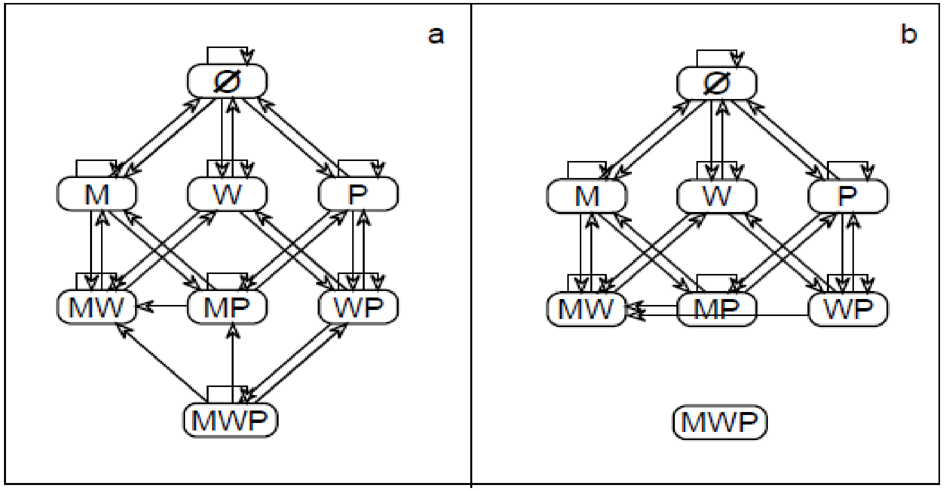
Safety property in LTL: $G\neg(CrMeth \wedge HiWat \wedge PumpOn)$

Partial system spec (background knowledge; s = sensor):

$$\bigcirc CrMeth \leftarrow sCMOn.$$
$$\bigcirc CrMeth \leftarrow CrMeth, \sim sCMOff.$$
$$\bigcirc HiWat \leftarrow sHiW.$$
$$\bigcirc HiWat \leftarrow CrMeth, \sim sLoW.$$
$$\bigcirc PumpOn \leftarrow TurnPOn.$$
$$\bigcirc PumpOn \leftarrow CrMeth, \sim TurnPOff.$$

Example: extraction of state transition

CrMeth = M (level of methane is critical)
HiWat = W (level of water is high)
PumpOn = P (pump is turned on)



Extracting Automata from Recurrent Neural Networks Using Queries and Counterexamples, Gail Weiss, Yoav Goldberg, Eran Yahav, 2017

Learning and Representing Temporal Knowledge in Recurrent Networks, Rafael V. Borges, Artur d'Avila Garcez, Luis C. Lamb, IEEE TNNLS, 2011

S. Tran and A. S. d'Avila Garcez. Deep Logic Networks: Inserting and Extracting Knowledge from Deep Belief Networks. IEEE TNNLS, 2016

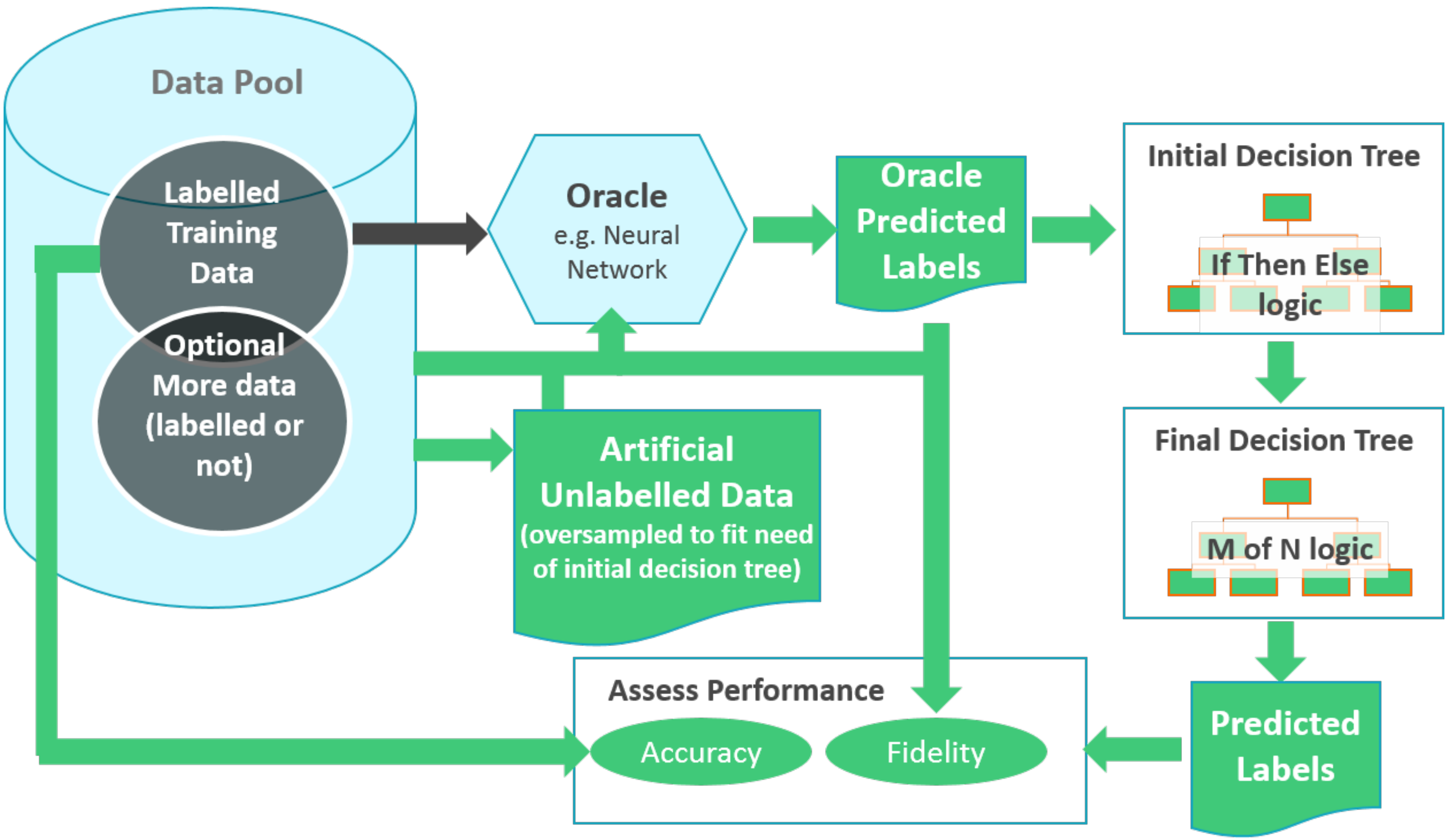
TREPAN (Craven and Shavlik, NIPS 1995)

- Unsound but efficient; extracts decision trees from trained neural networks
- Treats neural net as black-box (oracle) from which to query for input/output patterns
- Samples data from the training set or synthetic data to generate examples for the decision tree training
- Simplifies the rules in the trained decision tree into MofN rules

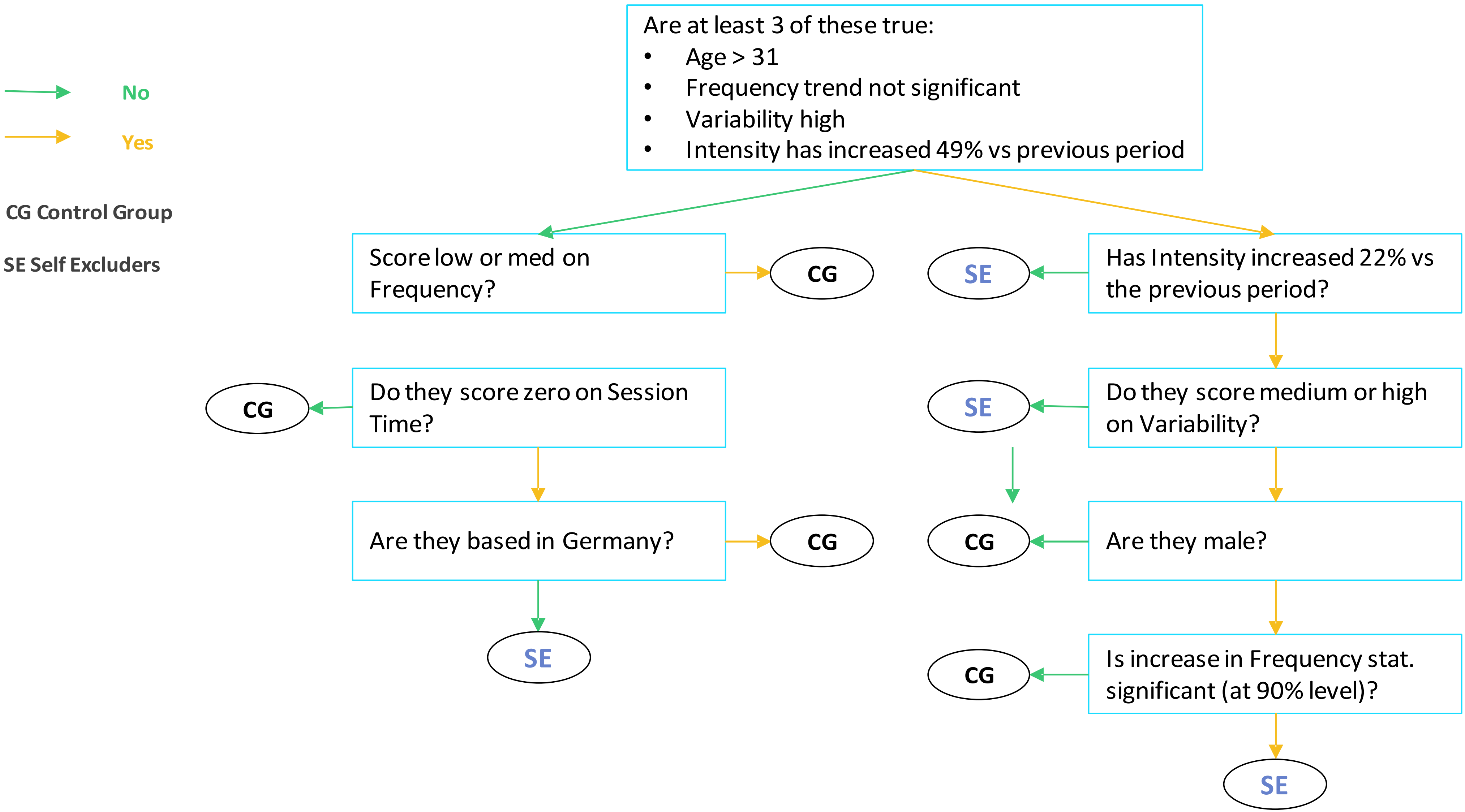
Recent application: Reducing harm from gambling

- 2014-18 EPSRC/InnovateUK projects with BetBuddy ltd.
- Trained a neural net to predict whether player should self-exclude from the game based on transaction data: frequency of play, betting intensity, variation, etc. (altogether some 25 markers)
- Used self-exclusion as a proxy for potential harm in online gambling (avoids use of much more complex model of addiction)

Variations of TREPAN



Example of Decision Tree Extracted



Reducing harm from gambling (cont.)

- Neural nets and Random Forests performed considerably better than logistic regression and Bayesian nets
- BetBuddy ltd. (recently acquired by FTSE 250 listed Playtech plc) is required to provide explanations to the regulator, gambling operator and the player
- Extracted decision tree can help debug the system and improve results too: “Are they based in Germany?”

The need for knowledge extraction (continued)

- How explainable is a decision tree with millions of nodes?
- Knowledge extraction allows a knowledge-base to be queried (goal-directed) producing a proof history (i.e. an explanation one gets for free with the answer)
- E.g. Suppose A implies B, B implies C, A is true. Is C true? Is B true? Is A true? Check!

Knowledge Consolidation

Challenge: efficient extraction of sound, comprehensible knowledge from large-scale networks (100's of neurons; 1000's of connections)

How about millions of connections?

What makes knowledge comprehensible?

Machine Learning with Constraints:

I. Donadello, L. Serafini and A. S. d'Avila Garcez. Logic Tensor Networks for Semantic Image Interpretation. In Proc. IJCAI'17, Melbourne, Australia, Aug 2017.

Knowledge-based Transfer Learning:

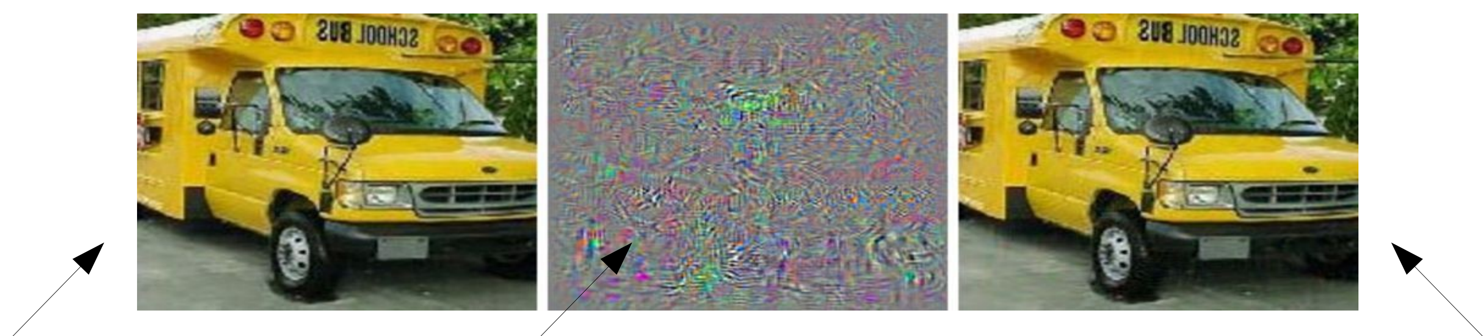
S. Tran and A. S. d'Avila Garcez. Deep Logic Networks: Inserting and Extracting Knowledge from Deep Belief Networks. IEEE TNNLS, Nov, 2016

Knowledge Extrapolation

Neural nets can perform many forms of reasoning (nonmonotonic, abductive, modal, temporal, epistemic, etc)

d'Avila Garcez, Broda and Gabbay. Neural-Symbolic Cognitive Reasoning, Springer, 2009

But can be very bad at extrapolation and sometime are not even robust...

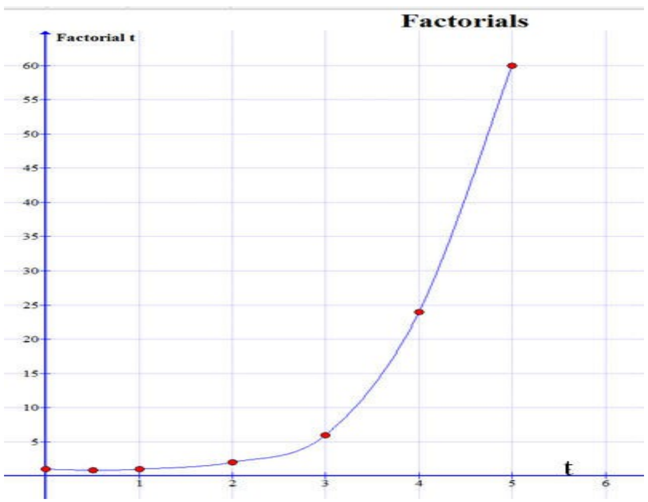


c.f. Intriguing Properties of Neural Networks, Szegedy et al., https://arxiv.org/abs/1312.6199, 2014

Deep Learning vs. Symbolic ML

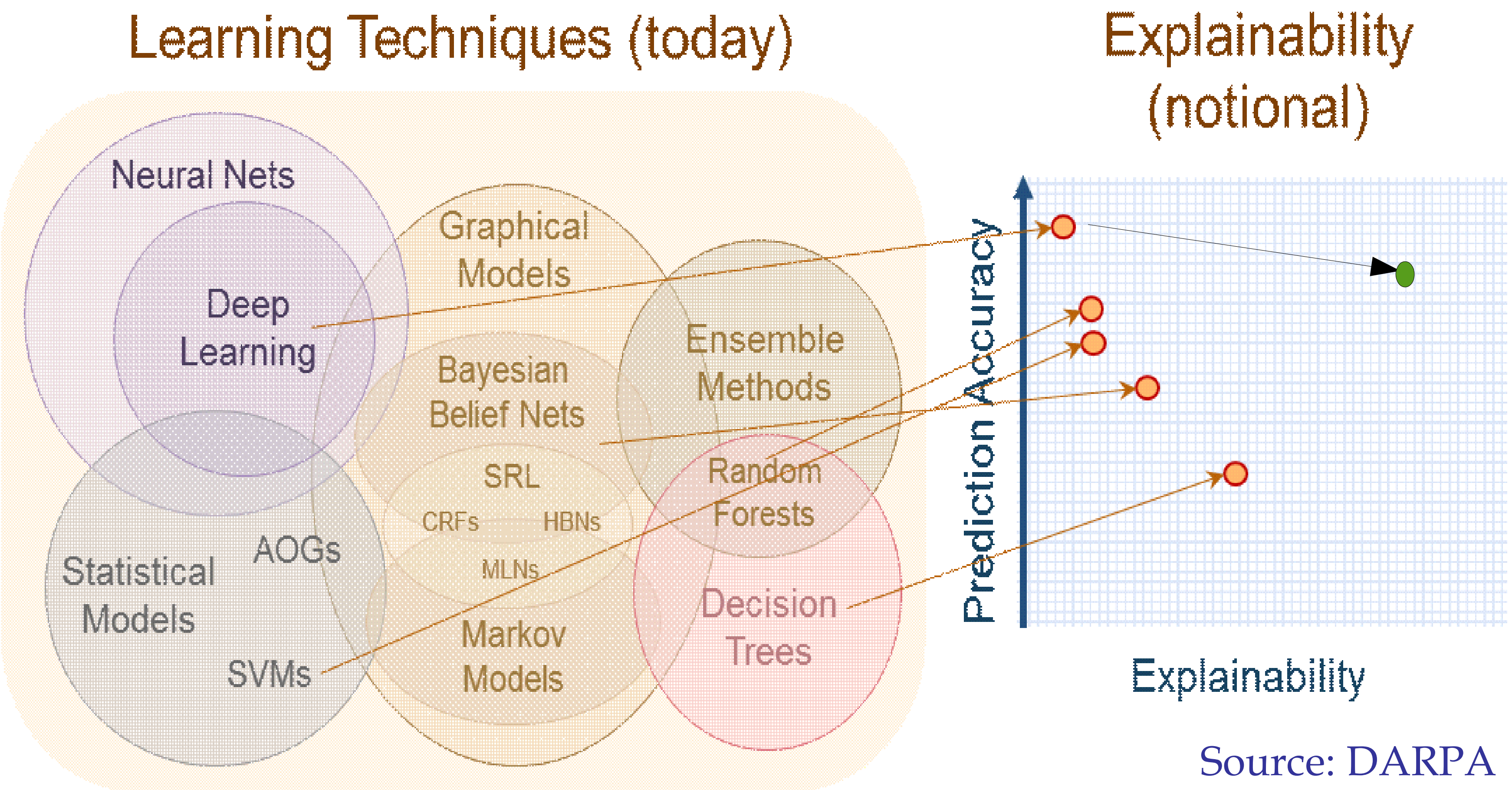
- We want good classification and prediction but also useful descriptions... e.g.: learning factorial (*n!*)

$$
\begin{gathered}
0! = 1 \\
1! = 1 \\
2! = 1 \cdot 2 = 2 \\
3! = 1 \cdot 2 \cdot 3 = 6 \\
4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24 \\
5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120 \\
6! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720
\end{gathered}
$$




```
factorial(0,1).
factorial(N,F) :- N>0, N1 is N-1,
                  factorial(N1,F1),
                  F is N * F1.
```

Explainable AI = ML + KR



Source: DARPA

Relevant/Recent Developments

- MIT's neural-symbolic system for image domains
- Rich Caruana's teacher/student models
- Neural-Symbolic Computing (NeSy 2019 at IJCAI, Macau, Aug 2019, www.neural-symbolic.org):
  - Deep Logic Networks,
  - Logic Tensor Networks,
  - Layerwise Knowledge Extraction as Learning
  - Counterfactual Local Explanations
- Information Bottleneck
- Disentangled Variational Autoencoders
- Binarized nets and other prunning methods
- Explanation by design (as a requirement or as constraints)

Conclusion: Why Neurons and Symbols

To study the statistical nature of learning and the logical nature of reasoning.

To provide a unifying foundation for robust learning and efficient reasoning.

To develop effective computational systems for AI applications.

Thank you!

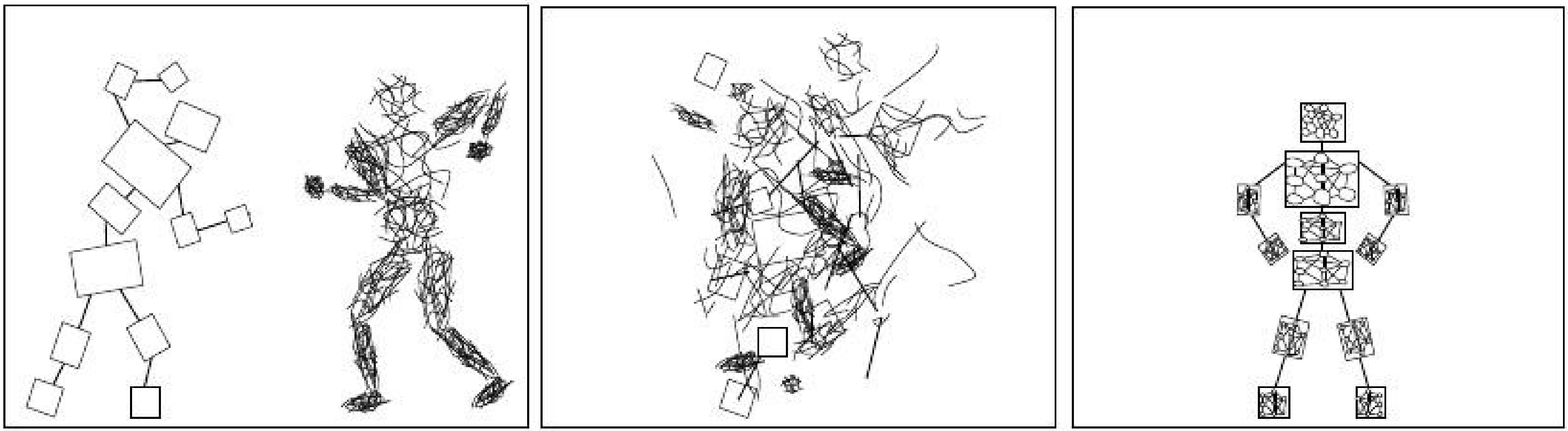


Figure 1. Conflict between theoretical extremes.

Attributed to Marvin Minsky

Logical Versus Analogical or Symbolic Versus Connectionist or Neat Versus Scruffy, AI Magazine, 1991