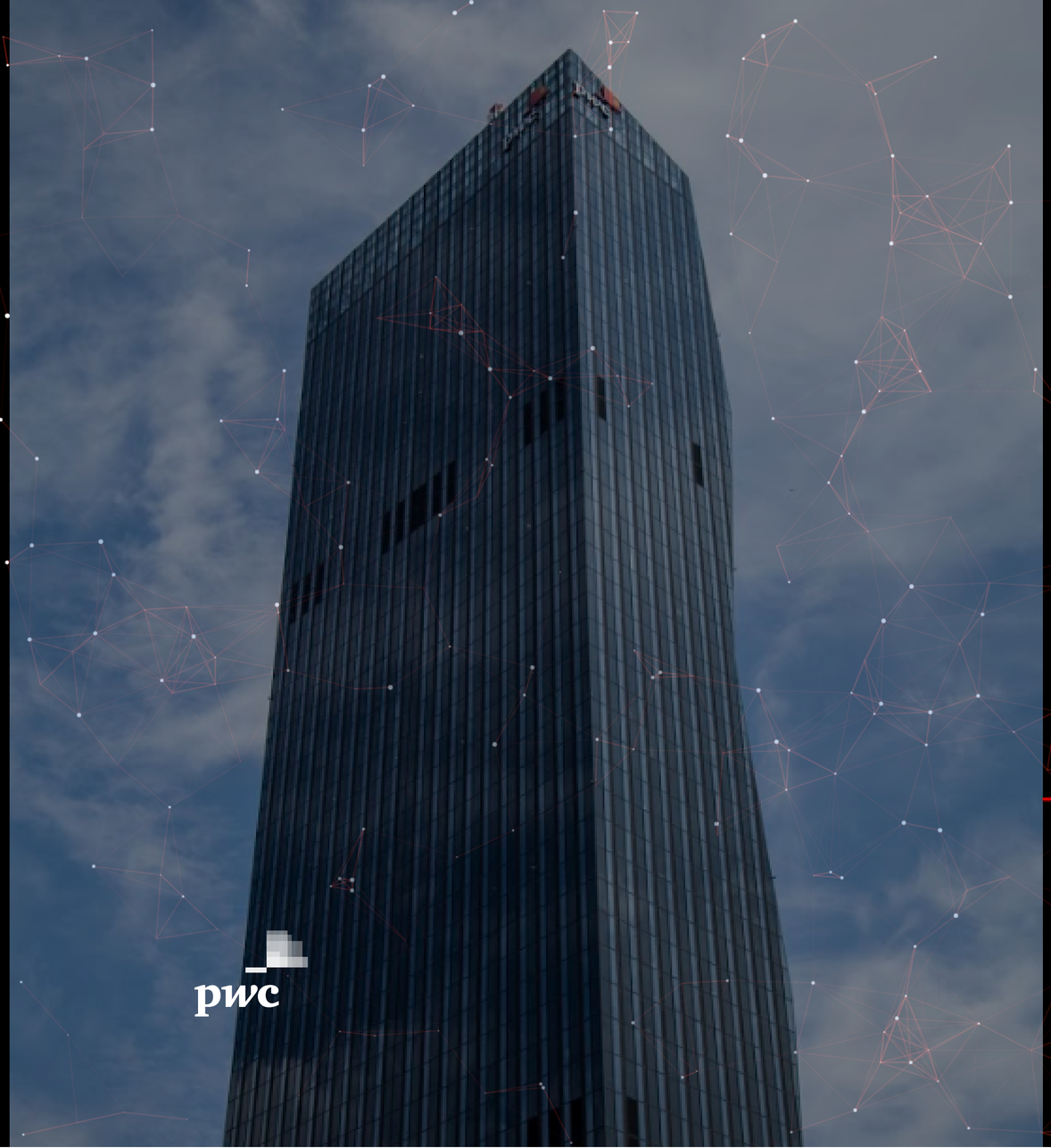


Explainable Neural Symbolic Learning

Ahmad Haj Mosa

ahmad.haj.mosa@pwc.com

LinkedIn: ahmad-haj-mosa

Fabian Schneider

schneider.fabian@pwc.com

LinkedIn: fabian-schneider-24122379

The definition of intelligence

Our minds contain processes that enable us to solve problems we consider difficult.

"Intelligence" is our name of those processes we don't yet understand.

Marvin Minsky

Interpretation vs explanation

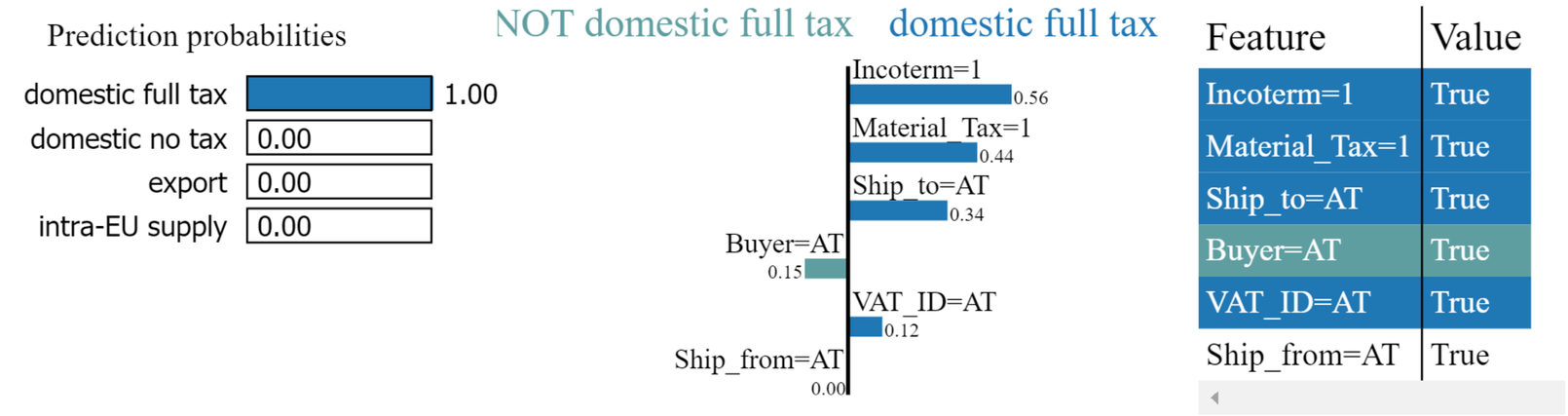
Interpretation	Explanation
Between AI and its designer	Between AI and its consumer
External analytics	Learned by the model
Model specific or agnostic	Model specific
Local or global	local
examples: LIME, SHAP and TSNE	examples: Attention Mechanism

Classification of Value Added Tax

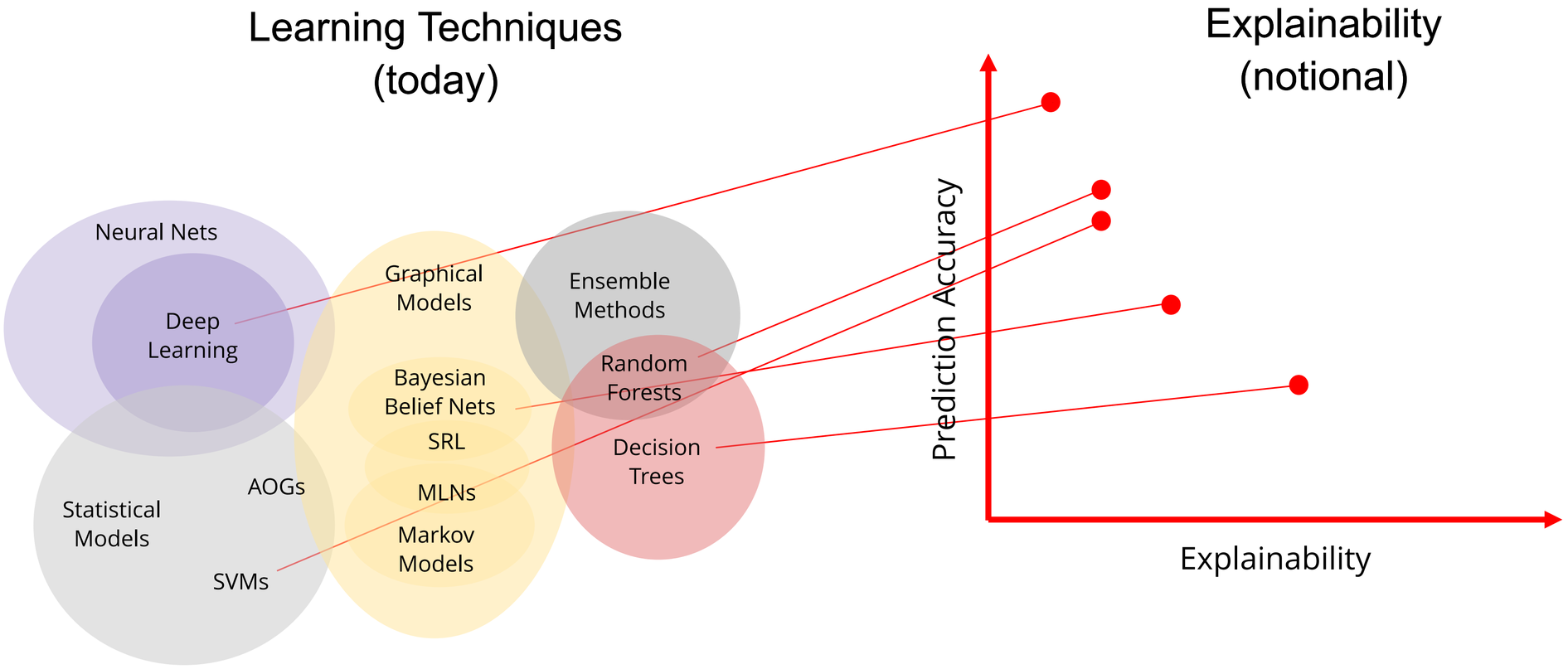
The first rule is: if your are shipping a product from Austria to Austria and the shipped material is according to law taxable, then the Taxcode is: domestic full tax -> 20%

	Ship_from	Ship_to	Buyer	Material_Tax	Incoterm	VAT_ID	Taxcode	Taxrate
114	AT	AT	AT	1	1	AT	domestic full tax	0.2

Neural Net

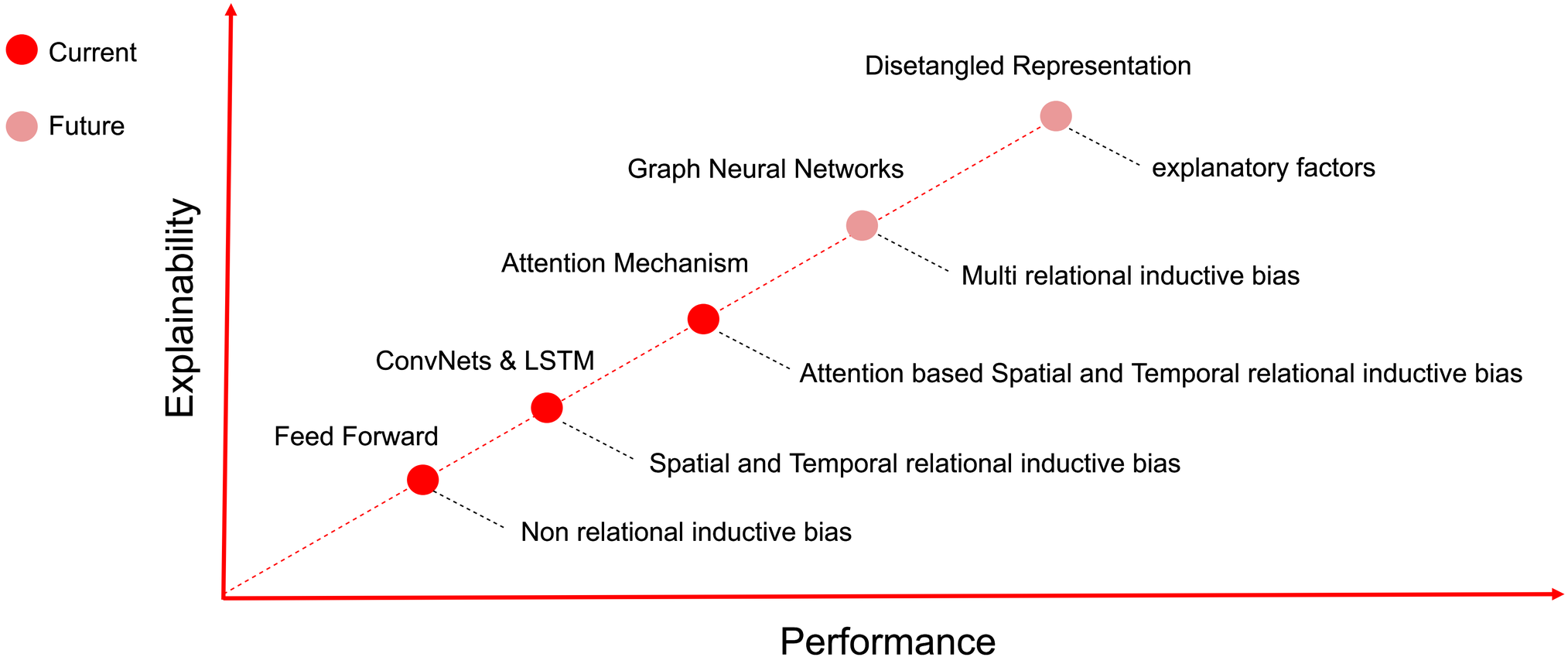


Is it explainability vs accuracy?



source: DARPA

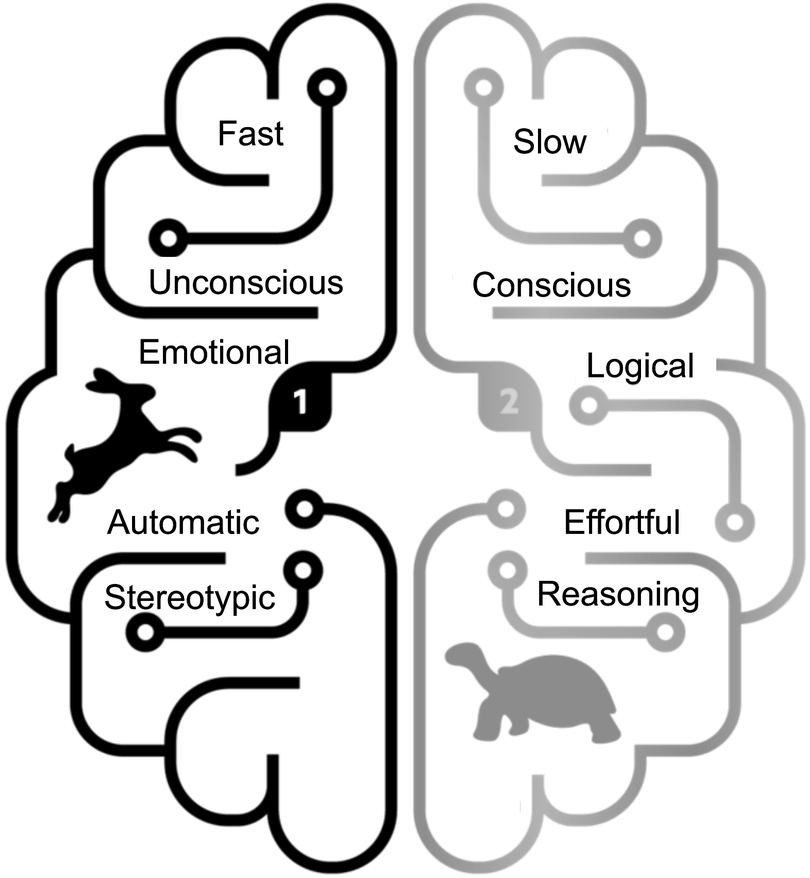
Is it explainability vs accuracy?



What is next in Deep Learning?

Thinking fast and slow

System 1	System 2
drive a car on highways	drive a car in cities
come up with a good chess move (if you're a chess master)	point your attention towards the clowns at the circus
understands simple sentences	understands law clauses
correlation	causation
hard to explain	easy to explain

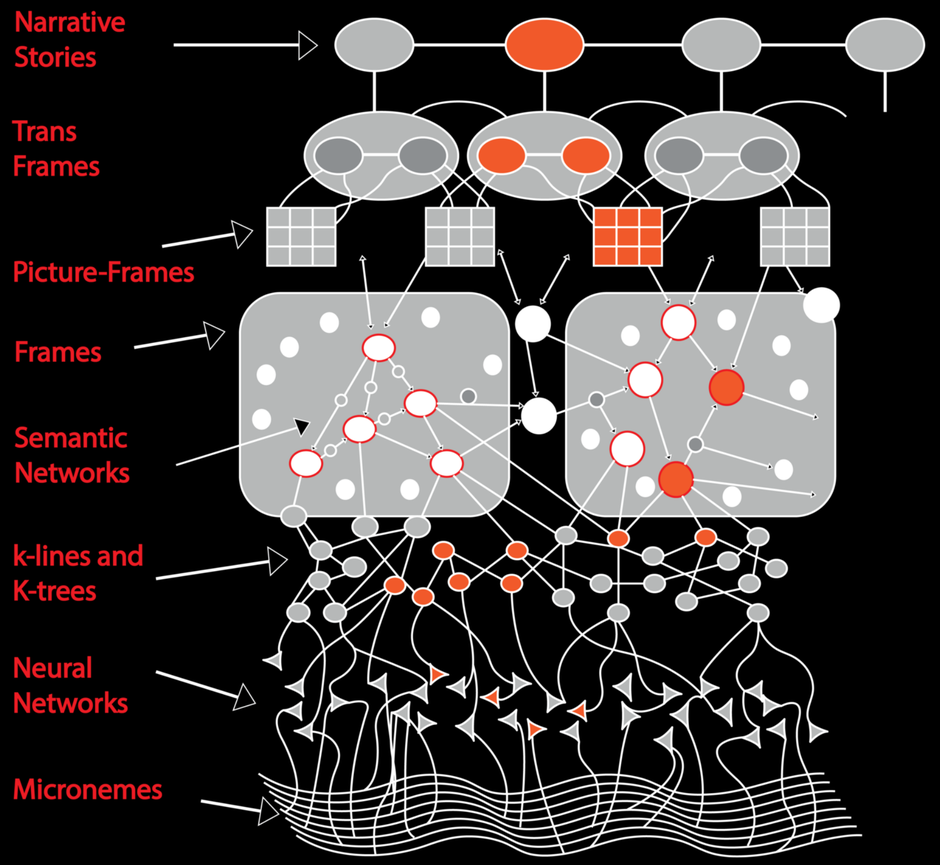


source: Thinking fast and slow by Daniel Kahneman

A framework for representing knowledge

easy
explanation
learning fast

hard
explanation
learning slow

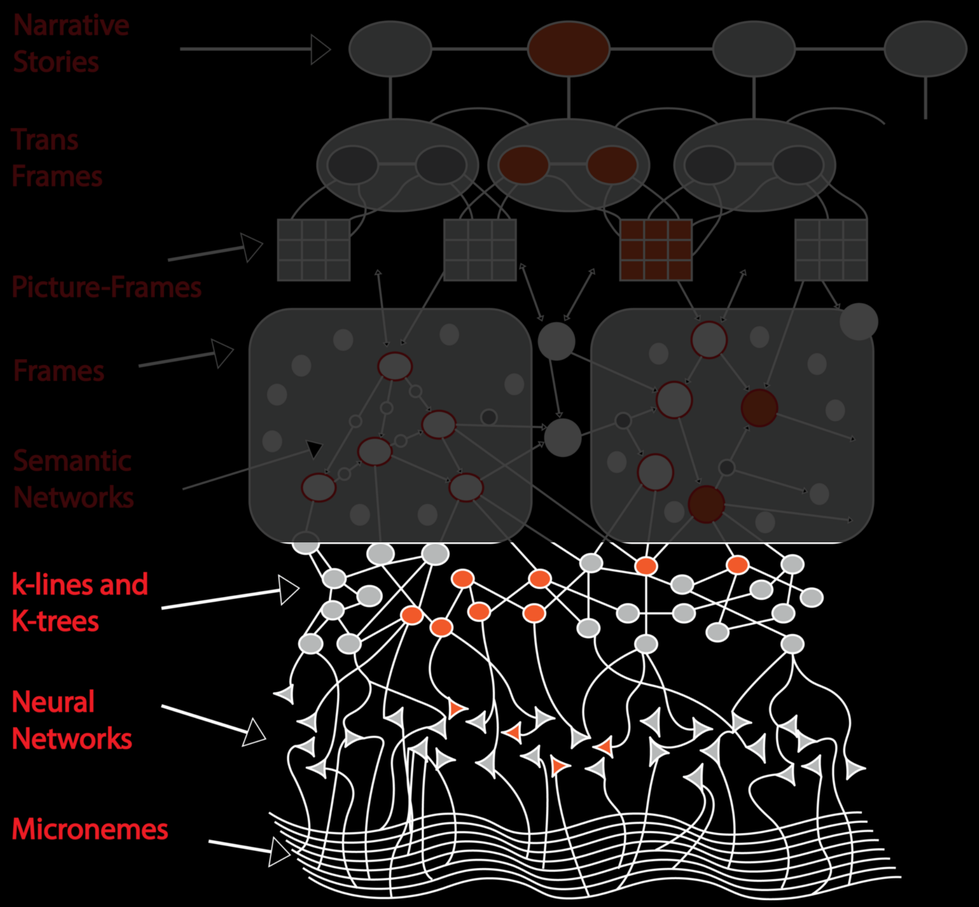


consciousness prior

thinking slow

thinking fast

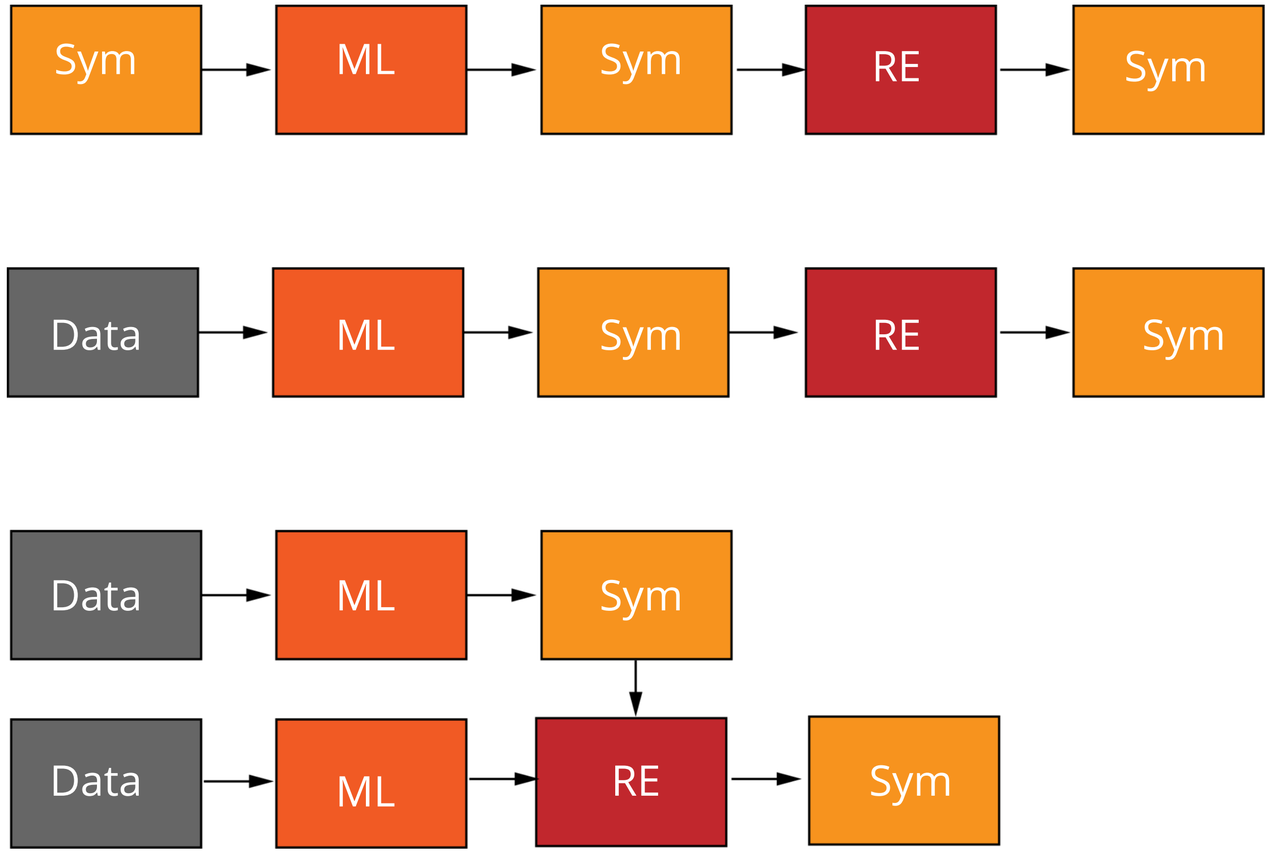
pwc



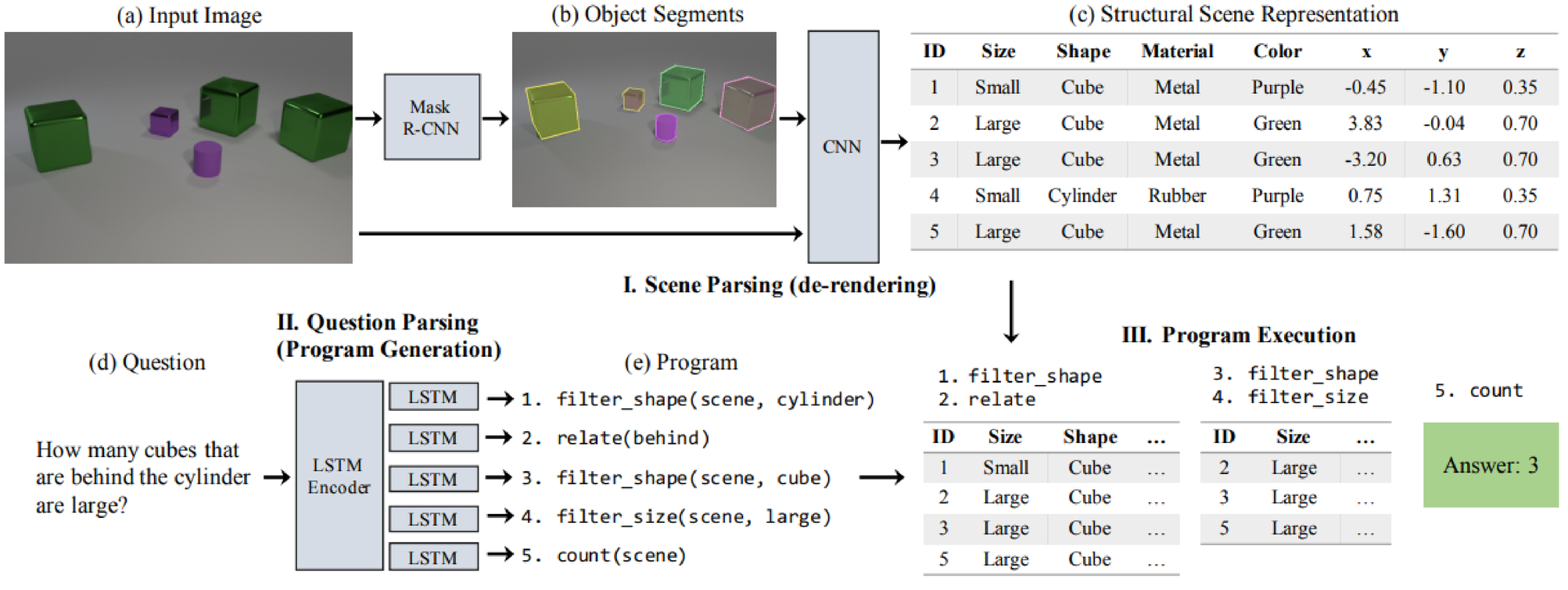
Neural Symbolic Learning

Neural Symbolic AI Models

- Sym: Symbolic representation (graph and/or knowledge base)
- ML: Statistical Machine Leaning (mainly deep learning)
- RE: Logical Reasoning Engine
- Data: raw data (image, voice, structured data .etc)



Neural symbolic and disentangled reasoning

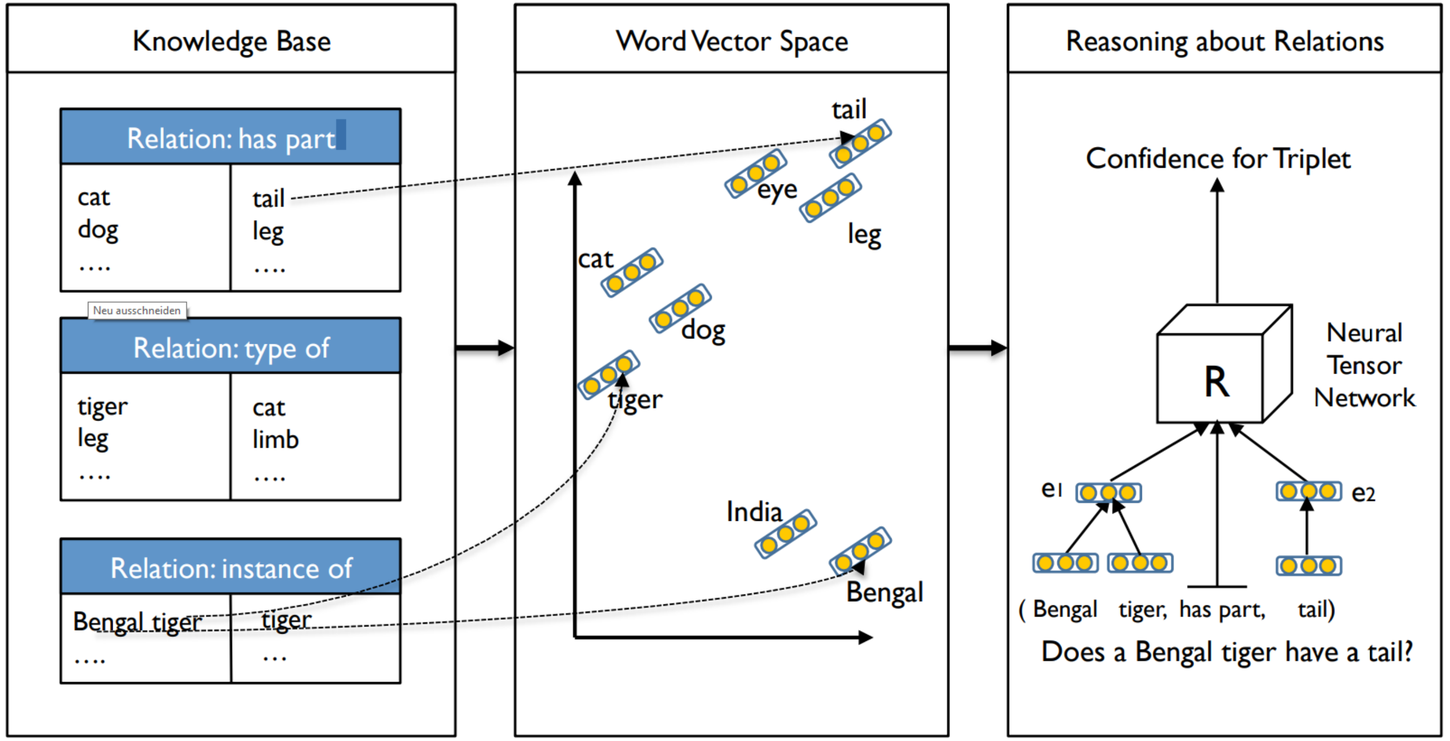


source: arXiv:1810.02338: Neural-Symbolic VQA: Disentangling Reasoning from Vision and Language Understanding

Logic Tensor Networks

- Defined FOL for real valued vectors (RealLogic)
- Defined semantic grounding for RealLogic
- Using Neural Tensor Network (NTN) for inferring new knowledge

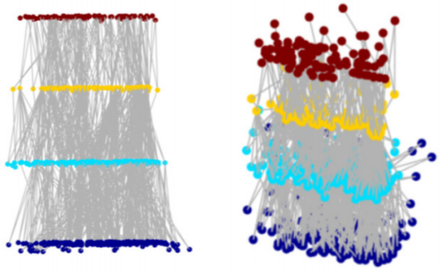
NTN: learning representation of entities and their relations in a KB unsupervised using tensor representation

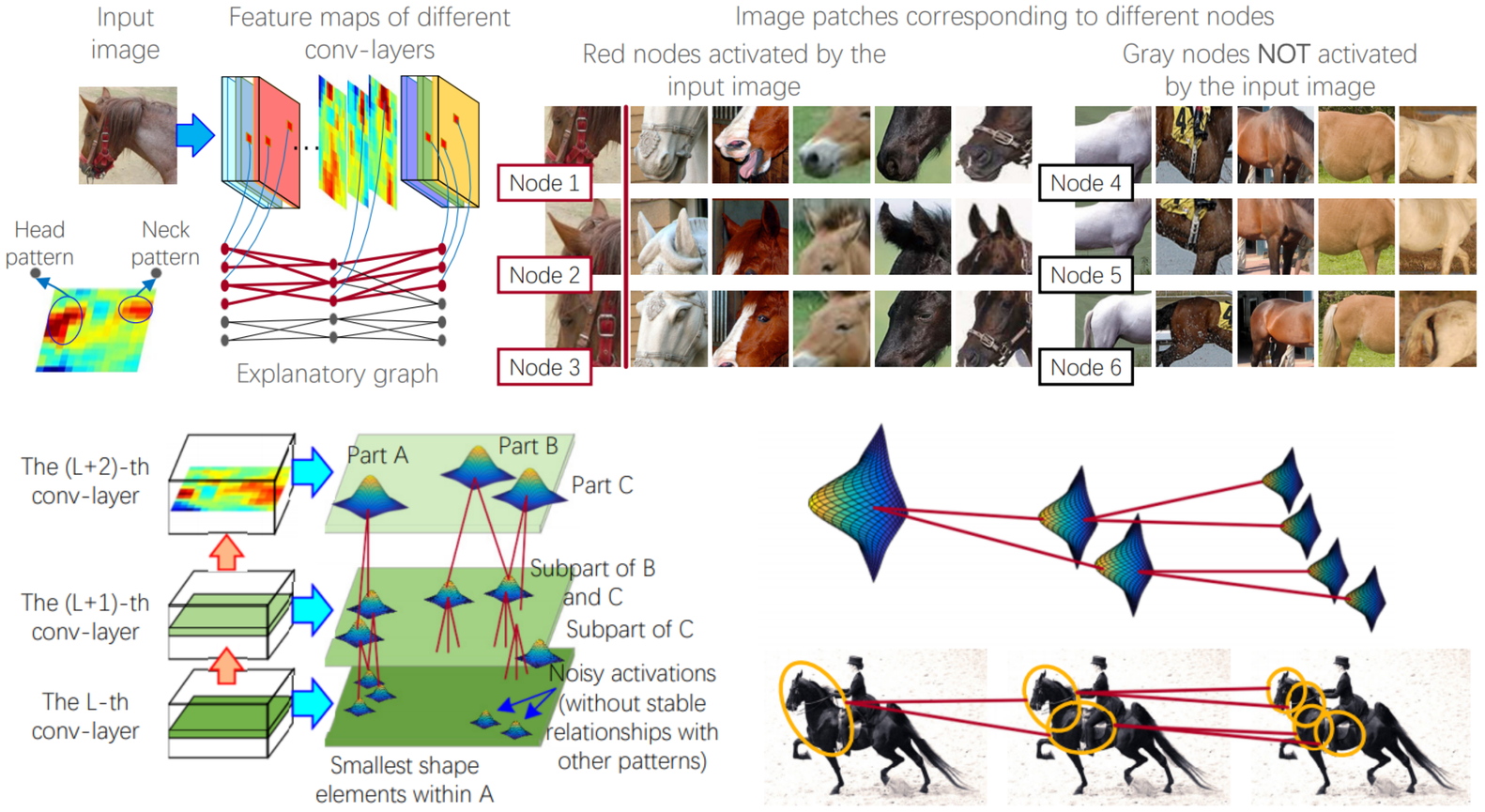


Socher, Richard, et al. "Reasoning with neural tensor networks for knowledge base completion." *Advances in neural information processing systems*. 2013.

Serafini, Luciano, and Artur d'Avila Garcez. "Logic tensor networks: Deep learning and logical reasoning from data and knowledge." *arXiv preprint arXiv:1606.04422* (2016).

Explanatory Graphs for CNNs





source: arXiv:1812.07997: Explanatory Graphs for CNNs

Framework 1: Logical Simplification using Tree Search, Graph Neural Network and Reinforcement Learning

Logical Simplification using TS, GN and RL

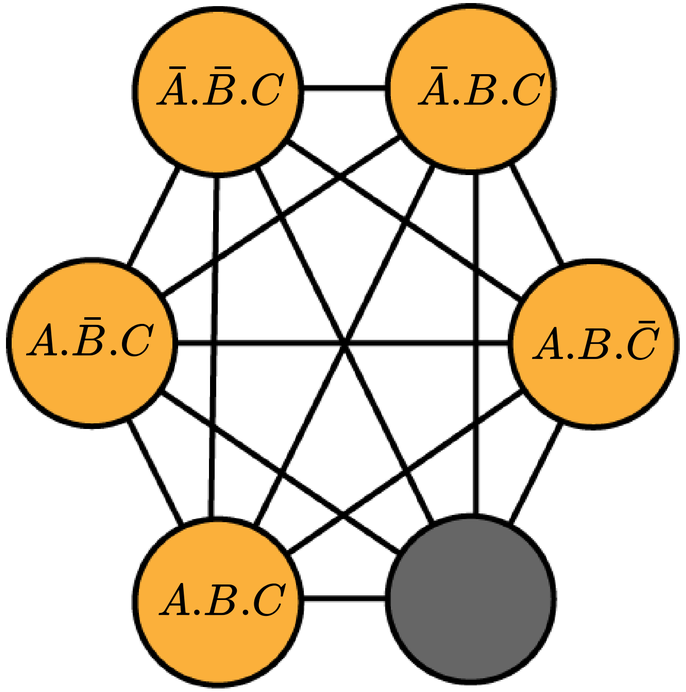
A	B	C	Y
0	0	0	0
0	0	1	1
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

Truth Table

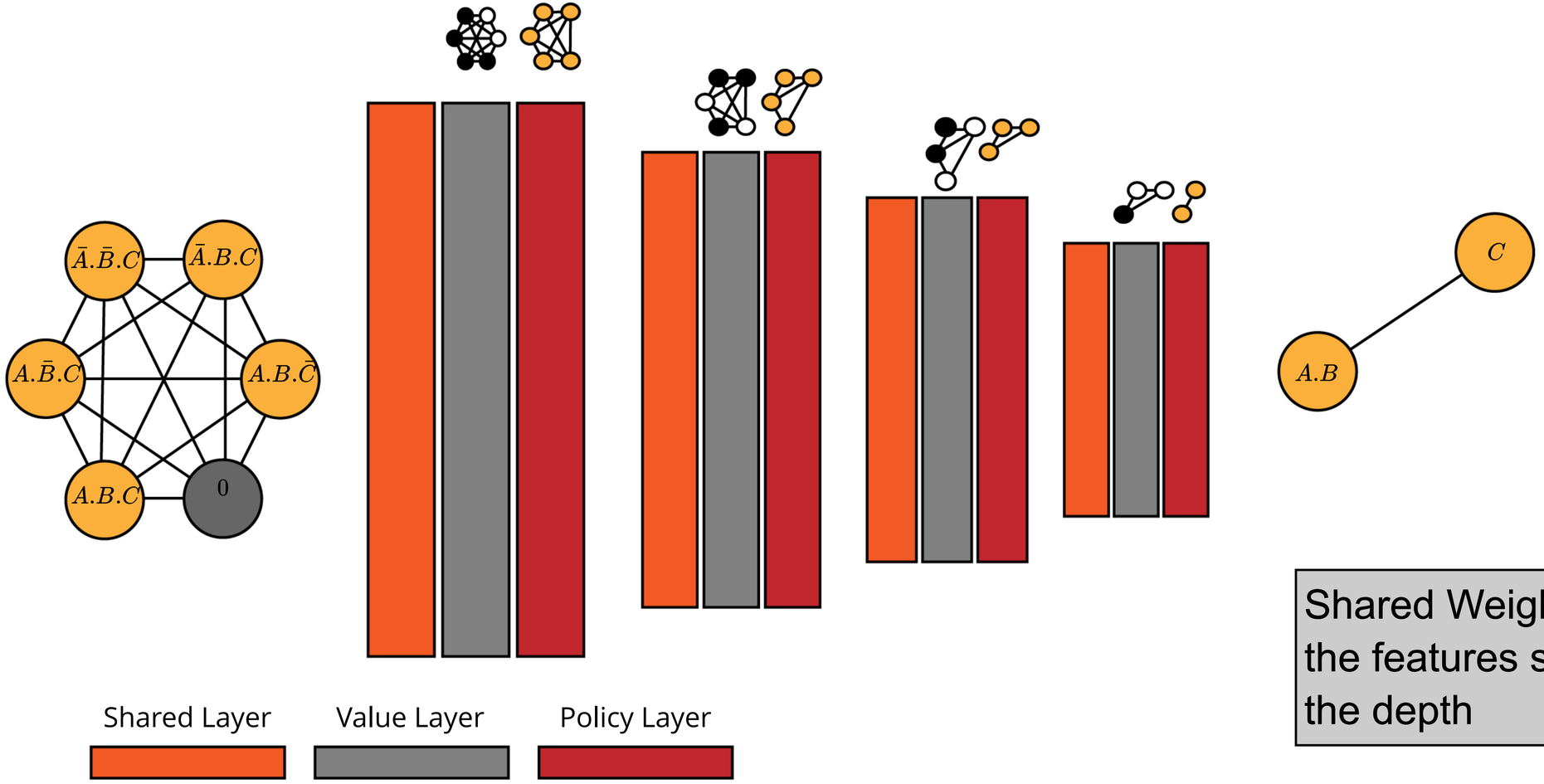
Initial $Y = \bar{A}.\bar{B}.C + \bar{A}.B.C + A.\bar{B}.C + A.B.\bar{C}. + A.B.C$

Target $Y = A.B + C$

Graph Representation

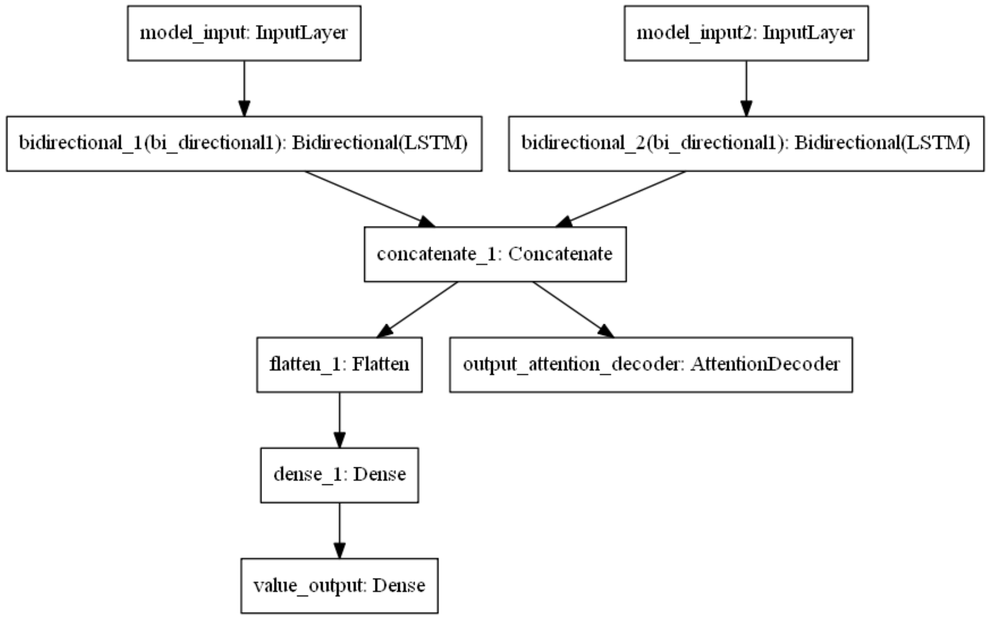


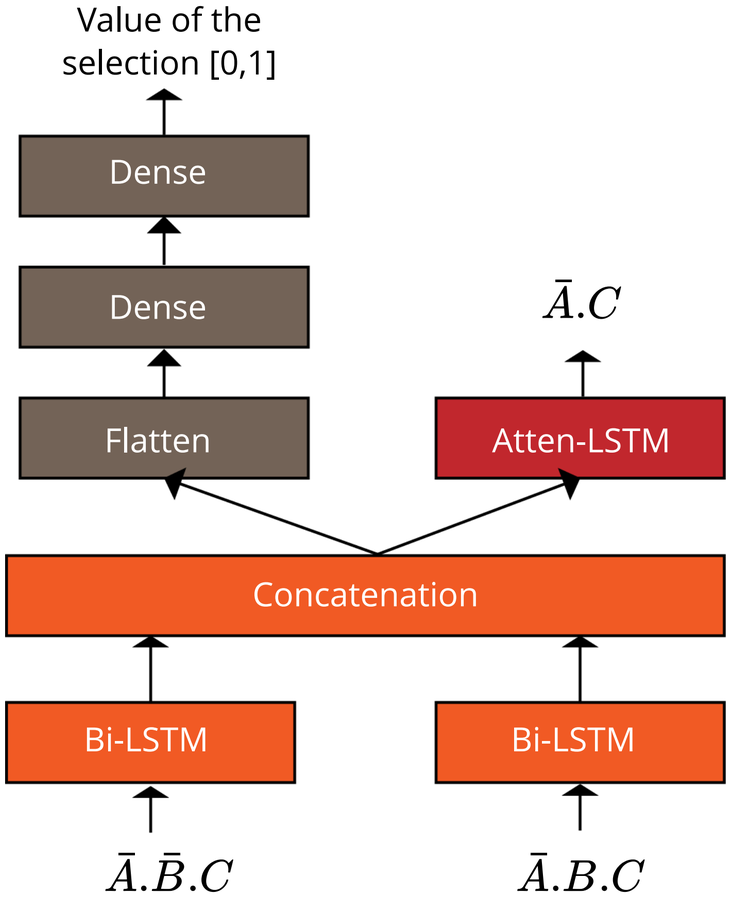
Logical Simplification using TS, GN and RL



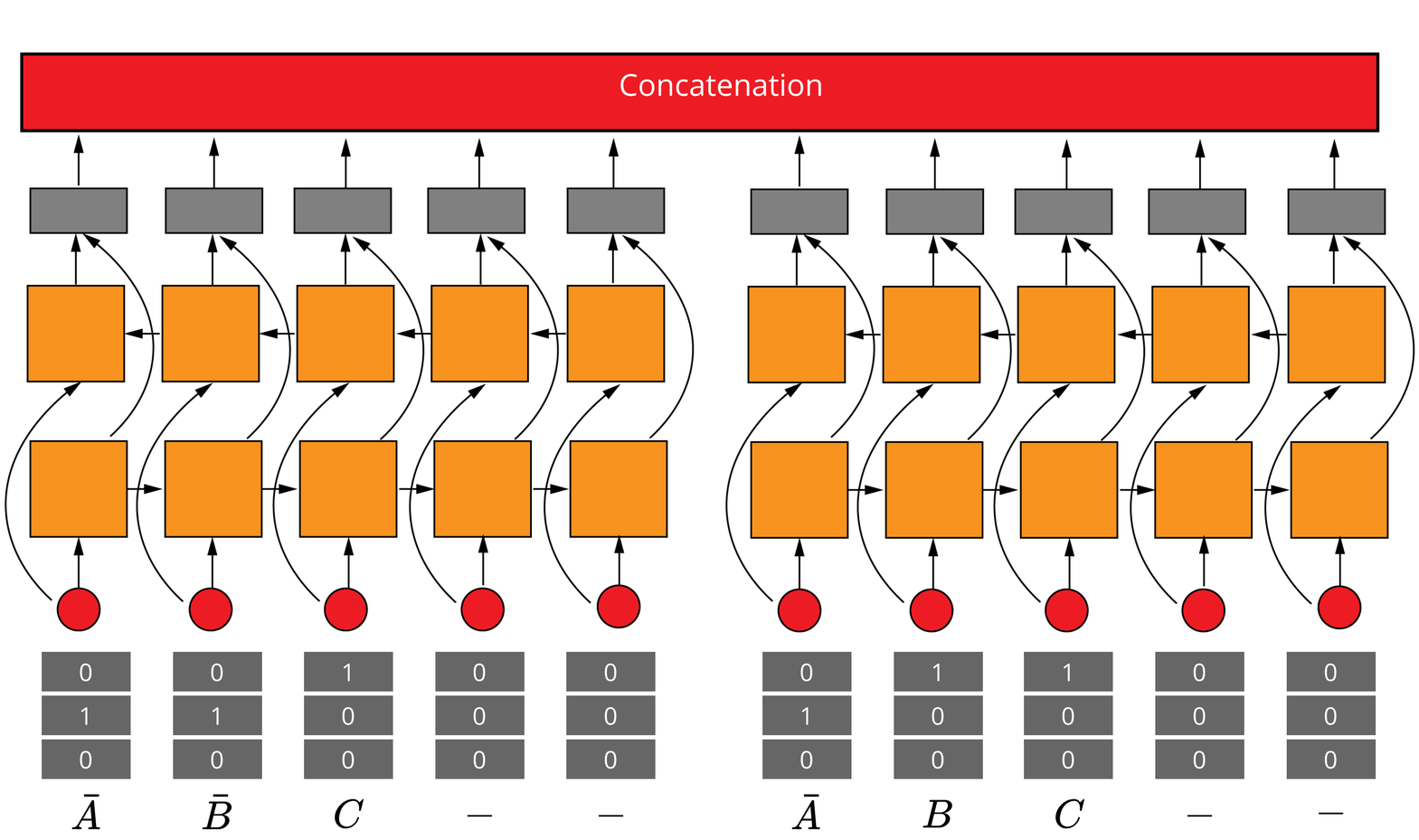
Shared Weights along the features space and the depth

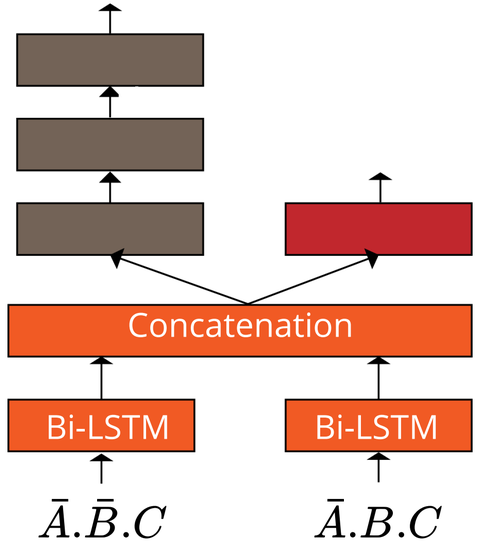
Simplifier Block





Simplifier Block





Node Features

1	0	0	0
0	1	0	0
0	0	0	0
A	$\bar{B}$	–	–

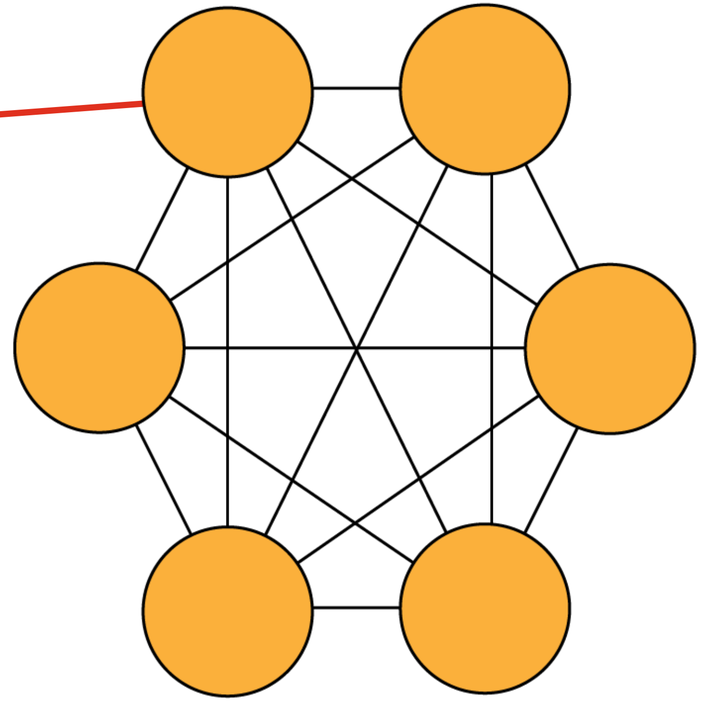
0	1	1	0
0	0	0	0
0	0	0	0
–	B	C	–

Binary Logic

ST	SF	TC	IT
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	0
AT	DE	$A0$	–

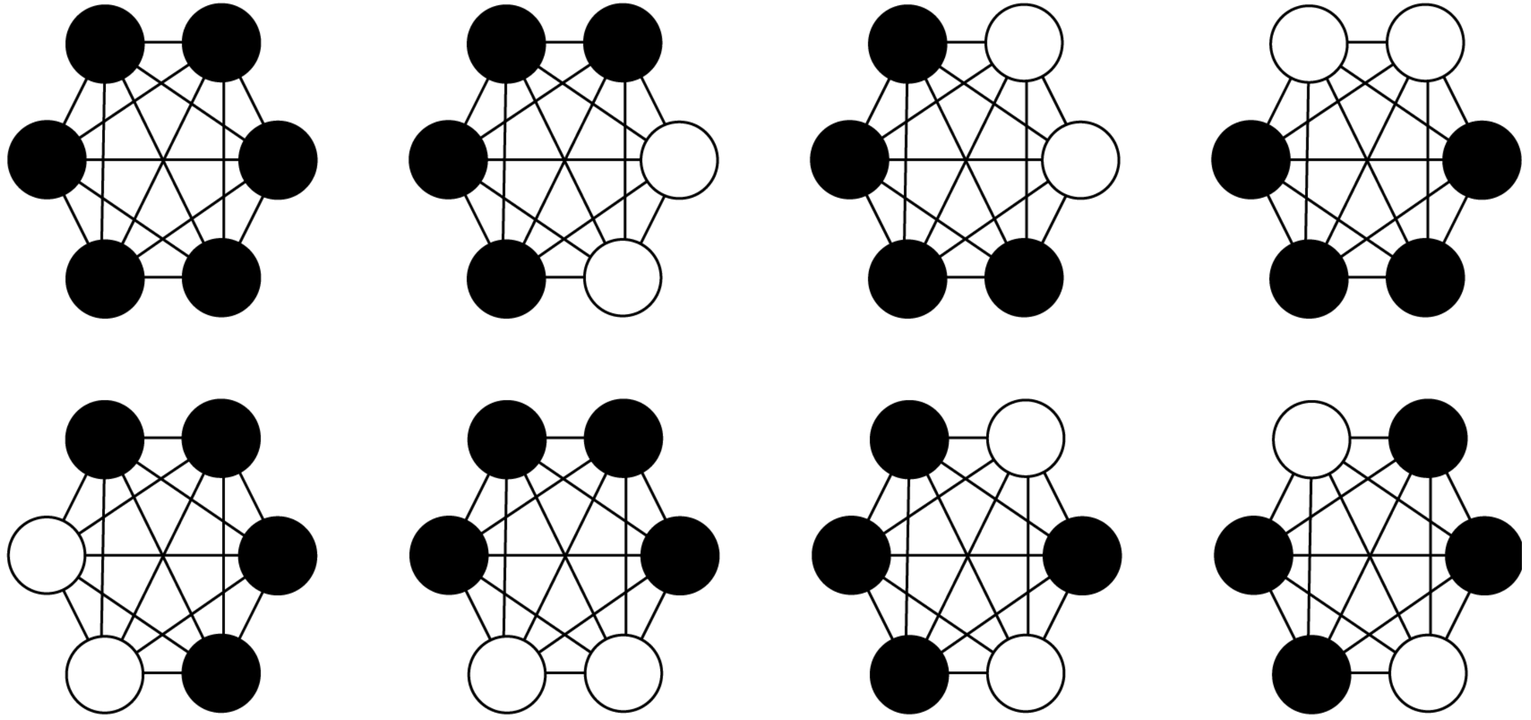
ST	SF	TC	IT
0	0	1	0
1	1	0	0
0	0	0	0
0	0	0	0
FR	DE	$A1$	–

Multivariate Logic

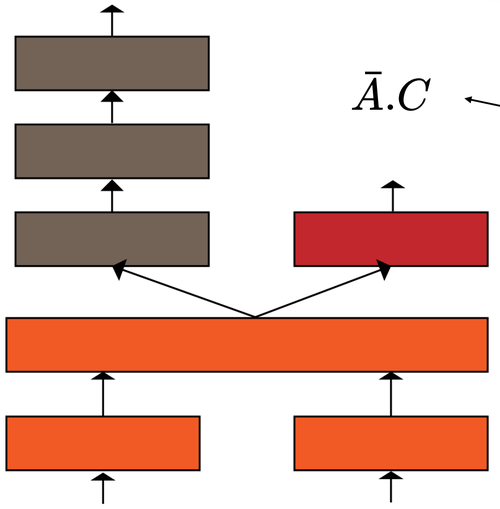


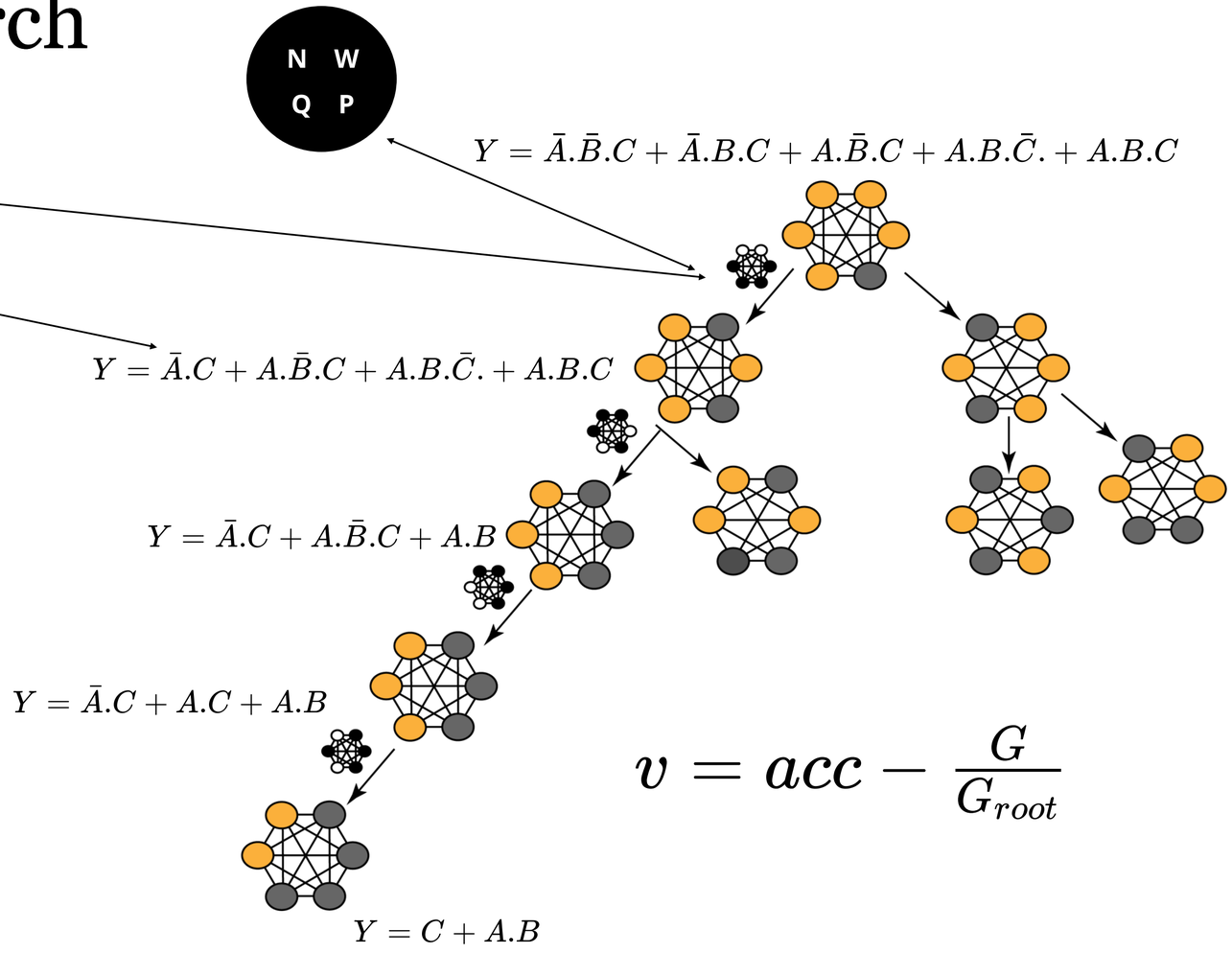
Convolution over the Graph

- Apply the simplifier block on every pair of nodes
- In case of large graph, apply the block locally on adjacent nodes



Monte Carlo Tree Search





Accumulated value W = W + v

Number of visits N = N + 1

Mean value Q = W / N

Prior probability of an action P

Active nodes in the graph: G

MCT + RL

Repeate the follwoing steps for M time:

- Generate Binary data
- Generate Target Rule
- Simulate the MCT:
  - Choose the action that maximises Q + U
  - Continue until the leaf node is reached
  - Rollout W, N, Q
  - Train the policy layer using policy gradient
  - Train the value layer by minimizing MSE between W and the predicted W

$$v = acc - \frac{G}{G_{root}}$$

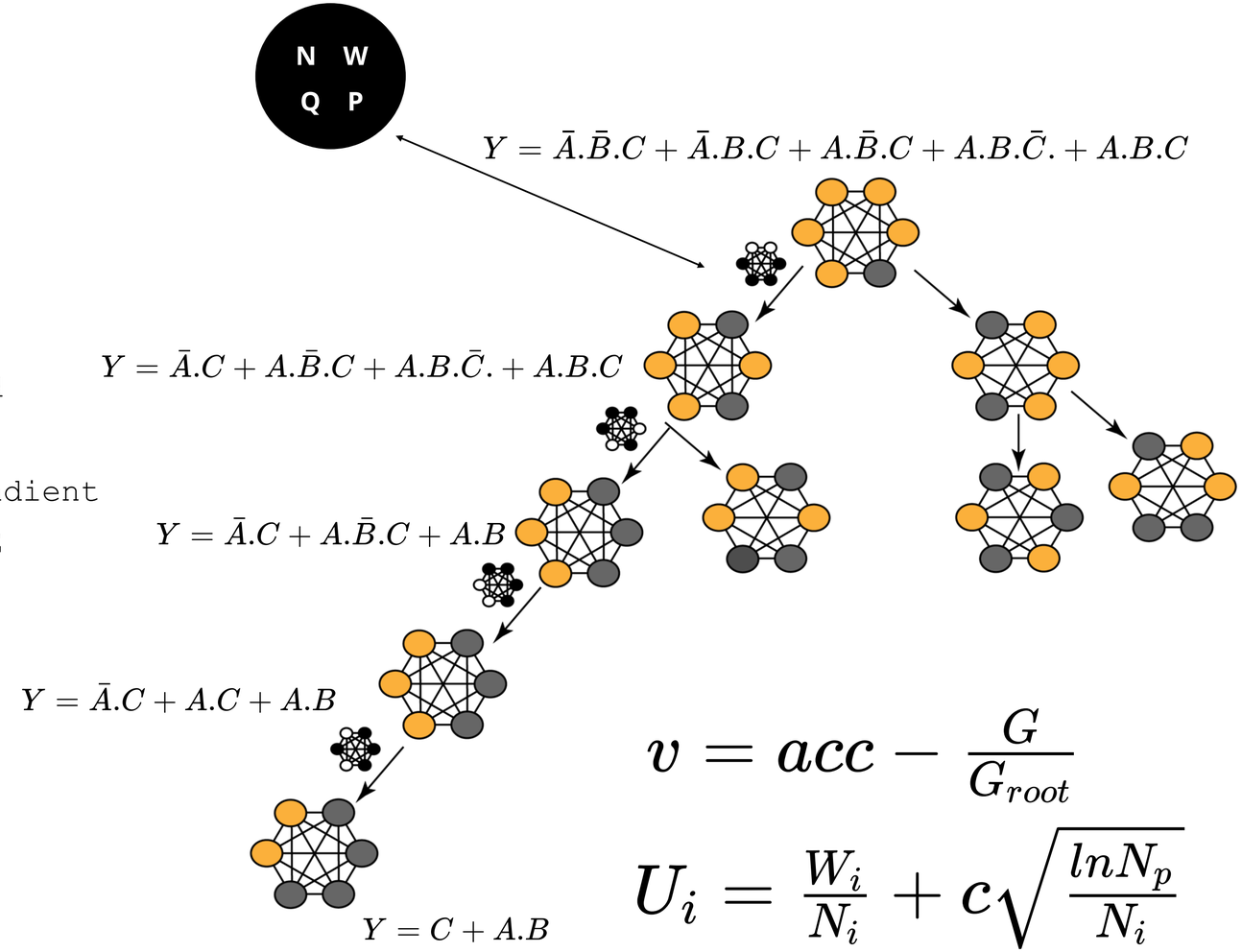
Accumulated value W = W + v

Number of visits N = N + 1

Mean value Q = W / N

Prior probability of an action P

Active nodes in the graph: G



$$v = acc - \frac{G}{G_{root}}$$

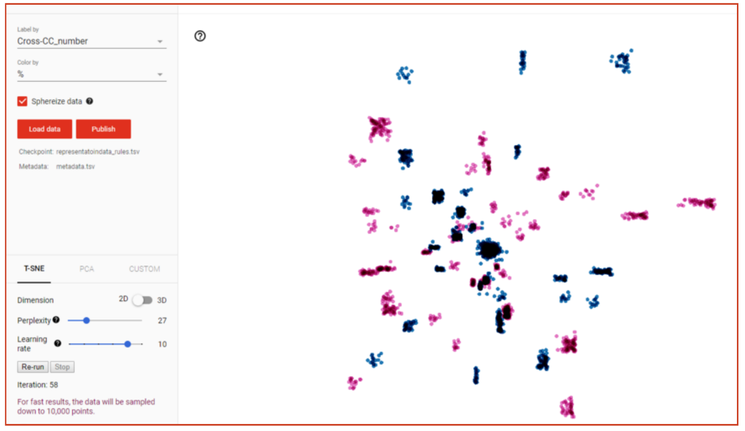
$$U_i = \frac{W_i}{N_i} + c\sqrt{\frac{lnN_p}{N_i}}$$

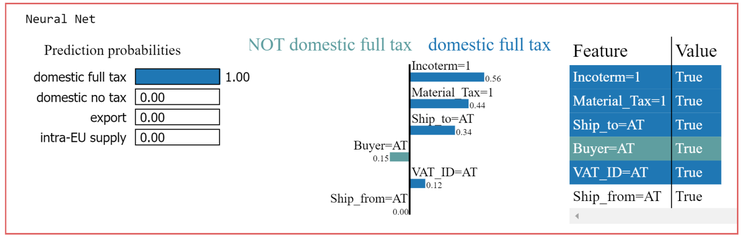
Framework 2: DL Disentangled Representation

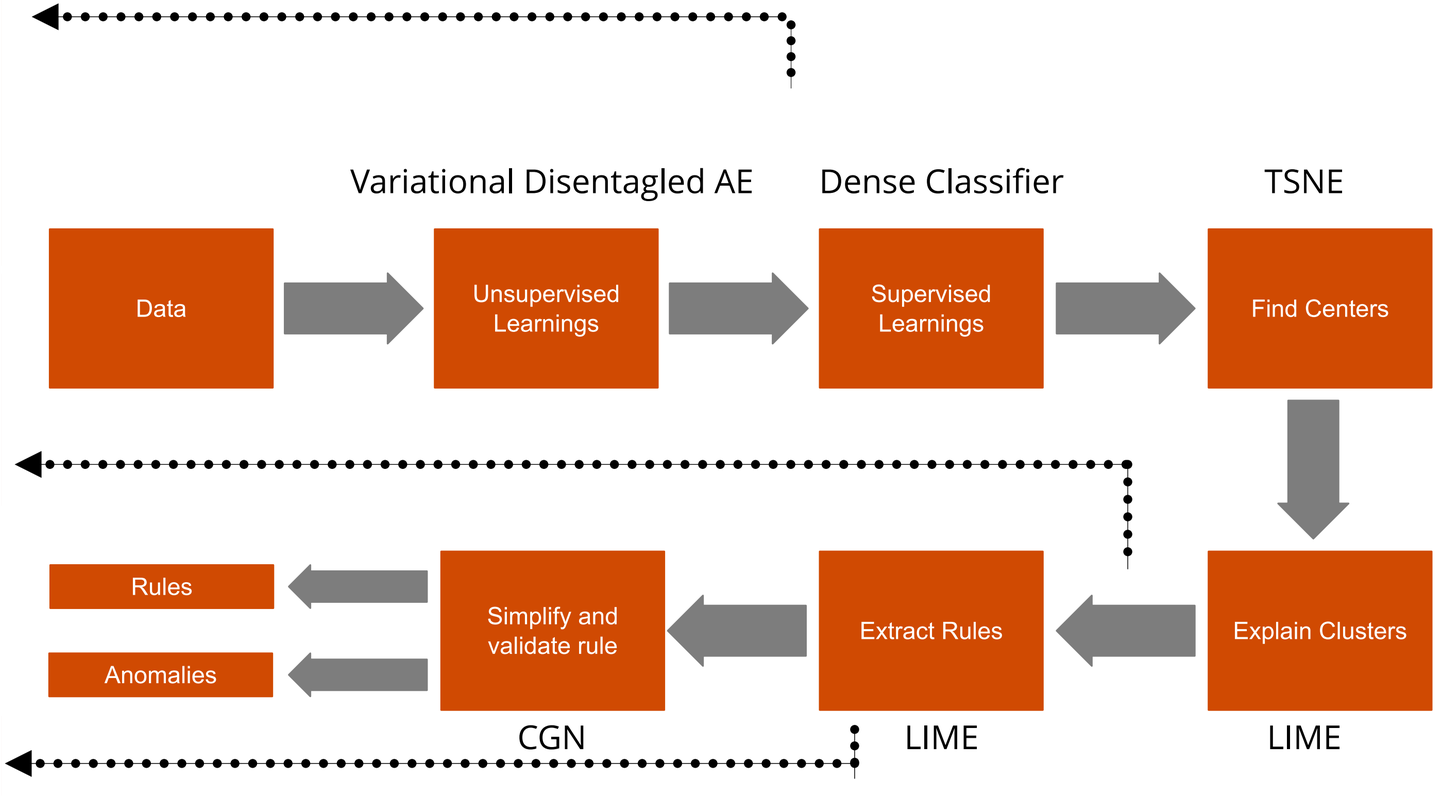
+ Functional Programming Computaiotnal

Graph

Framework 2: DL representation + CGN reasoning

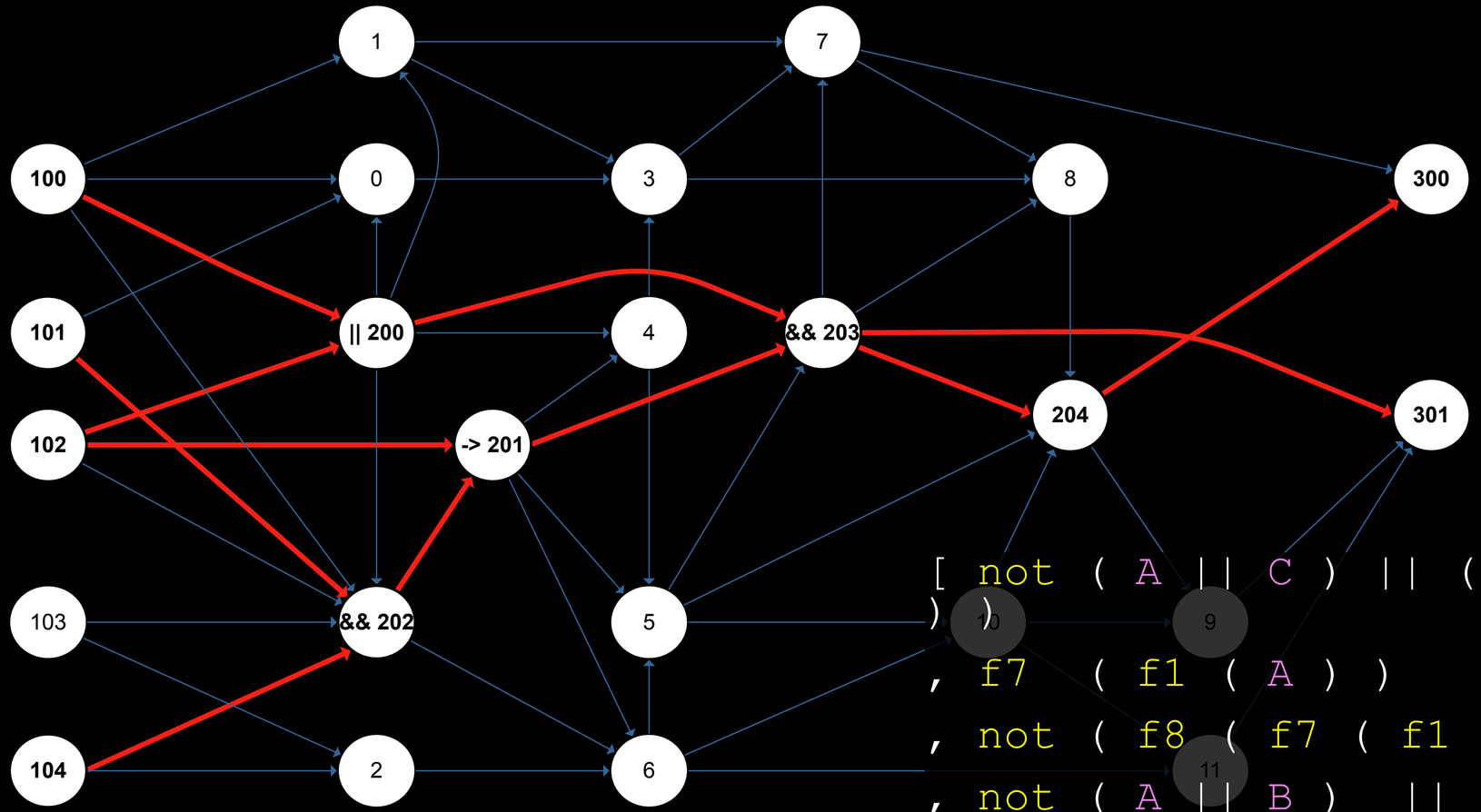






(Ship_to = AT) & (Material_Tax =1) & (IncoTerm =1)
& (VAT_ID =AT) --> domestic full tax

Computational Graph Network




```
[ not ( A || C ) || ( C && not ( B && E
) )
, f7  ( f1 ( A ) )
, not ( f8 ( f7 ( f1 ( A ) ) ) )
, not ( A || B )  || implication( C, D
)
, ... ]
```

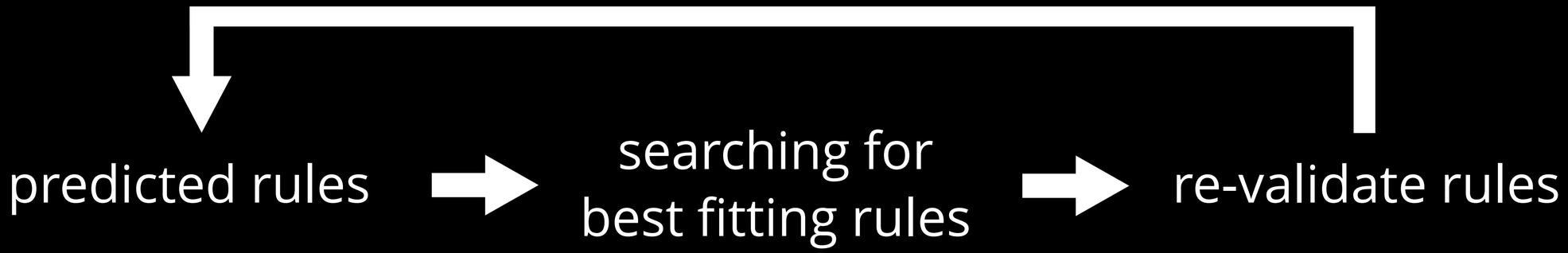
Haskell

$$\varphi(x) = f \circ g\ (x,?)? = f(g(x,?))?$$

apply φ to input so we get result

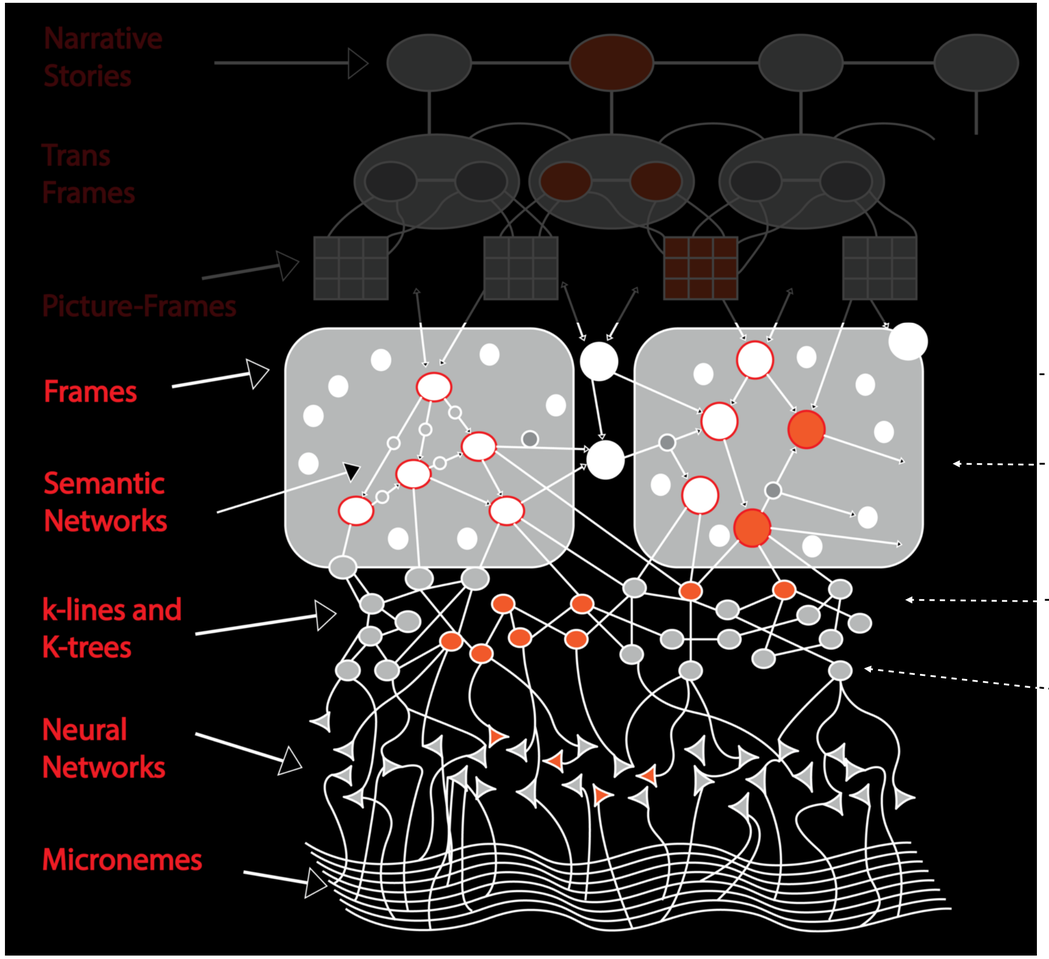
```
(.-) [[a] -> [b]] -> ([[b]] -> [a] -> [c]) -> [a] -> [c]
(.-) funcList          newFunc                  inp =
                       newFunc (map ($inp) funcList) inp
-- where
--     funcList = Previous Node Function List
--     newFunc  = Node Function
--     inp = Sample Input
```

Usage of CGN


predicted rules
searching for
best fitting rules
re-validate rules

A framework for representing knowledge

"When you "get an idea," or "solve a problem" ... you create what we shall call a K-line. ... When that K-line is later "activated", it reactivates ... mental agencies, creating a partial mental state "resembling the original."



High Level SR

Consciousness priors

Low Level SR

Disentangled, distributed or semantic representation

Conclusion

- Accuracy is not enough to trust an AI
- Accuracy vs Explainability is not a trade-off

Look for solutions in weird places:

- Try Functional Programming & Category Theory

Trends in XAI:

- closing the gap between Symbolic AI and DL
- disentangled representation
- Object-Oriented-Representation
- Computational Graph Networks

+ Don't let your robot read legal texts ;)

Thank you for your attention

Ahmad Haj Mosa
AI Researcher
ahmad.haj.mosa@pwc.com

Fabian Schneider
PoC-Engineer & Researcher
schneider.fabian@pwc.com