

Neuro-symbolic integration for ontology-based classification of structured objects

Till Mossakowski



OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG



FAKULTÄT FÜR
INFORMATIK

Ontolog summit, March 20, 2024

Overview

- 1 Symbolic knowledge representation
- 2 Neuro-symbolic extension of ontologies
- 3 Transformers for more performance and interpretability
- 4 Ontology pre-training can improve Transformer performance
- 5 Outlook

Overview

1 Symbolic knowledge representation

2 Neuro-symbolic extension of ontologies

3 Transformers for more performance and interpretability

4 Ontology pre-training can improve Transformer performance

5 Outlook

Ontologies provide high-quality symbolic knowledge: e.g. Ontology of Chemical Entities of Biological Interest (ChEBI)

 **ChEBI**

Search
 Advanced

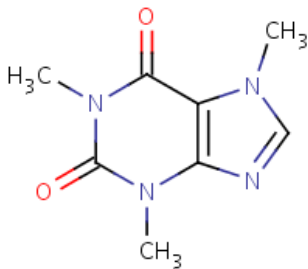
Examples: [iron*](#), [InChI=1S/CH4O/c1-2/h2H,1H3](#), [caffeine](#) ☐ ★★☆☆

Home Advanced Search Browse Documentation Download Tools About ChEBI Contact us Submit

ChEBI > Main

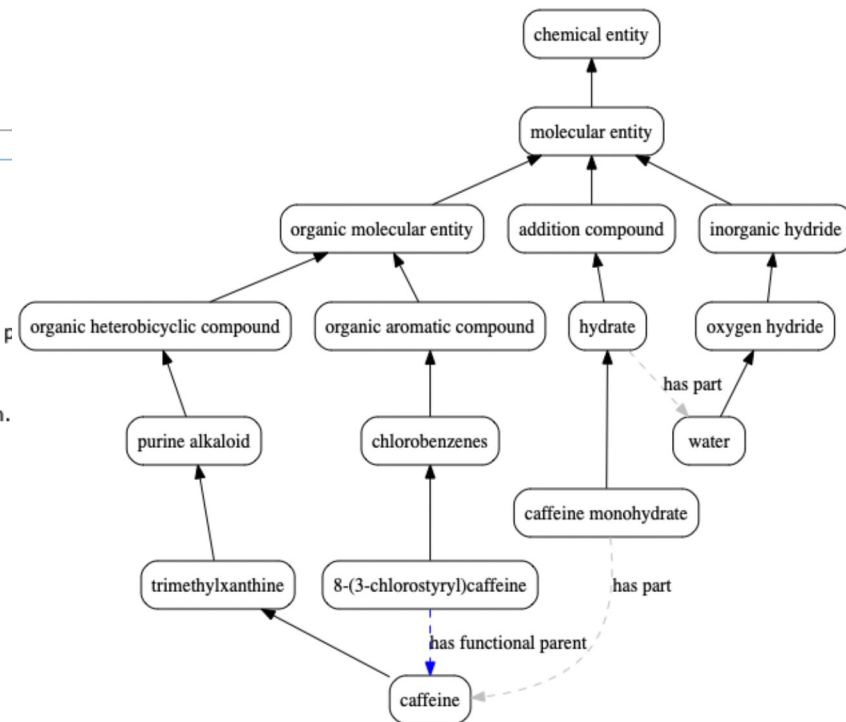
CHEBI:27732 - caffeine

Main ChEBI Ontology Automatic Xrefs Reactions Pathways Models



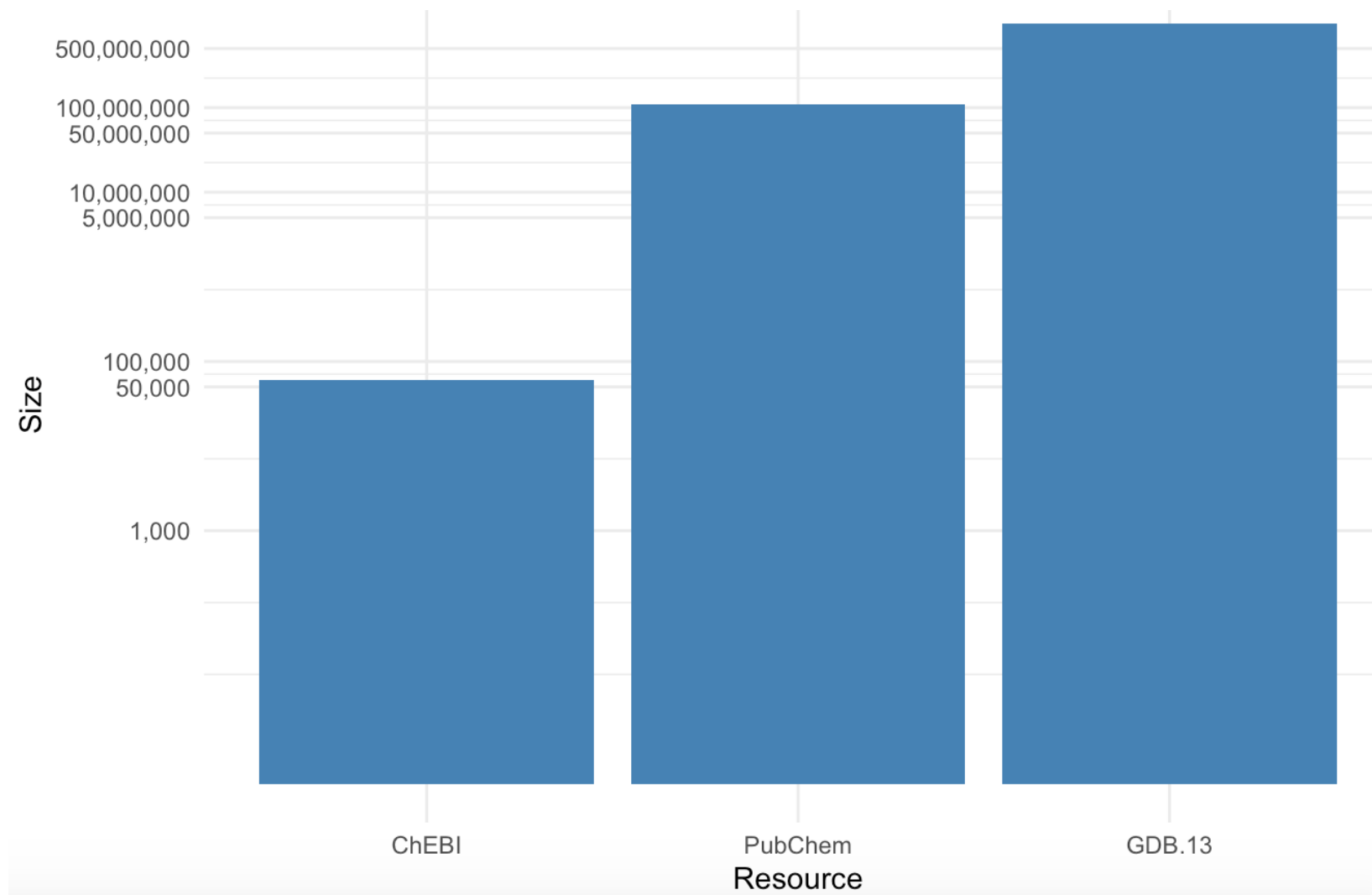
ChEBI Name	caffeine
ChEBI ID	CHEBI:27732
Definition	A trimethylxanthine in which the three methyl groups are located at p occurs naturally in tea and coffee.
Stars	★★★★ This entity has been manually annotated by the ChEBI Team.
Secondary ChEBI IDs	CHEBI:3295, CHEBI:41472, CHEBI:22982
Supplier Information	 eMolecules:493944 , eMolecules:27517656 , ZINC000000001084
Download	 Molfile XML SDF

SMILES Cn1cnc2n(C)c(=O)n(C)c(=O)c12

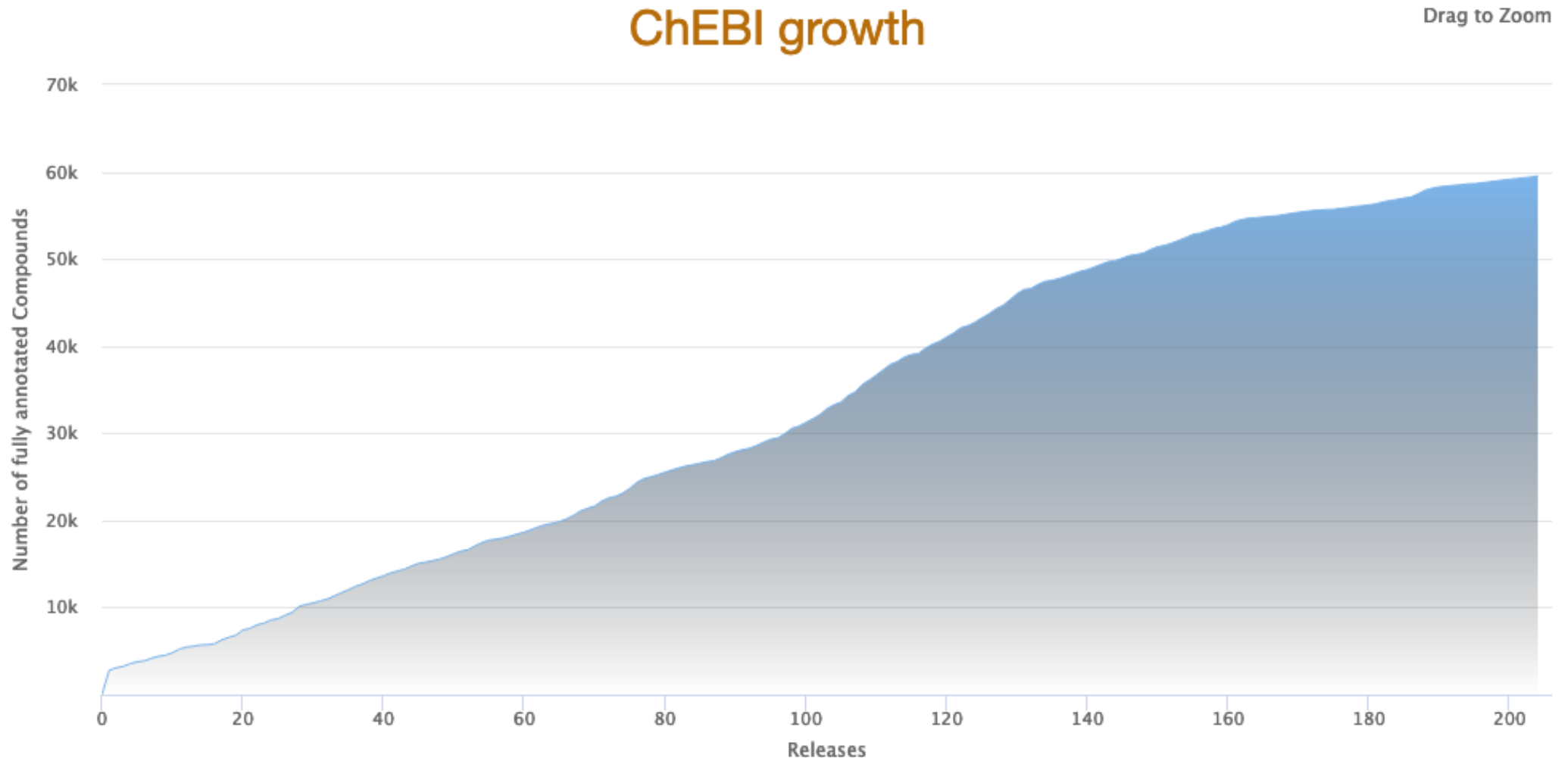


Hastings et al., *Nucleic Acids Research* 2016

Chemical “space” is huge (the numbers of possible molecules that might be included in the ontology in the future)



A problem looking for a solution: Manually maintained ontologies such as ChEBI ensure high quality, but scale poorly

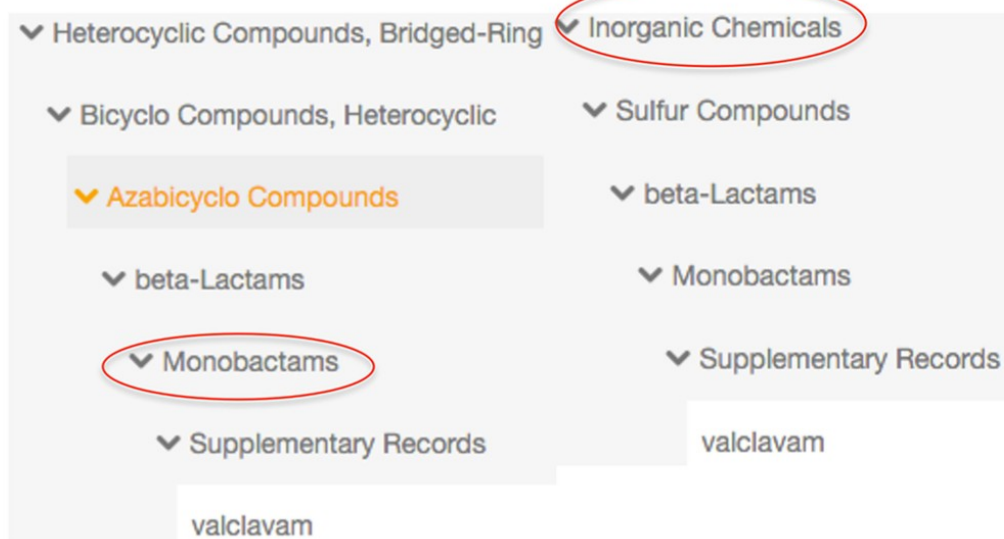


Reasons why automated text mining fails to replace curated ontologies

- Ontologies are community-**shared consensus** knowledge resource

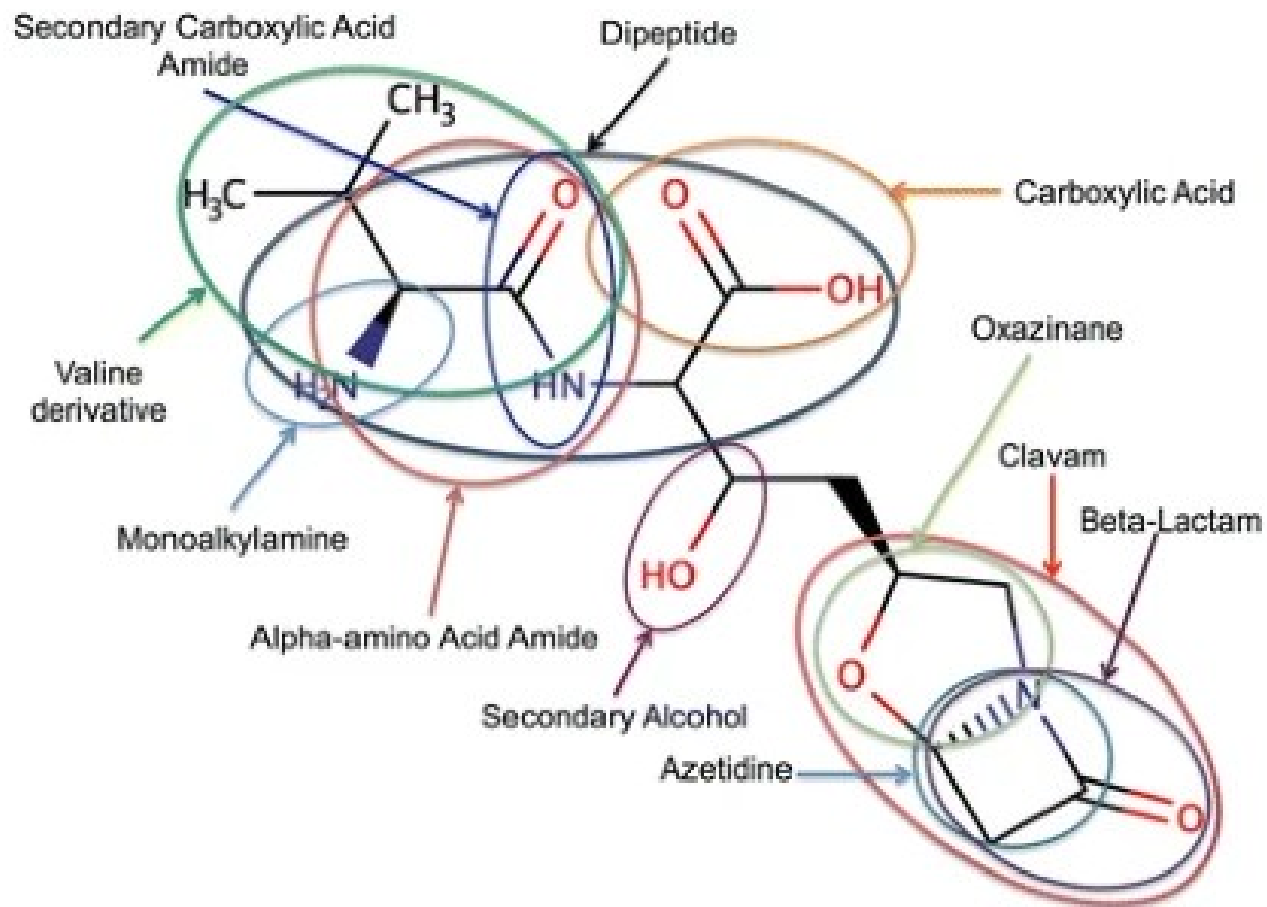


- Algorithmic approaches to cleaning up noise are extensively used in PubChem. Still, **PubChem** contains numerous **incorrectly assigned classes**, and classes are missing.



PubChem: Valclavam (CID126919)

ClassyFire: rule-based extension of chemical ontologies based on structural features of the molecules

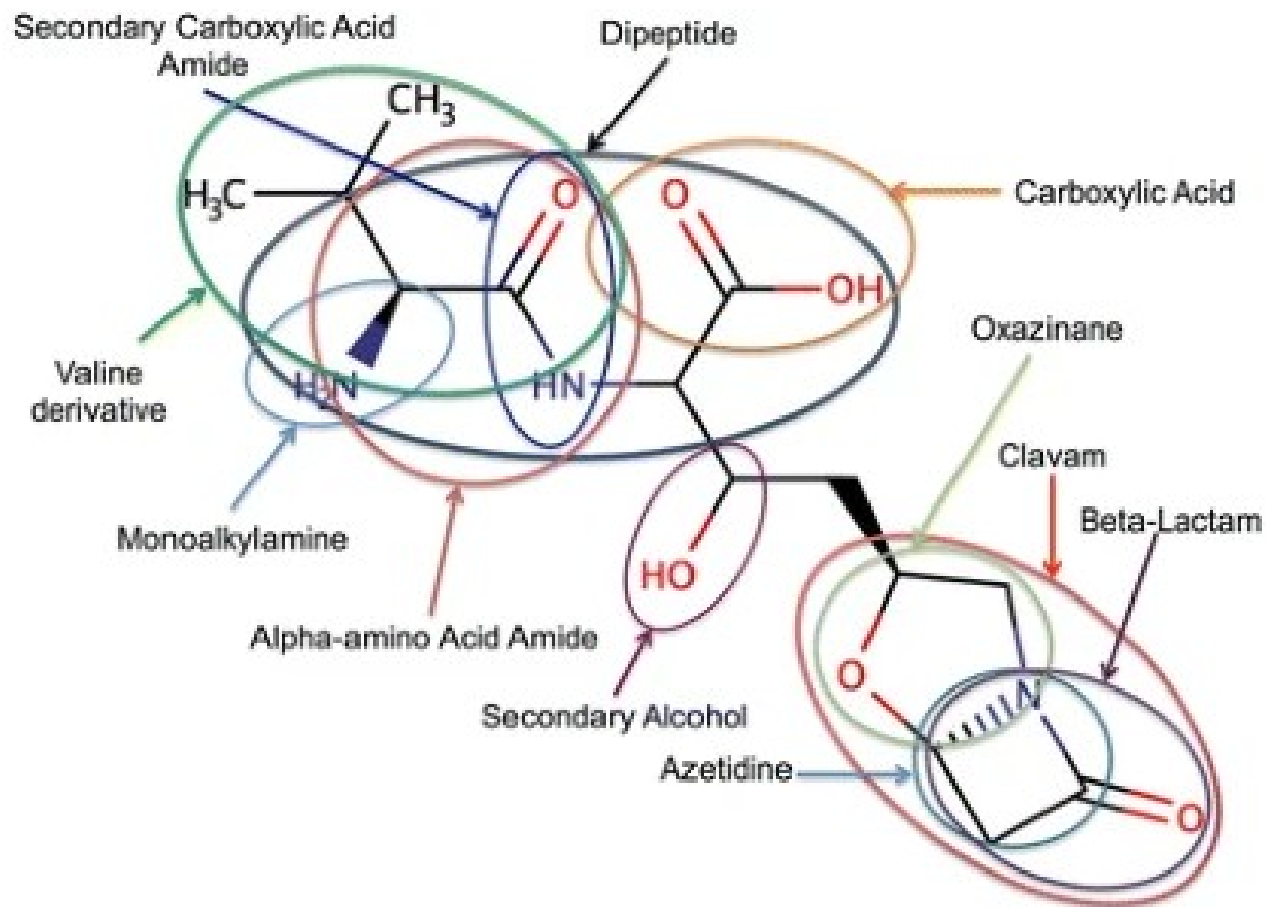


ClassyFire:

- 9,000 rules => algorithmic “black box”
- rule set is difficult to maintain
- ChemOnt is used, with 4,825 classes
- rules are not integrated with ChEBI's definitions of ontology classes

Djoumbou Feunang et al., “ClassyFire:...”, *Journal of Cheminformatics* 2016

ClassyFire: rule-based extension of chemical ontologies based on structural features of the molecules



ClassyFire:

- 9,000 rules => algorithmic “black box”
- rule set is difficult to maintain
- ChemOnt is used, with 4,825 classes
- rules are not integrated with ChEBI's definitions of ontology classes

Limitation of symbolic methods!

Djoumbou Feunang et al., “ClassyFire:...”, *Journal of Cheminformatics* 2016

Overview

1 Symbolic knowledge representation

2 Neuro-symbolic extension of ontologies

3 Transformers for more performance and interpretability

4 Ontology pre-training can improve Transformer performance

5 Outlook

The Gartner Hype Cycle for Artificial Intelligence, 2023



The Gartner Hype Cycle for Artificial Intelligence, 2023



Neuro-symbolic knowledge representation and classification



$$17 \times 24 = ?$$

Neuro-symbolic knowledge representation and classification

Subsymbolic ("Intuition")



Symbolic ("Ratio")

$$17 \times 24 = ?$$

Neuro-symbolic knowledge representation and classification

Subsymbolic ("Intuition")

- "System 1"
- Fast
- Data driven
- Adapt to noisy data
- Flexible
- Automatic



Symbolic ("Ratio")

- "System 2"
- Slow
- Knowledge driven
- Abstraction
- Reliable
- Explainable

$$17 \times 24 = ?$$

Neuro-symbolic knowledge representation and classification

Subsymbolic ("Intuition")

- "System 1"
- Fast
- Data driven
- Adapt to noisy data
- Flexible
- Automatic



**neuro-
symbolic**

Symbolic ("Ratio")

- "System 2"
- Slow
- Knowledge driven
- Abstraction
- Reliable
- Explainable

$17 \times 24 = ?$

Neuro-symbolic knowledge representation and classification

Subsymbolic ("Intuition")

- "System 1"
- Fast
- Data driven
- Adapt to noisy data
- Flexible
- Automatic



**neuro-
symbolic**

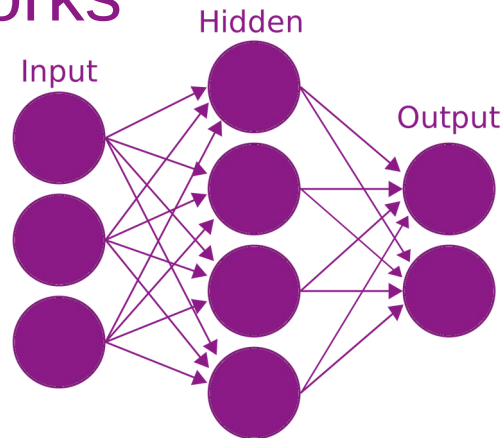
Symbolic ("Ratio")

- "System 2"
- Slow
- Knowledge driven
- Abstraction
- Reliable
- Explainable

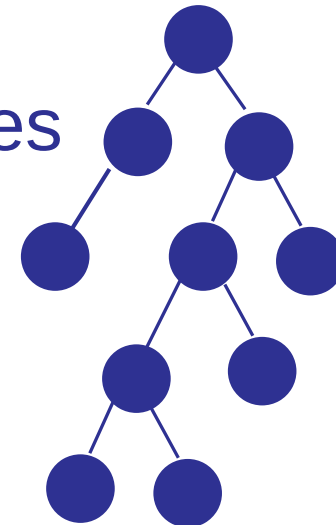
$$17 \times 24 = ?$$

here:

Neural networks



Ontologies

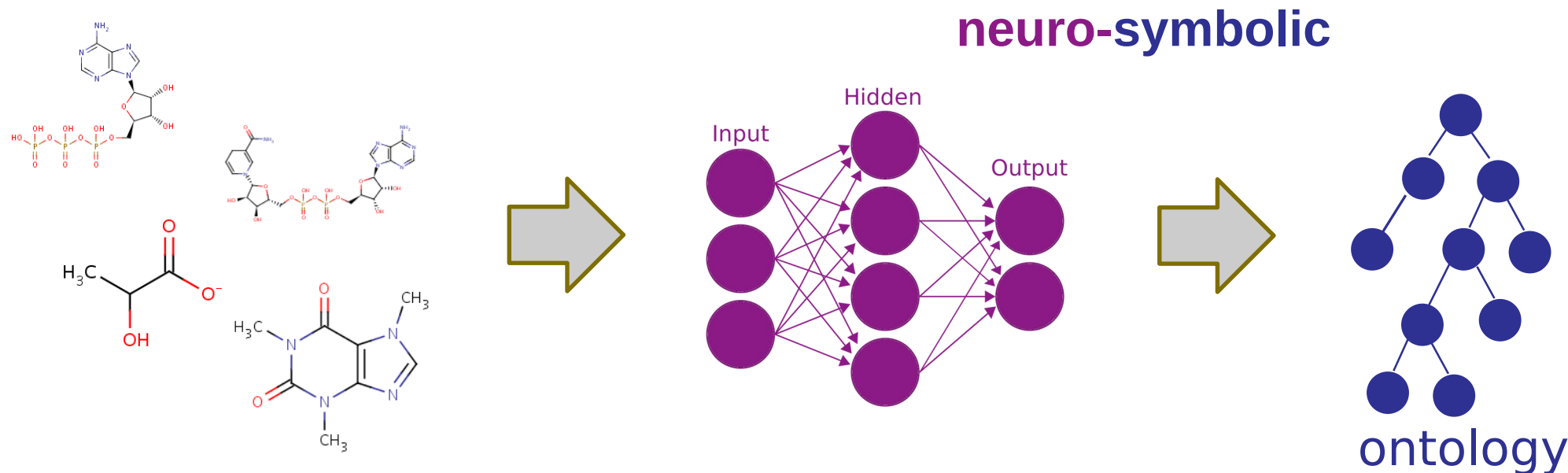


Research questions

- What are useful **neuro-symbolic** architectures for (grounded) knowledge representation?
- Can we beat purely **neural** and purely **symbolic** baselines?
- Can **neural networks** help with automated extension of **ontologies**?
- Can **ontologies** help improving the performance of **neural networks**?

neuro + symbolic ?

Can neural networks help with automated extension of ontologies?



Given the design decisions by the ontology developers:
how would they **extend** the ChEBI **ontology** to cover a **novel molecule**?

Knowing which classes a molecule belongs to helps to **predict** its **chemical behaviour** and uses, as well as enrich data and drive **discovery approaches**.

J. Hastings, M. Glauer, A. Memariani, F. Neuhaus, T. Mossakowski. “**Learning Chemistry: Evaluating machine learning for the task of structure-based chemical ontology classification**,” *Journal of Cheminformatics* 2021

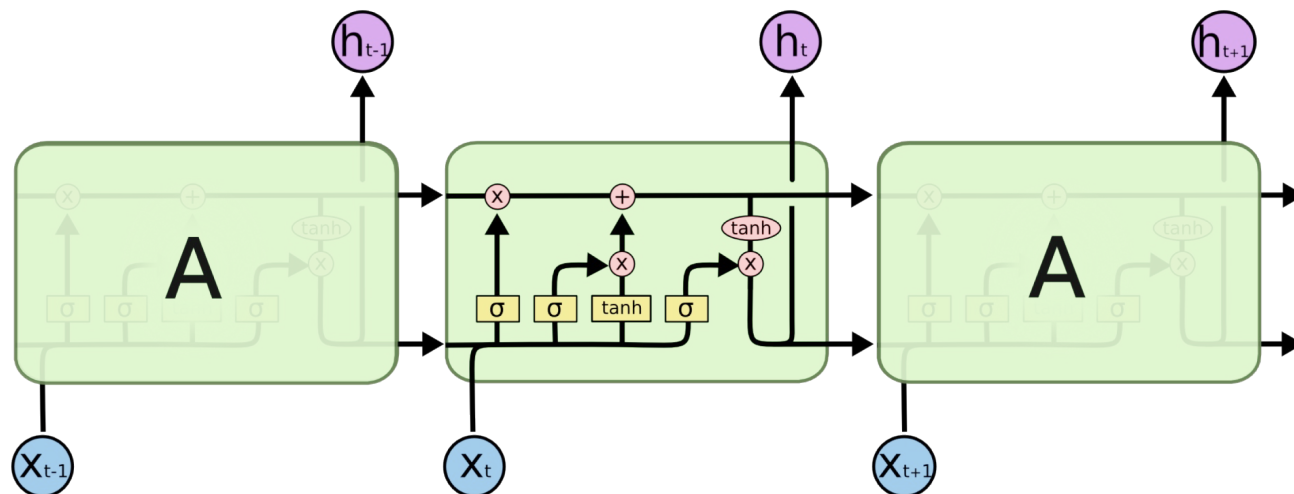
In initial exploratory research, we compared different ML algorithms

CLASSICAL

Logistic Regression
Random Forest
Naïve Bayes
Linear Discriminant Analysis
K-Nearest Neighbours
Support Vector Machine

(input: fixed-length chemical fingerprints)

DEEP

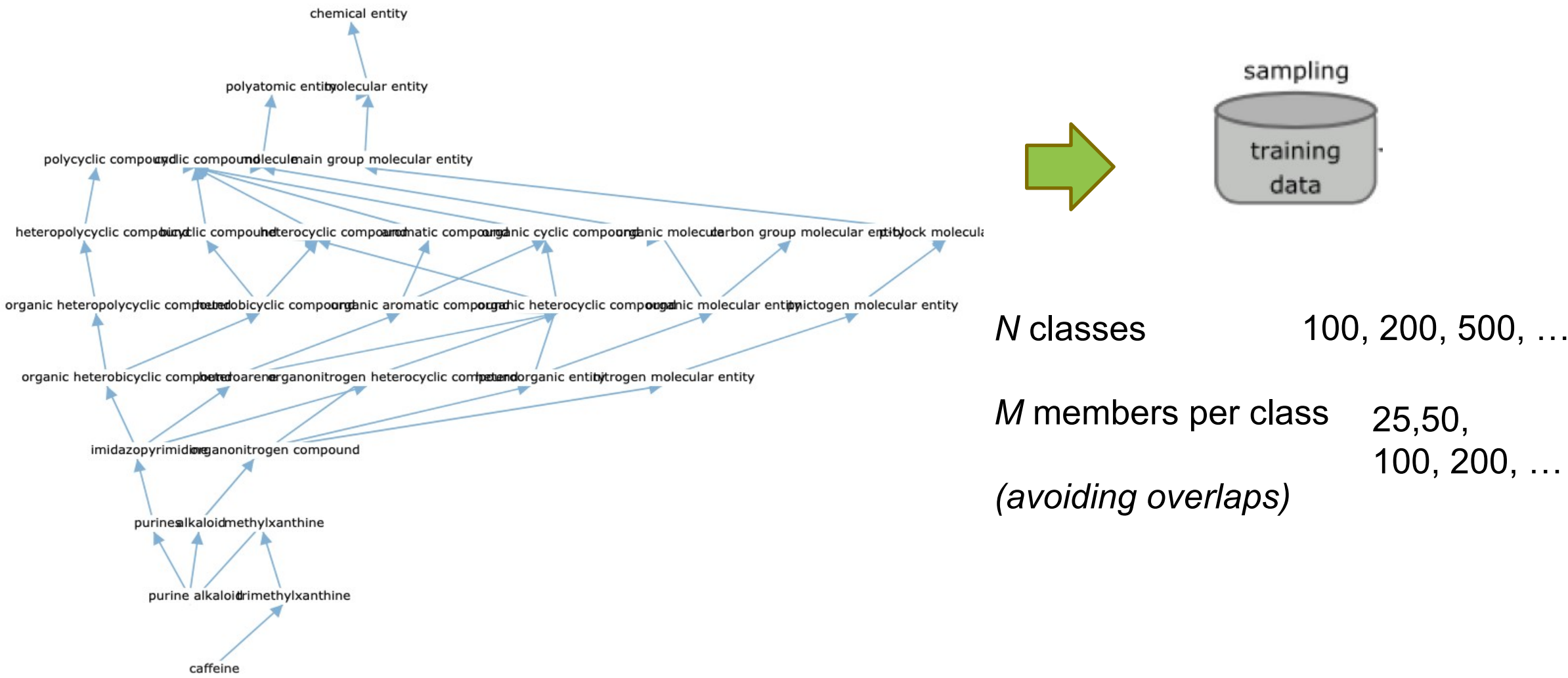


Long Short-Term Memory Network (LSTM)

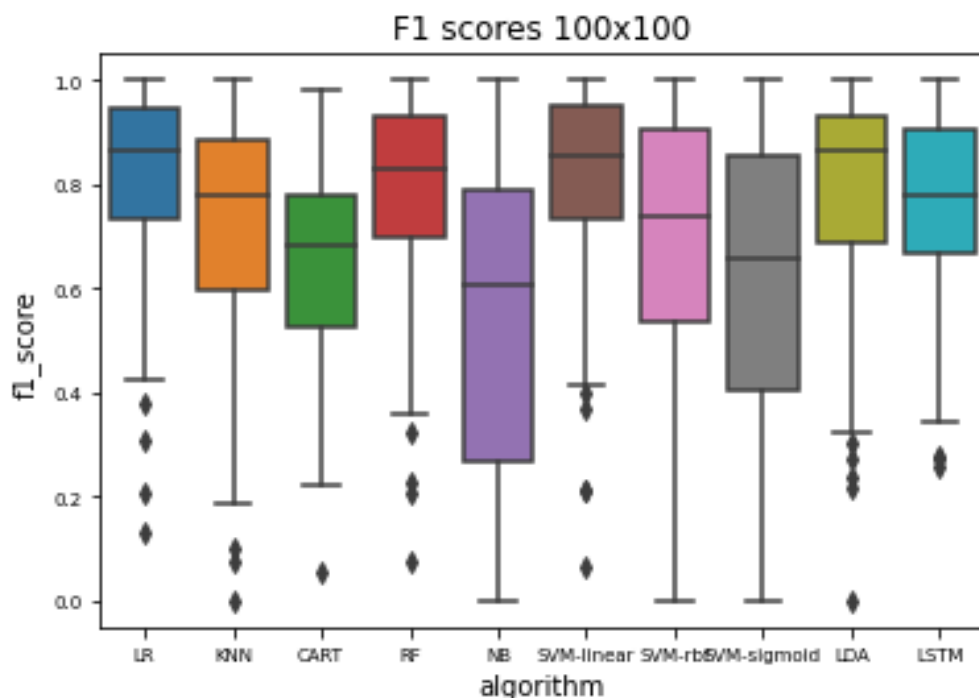
(input: SMILES)

J. Hastings, M. Glauer, A. Memariani, F. Neuhaus, T. Mossakowski. “**Learning Chemistry: Evaluating machine learning for the task of structure-based chemical ontology classification,**” *Journal of Cheminformatics* 2021

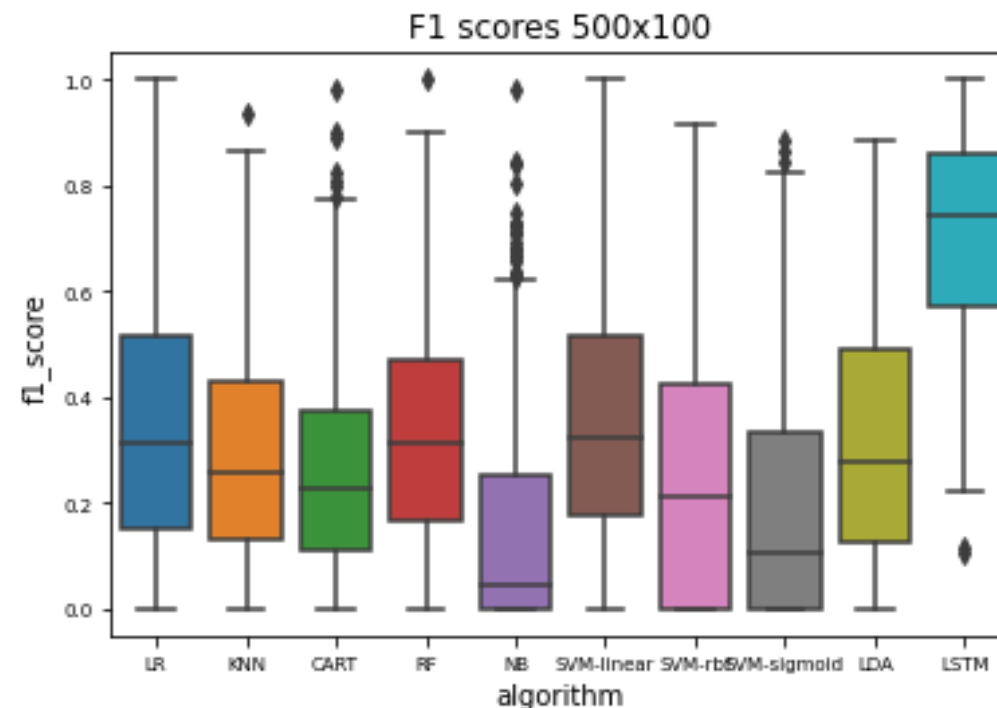
We used a sub-sampling approach to generate “learning-ready” datasets



Performance of various machine learning models



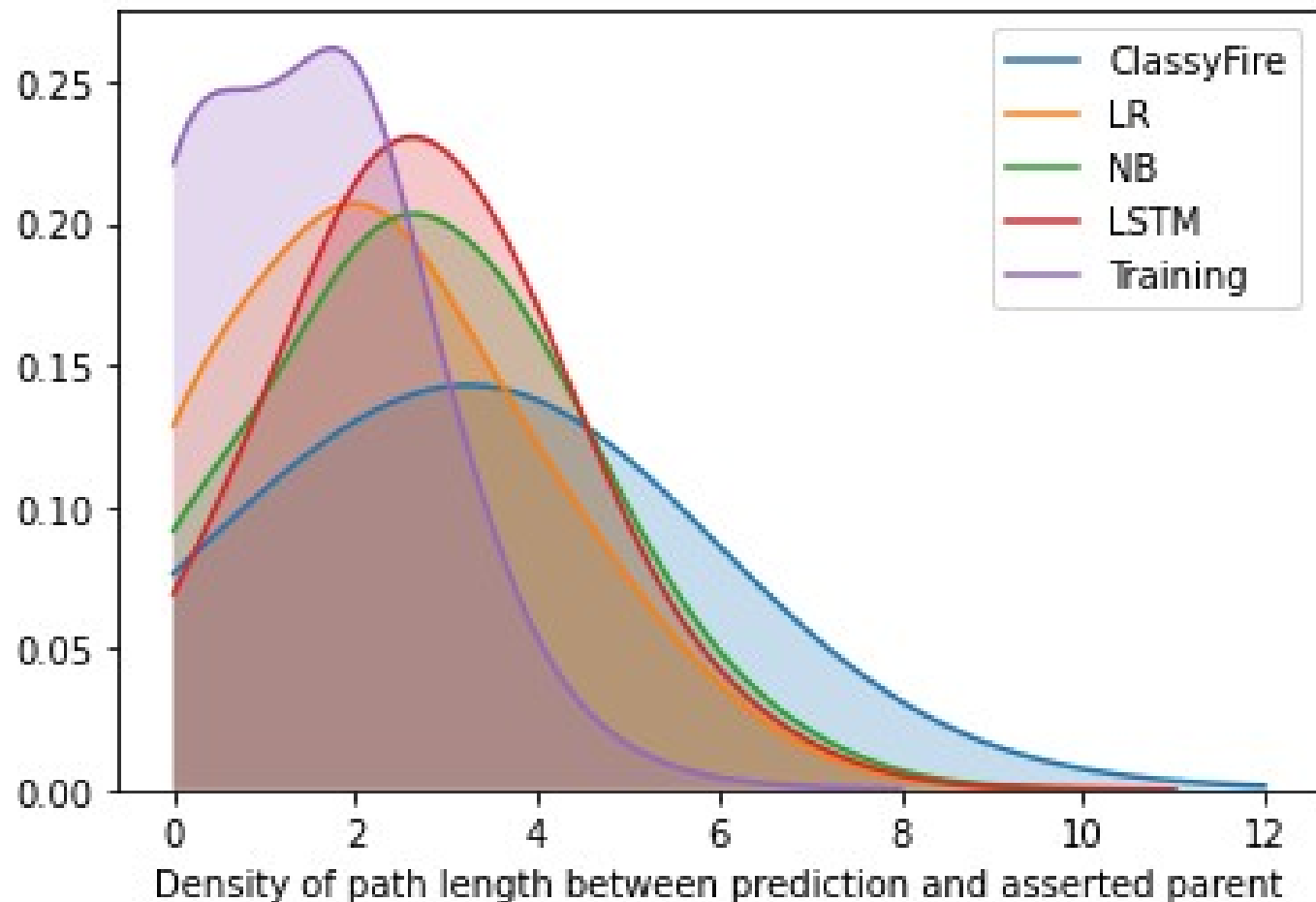
Restricted problems (subsampling)
“size” = number of classes
x number of members per class



LSTMs (a form of recurrent neural network)
are most robust to problem size

J. Hastings, M. Glauer, A. Memariani, F. Neuhaus, T. Mossakowski. “**Learning Chemistry: Evaluating machine learning for the task of structure-based chemical ontology classification,**” *Journal of Cheminformatics* 2021

Comparison to state of the art (ClassyFire)

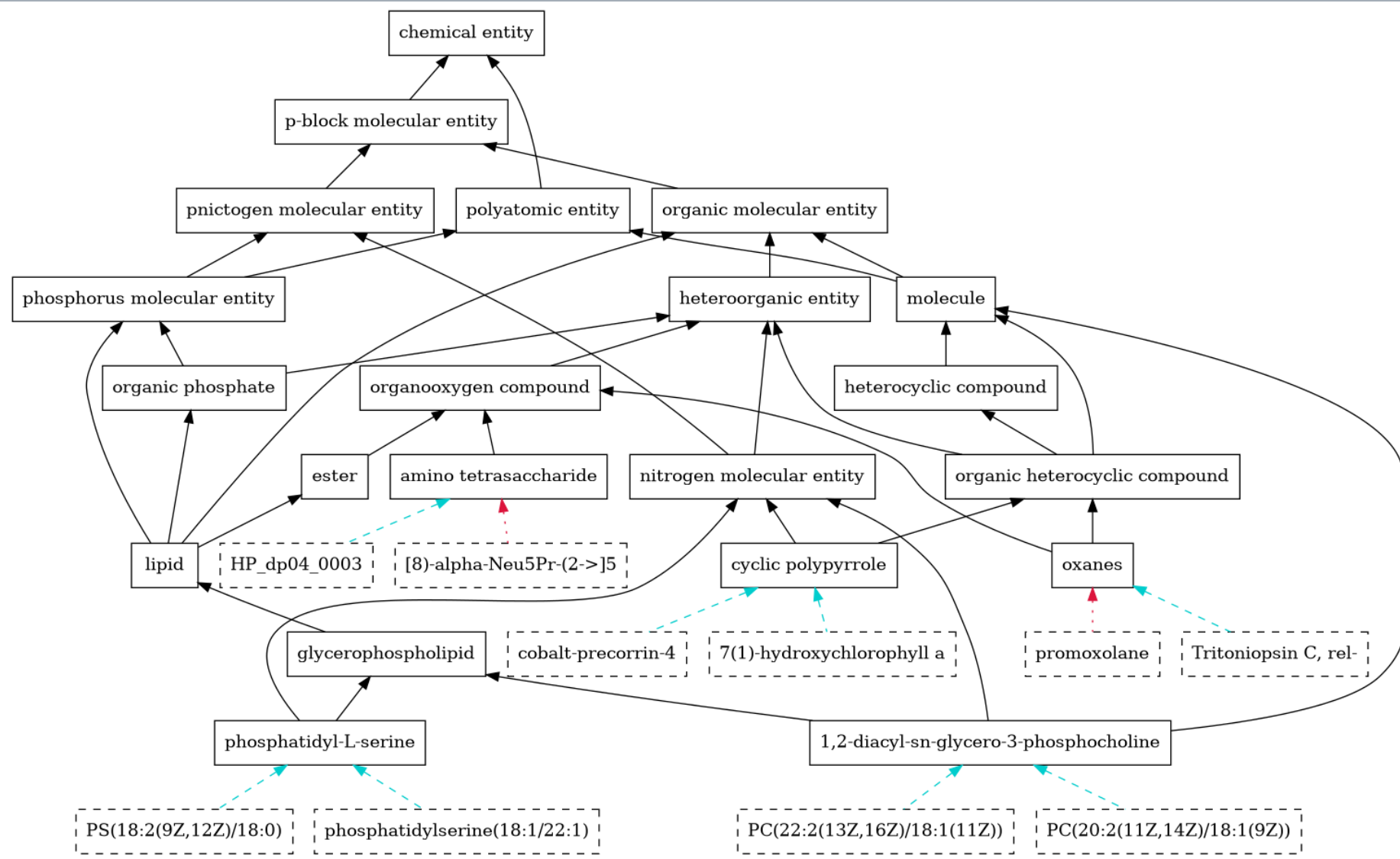


NOTES:

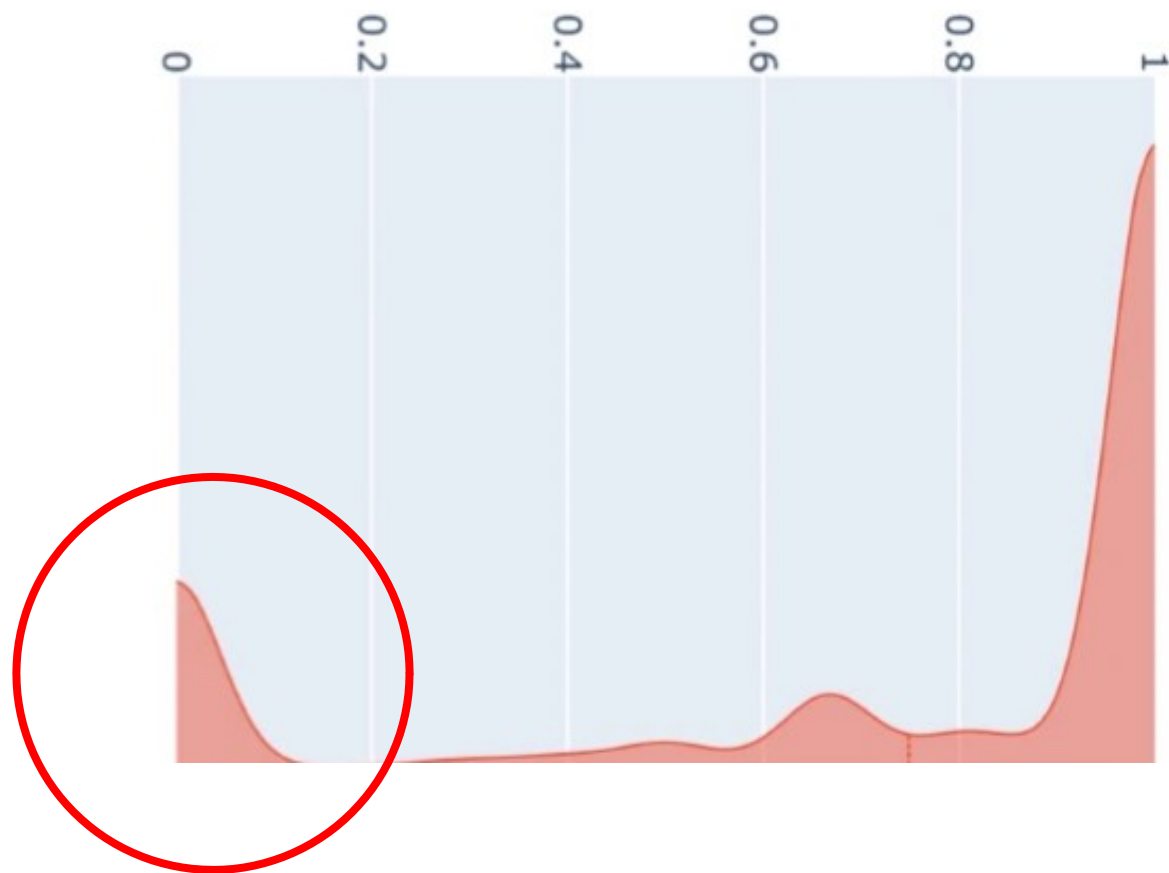
- a) ClassyFire mapping to ChEBI is “semantically lossy”
- b) ClassyFire already used in ChEBI maintenance

J. Hastings, M. Glauer, A. Memariani, F. Neuhaus, T. Mossakowski. “**Learning Chemistry: Evaluating machine learning for the task of structure-based chemical ontology classification,**” *Journal of Cheminformatics* 2021

Sample ontology extension



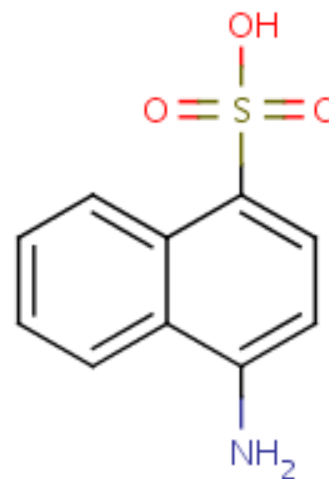
For some molecules, the deep neural network made no class prediction



“molecule-wise” average F1 scores

Classical ML approaches cannot learn the ontology hierarchy

The LSTM struggles with aromatic ring structures



Nc1ccc(c2ccccc12)S(=O)(=O)O
naphthionic acid

Overview

1 Symbolic knowledge representation

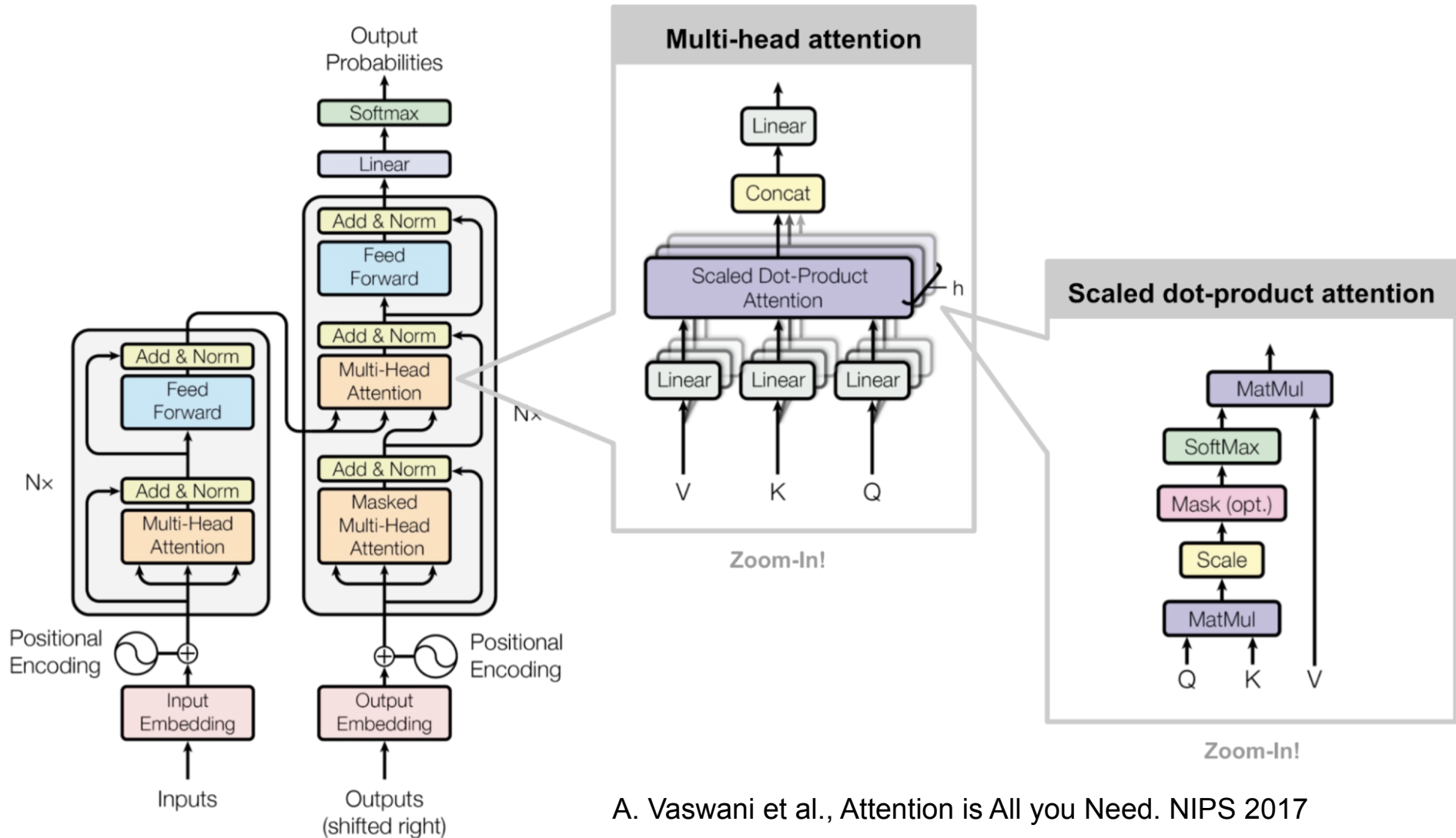
2 Neuro-symbolic extension of ontologies

3 Transformers for more performance and interpretability

4 Ontology pre-training can improve Transformer performance

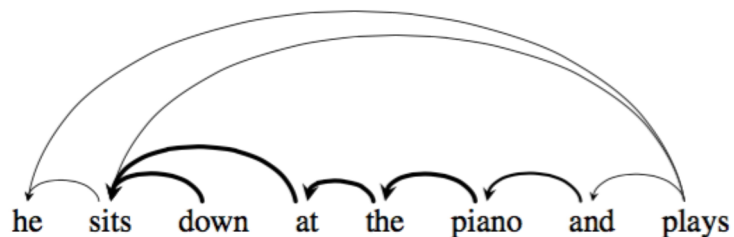
5 Outlook

Transformers revolutionised deep learning



A. Vaswani et al., Attention is All you Need. NIPS 2017

RoBERTa/ELECTRA transformer: self-attention, pre-training, fine-tuning



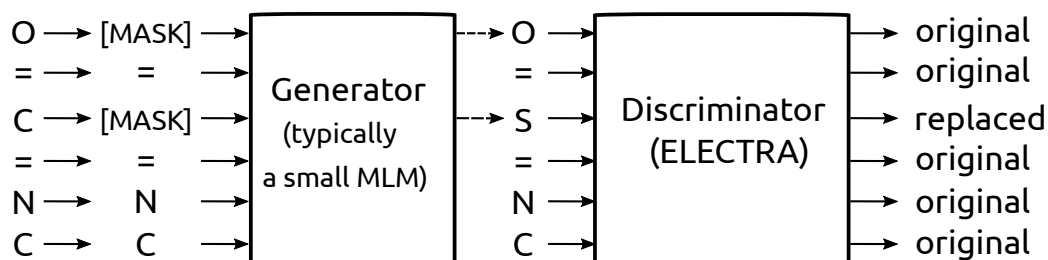
Self-attention in Transformer language models

BERT, RoBERTa:

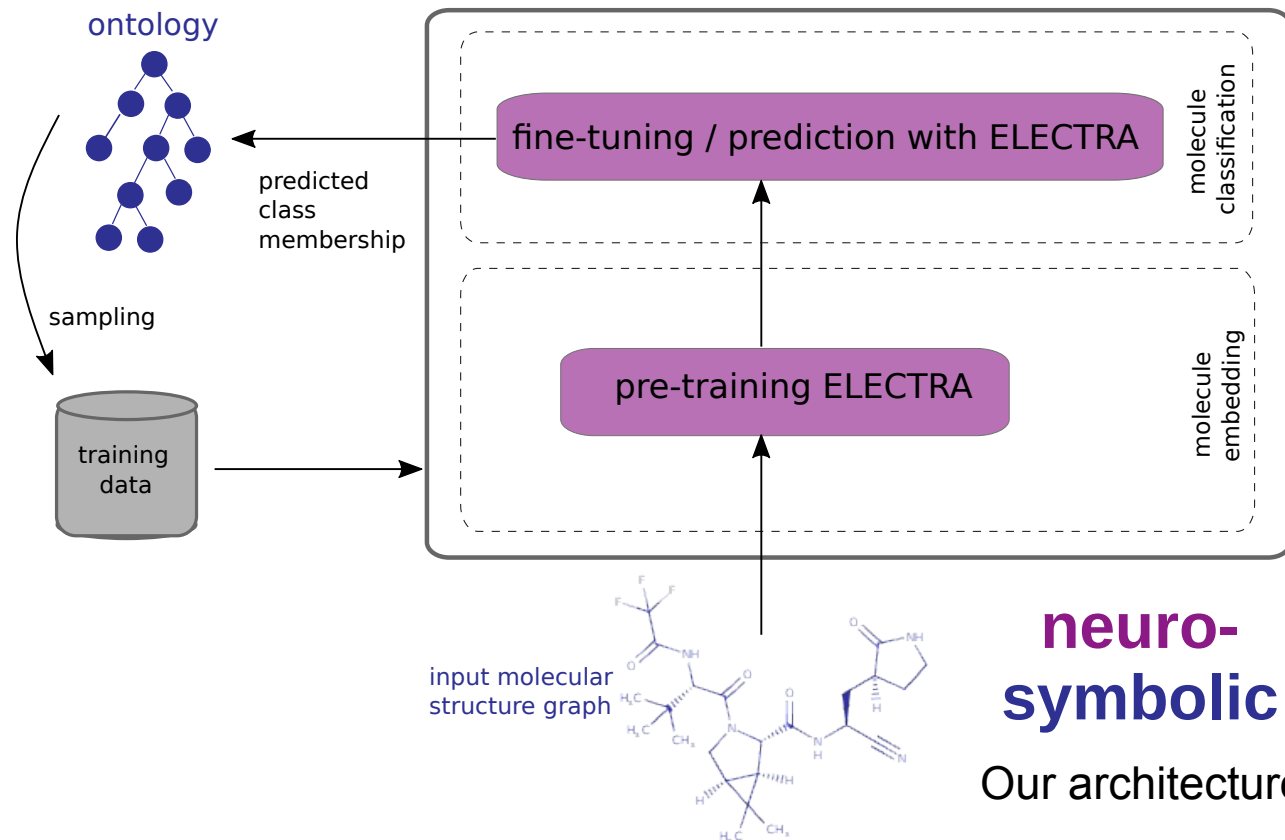
bi-directional attention, pre-training+fine-tuning

ELECTRA: pre-training using discrimination

(K. Clark et al., ICLR 2020)

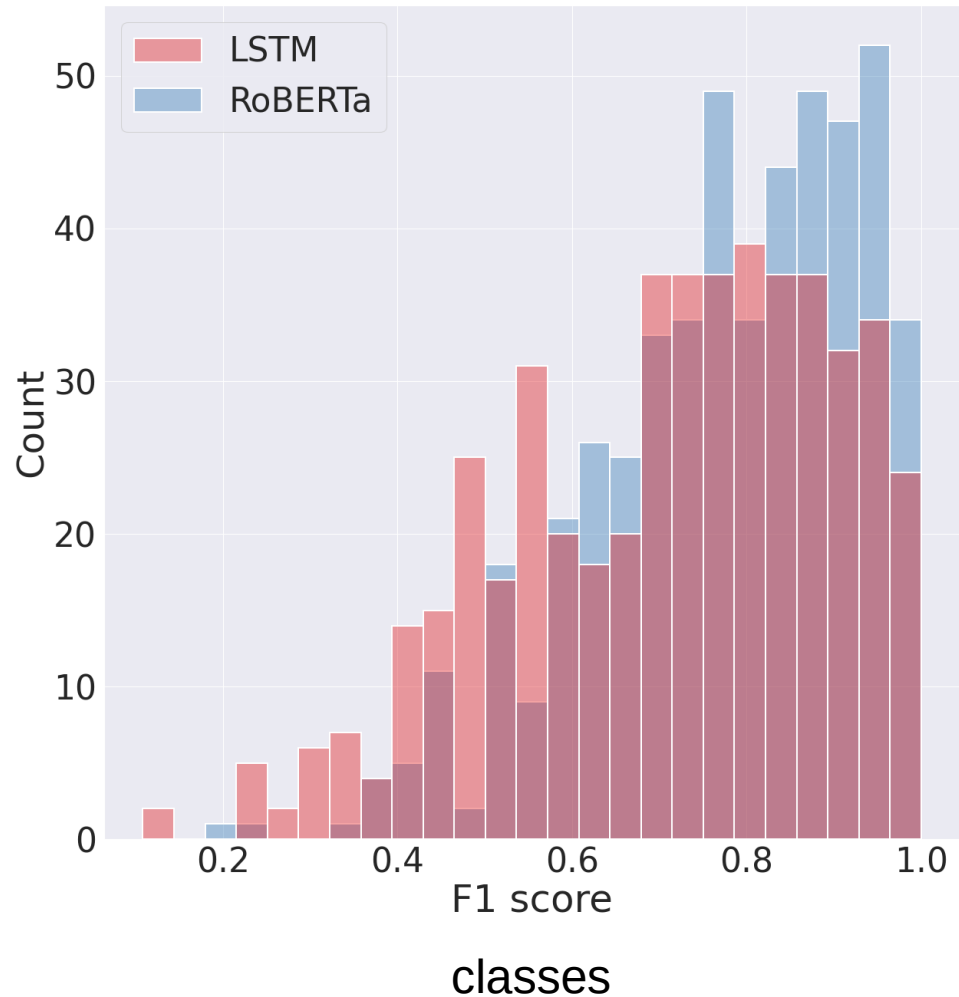
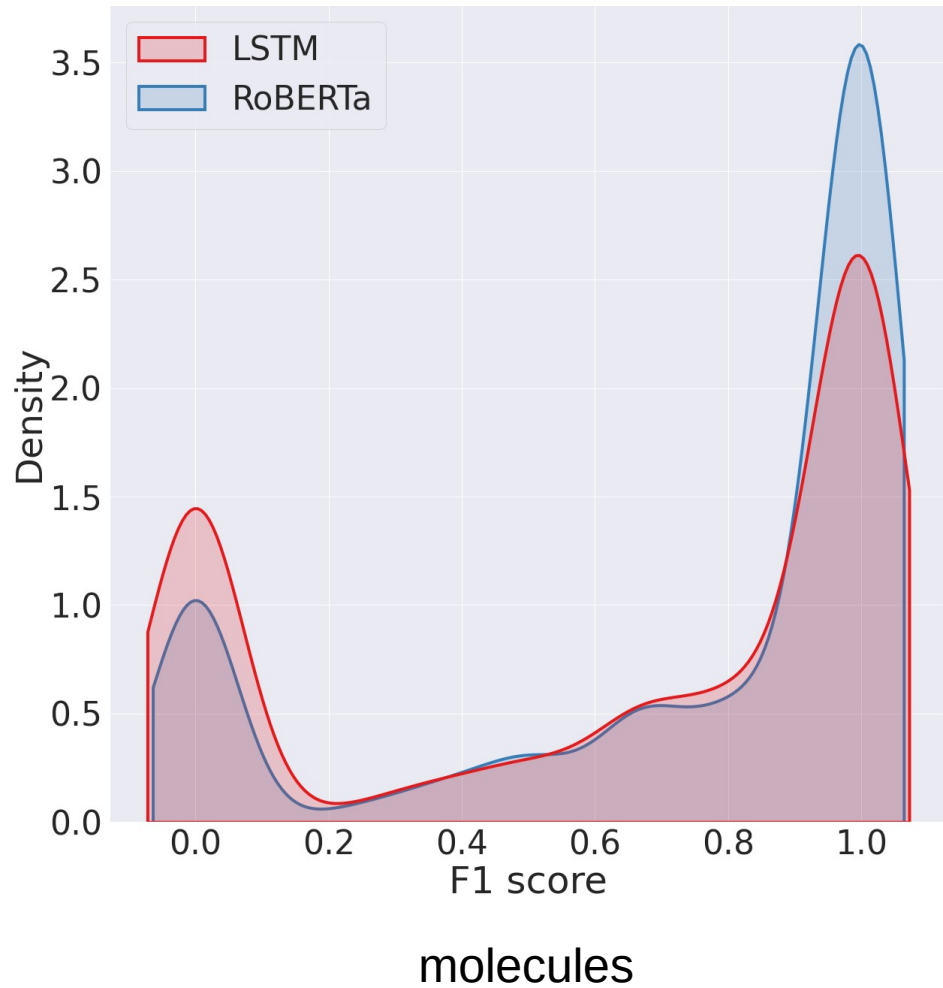


Our adaption of ELECTRA pre-training to SMILES strings



M. Glauer, A. Memariani, F. Neuhaus, T. Mossakowski, J. Hastings, Interpretable Ontology Extension in Chemistry, *SWJ* 2022

RoBERTa (and ELECTRA) outperforms the LSTM-based network



Newer ChEBI versions lead to better results

Reason: the newer version has more classes

	classes	Training set size	Micro F1 (all cl.)	Macro F1 (all cl.)	Macro F1 (common)
v148	709	80,639	0.8546	0.5133	0.5200
v200	854	129,187	0.9032	0.6070	0.6231

Table 2 Comparison of number of classes, training size and F1 scores with different aggregation methods for $ChEBI_{v148}^{709}$ and $ChEBI_{v200}^{854}$. The macro F1 score was calculated for all classes predicted by the model and, for comparison, also on the subset of 701 classes that appear in both ChEBI version. The best result for each score is highlighted in bold.

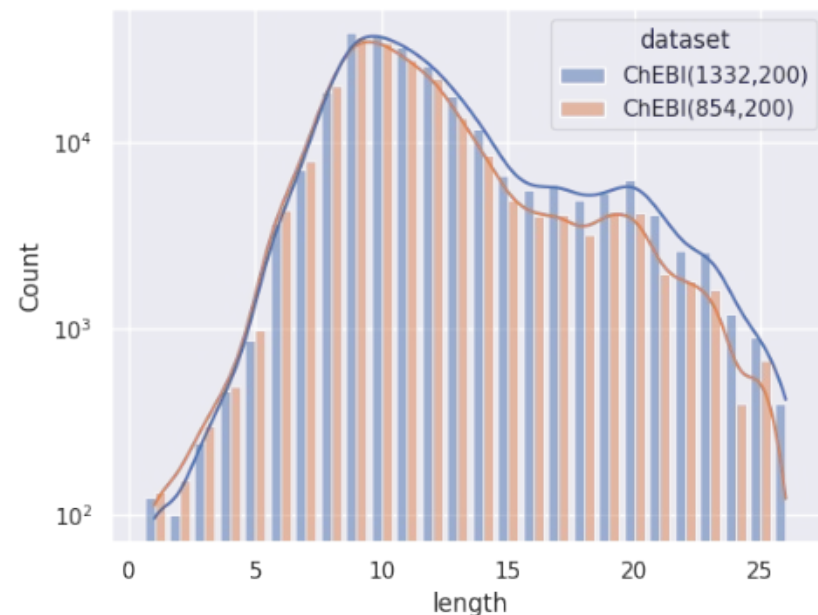


Fig. 8 Distribution of path lengths (log scale) amongst correct non-redundant predictions on molecular entities in the test set as achieved by the weighted models on both datasets

M. Glauer, F. Neuhaus, S. Flügel, M. Wosny, T. Mossakowski, A. Memariani, J. Schwerdt, J. Hastings:
Chebfier: Automatically classifying chemicals in ChEBI with semantically-enhanced AI.
Submitted for publication

Chebifier tool available online

Chebifier

Classify chemical structures using AI.

+ ADD SMILES ⬆️ UPLOAD FILE ➡️ PREDICT CLASSES ⬇️ DOWNLOAD JSON

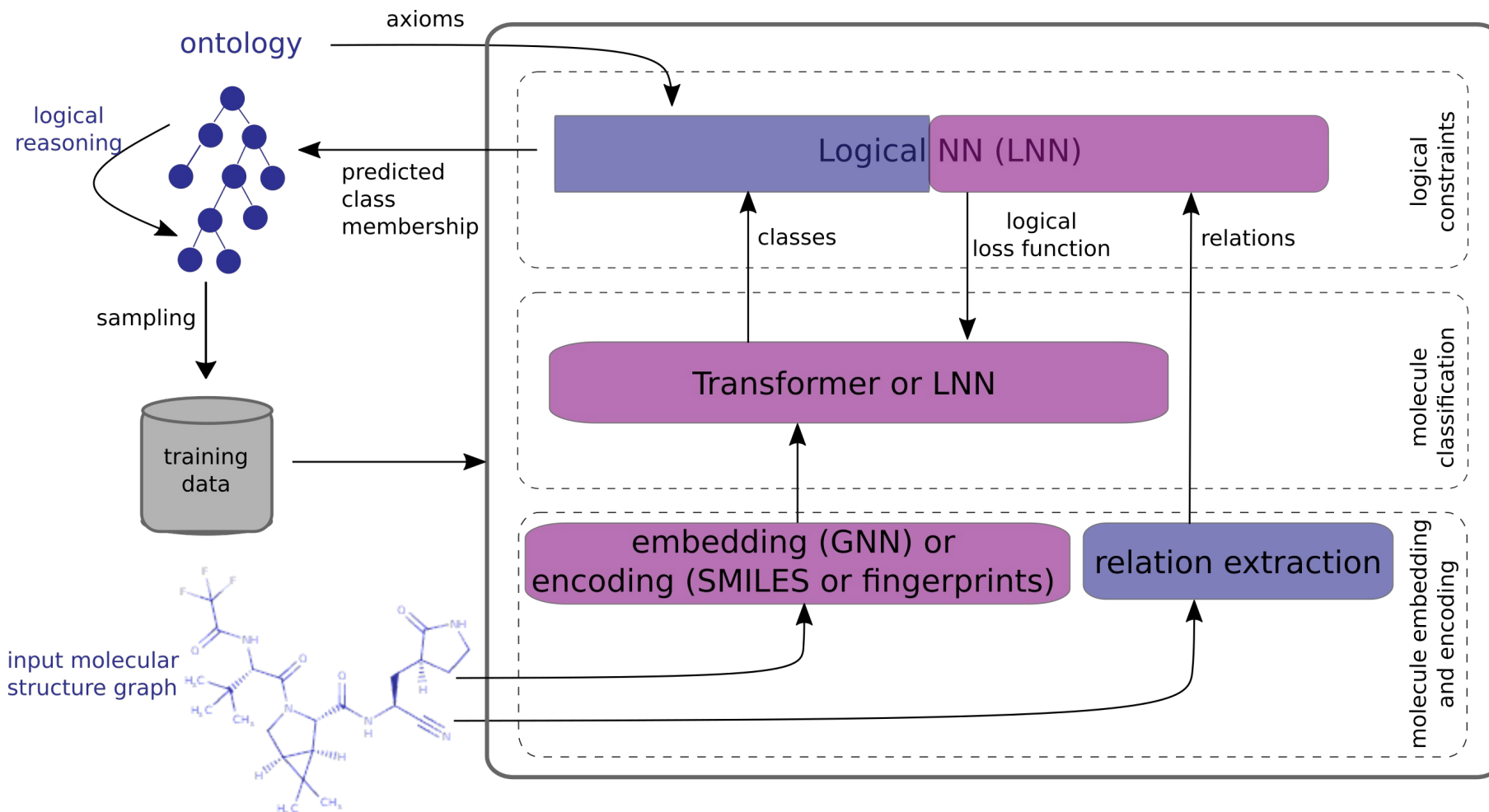
Smiles	Predicted Class	Actions
CCO	primary alcohol	  
C1=CC=C(C=C1)C=O	organic aromatic compound olefinic compound	  
CN1C2=C(C=C(C=C2)OC)SC1=NC(=O)/C=C/C3=CC=CC=C3	organosulfur heterocyclic compound organic heterobicyclic compound organonitrogen heterocyclic compound	  
COC1=C(C=CC(=C1)/C=C/C(=O)OCC(=O)NC2=CC(=C(C=C2)F)F)OC(F)F	benzenes	  
CCCCC(=O)C1=C(CC(CC1=O)(C)C)N2CCOCC2	organic molecule	  
CCOC(=O)C1=C(NC(=C1C)C(=O)/C=C\2/N(C(=O)/C(=C\3=CC=C(C=C3)[N+](=O)[O-])/S2)C4=CC=C(C=C4)C)/C#N)C	organonitrogen compound organochalcogen compound organic molecule	  
CC1=CC(=C(C=C1)C)NC(=O)CSC2=NN=C(N2CC3=CC=CC=...	anilide	  
CC1=C(C=CC(=C1)C)NC(=O)CSC2=NN=C(N2CC3=CC=CC=...	benzothiazole	  

Rows per page: 100 1–100 of 1469 < >

M. Glauer, F. Neuhaus, S. Flügel, M. Wosny, T. Mossakowski, A. Memariani, J. Schwerdt, J. Hastings:
Chebifier: Automatically classifying chemicals in ChEBI with semantically-enhanced AI.
Submitted for publication

<https://chebifier.hastingslab.org/>
<https://github.com/ChEB-AI>

DFG-SNSF project StrOntEx: Structured-entities Ontology Extension



M. Glauer, T. Mossakowski, F. Neuhaus, A. Memariani, J. Hastings: Neuro-Symbolic Semantic Learning for Chemistry. In P. Hitzler et al.: Compendium of Neurosymbolic Artificial Intelligence, IOS Press, 2023: 460-484

**neuro-
symbolic**

R. Riegel et al. Logical Neural Networks. CoRR abs/2006.13155 (2020)

Overview

1 Symbolic knowledge representation

2 Neuro-symbolic extension of ontologies

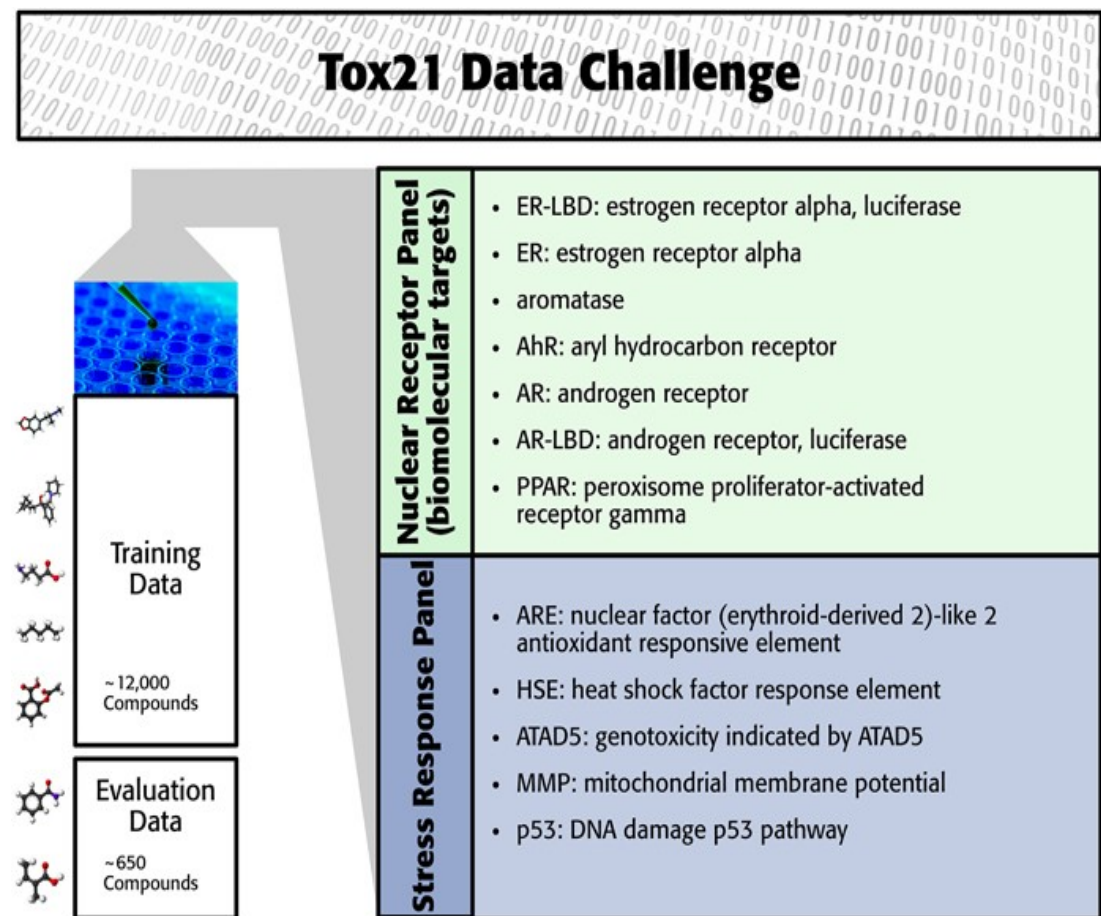
3 Transformers for more performance and interpretability

4 Ontology pre-training can improve Transformer performance

5 Outlook

Predicting toxicity from molecular structure: a hard task

Ontology pre-training as a method to embed prior knowledge

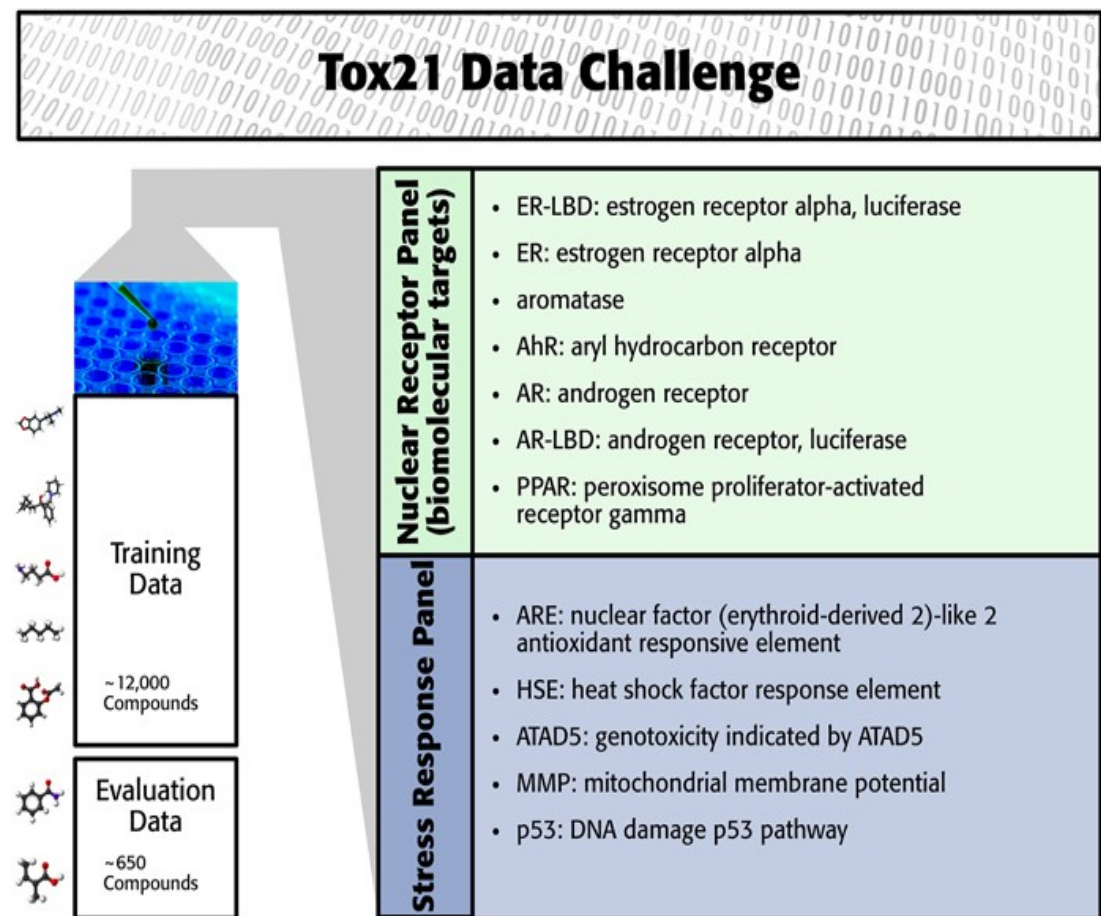


Can **ontology knowledge** improve the predictive quality of **neural networks**?

Martin Glauer, Fabian Neuhaus, Till Mossakowski, Janna Hastings:
Ontology Pre-training for Poison Prediction. KI 2023: 31-45. Best paper award

Predicting toxicity from molecular structure: a hard task

Ontology pre-training as a method to embed prior knowledge



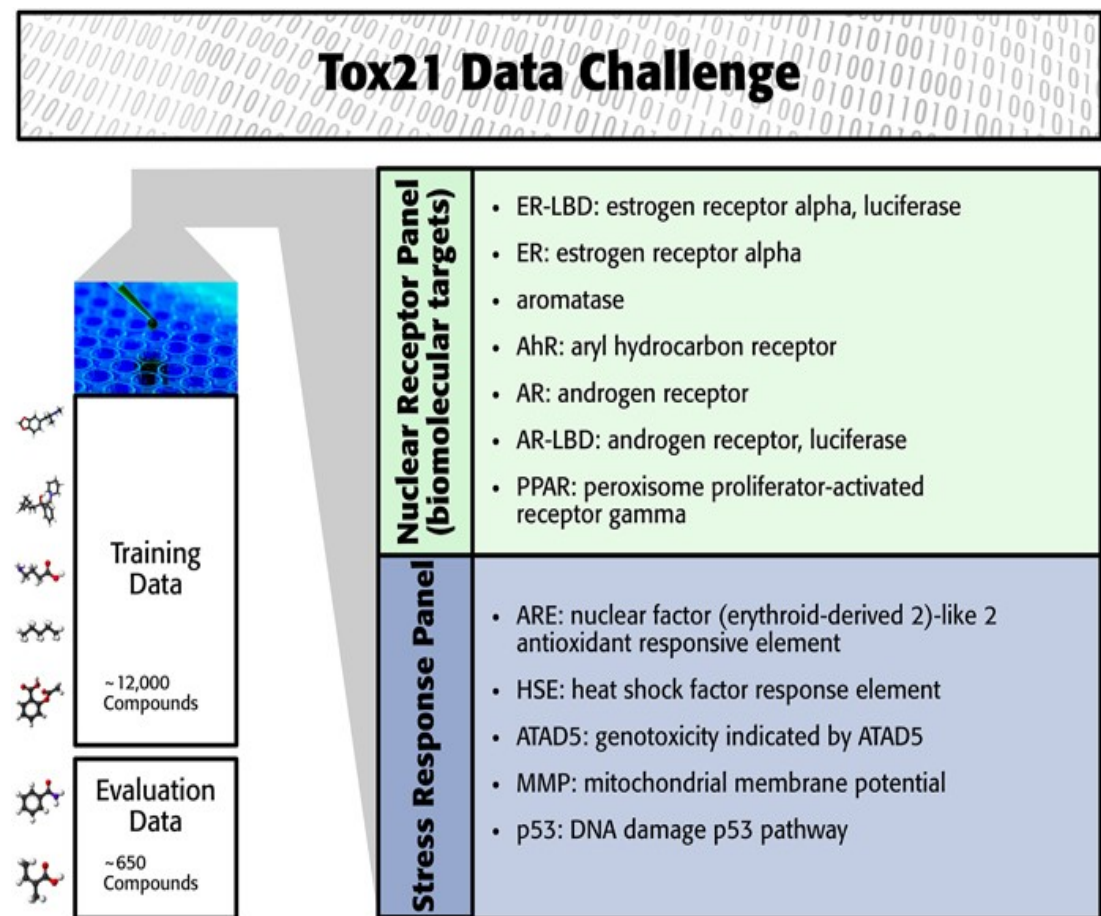
Can **ontology knowledge** improve the predictive quality of **neural networks**?

Ontology can provide **important features and distinctions** (e.g. organic versus inorganic) to **neural network**.

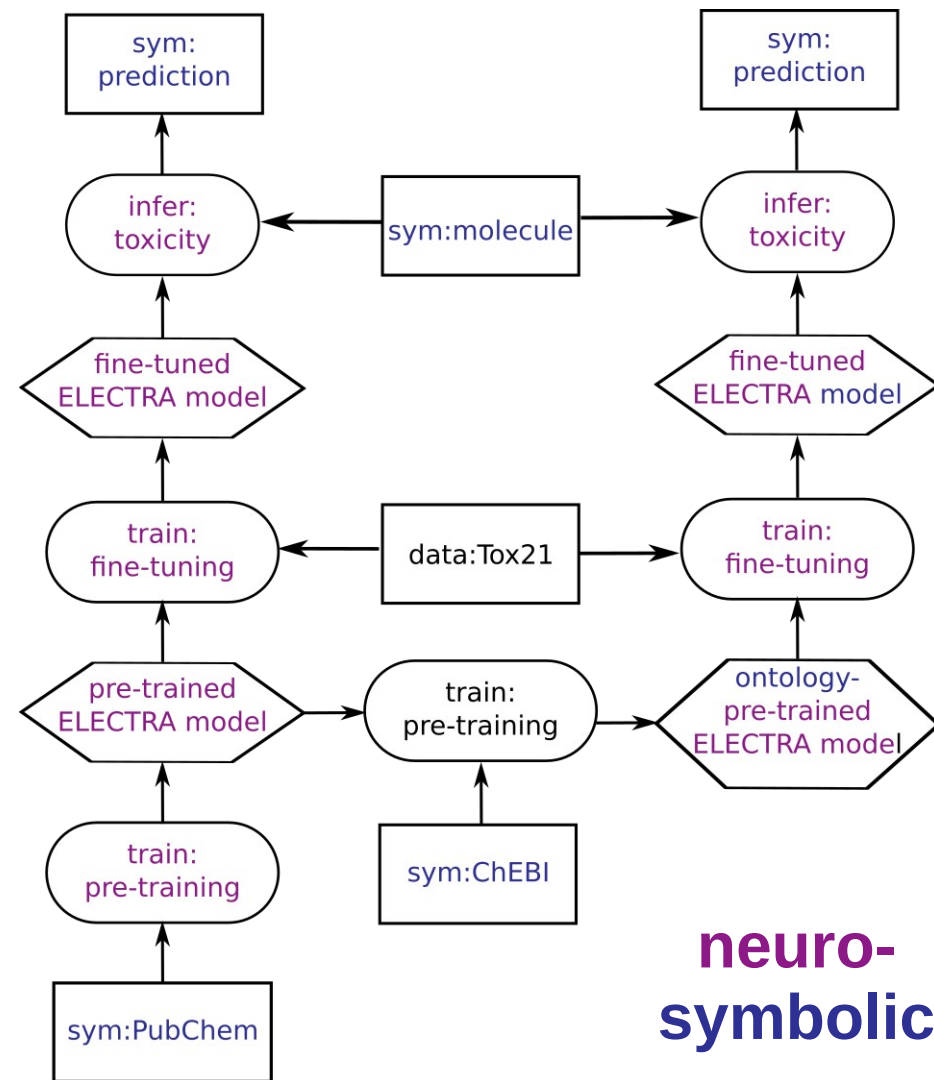
Martin Glauer, Fabian Neuhaus, Till Mossakowski, Janna Hastings:
Ontology Pre-training for Poison Prediction. KI 2023: 31-45. Best paper award

Predicting toxicity from molecular structure: a hard task

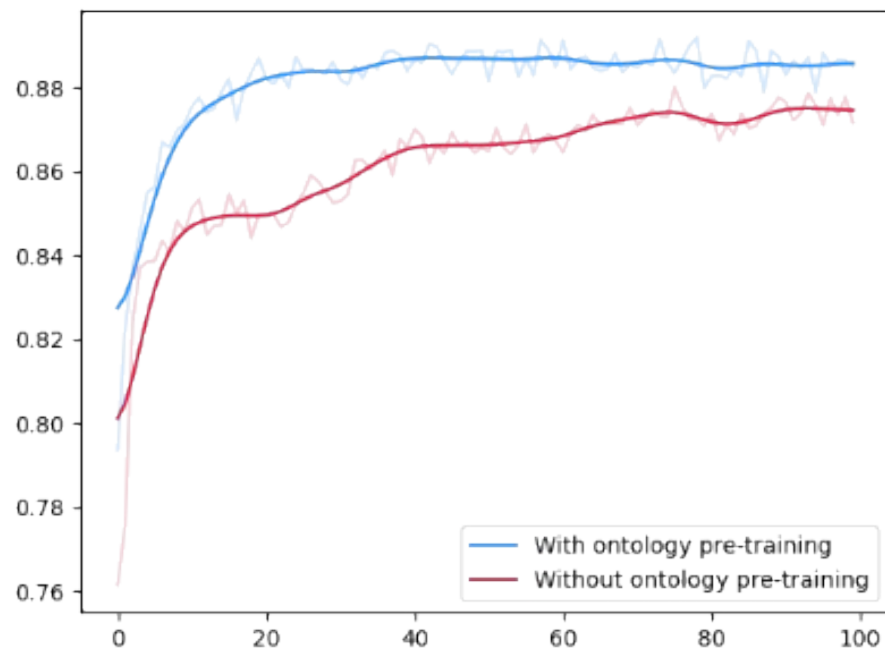
Ontology pre-training as a method to embed prior knowledge



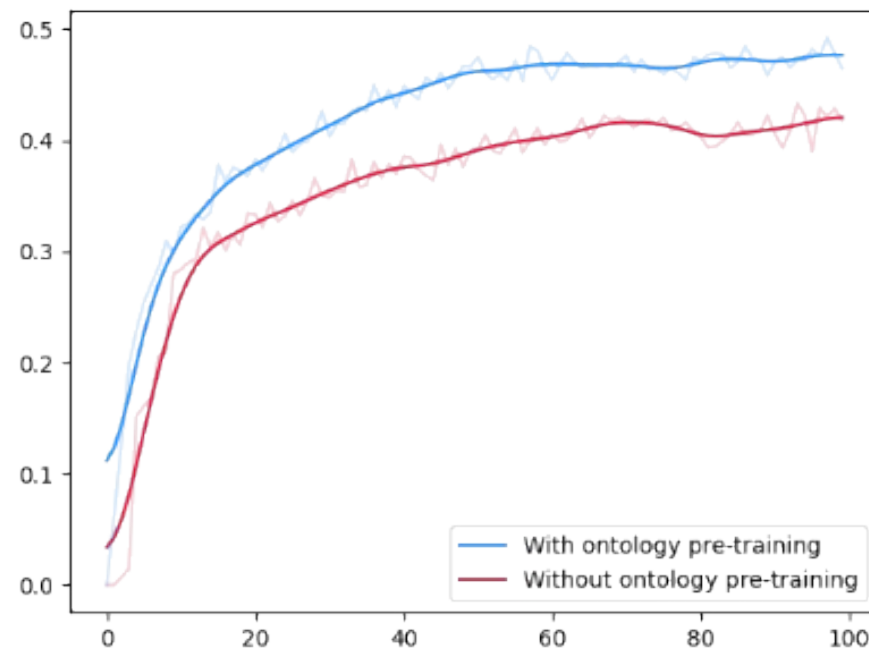
Martin Glauer, Fabian Neuhaus, Till Mossakowski, Janna Hastings:
Ontology Pre-training for Poison Prediction. KI 2023: 31-45. Best paper award



Ontology pre-training improved predictive performance and reduced training time



(a) ROC-AUC / MoleculeNet Tox21 dataset



(b) F1 score (micro) / MoleculeNet Tox21 dataset

Martin Glauer, Fabian Neuhaus, Till Mossakowski, Janna Hastings: Ontology Pre-training for Poison Prediction. KI 2023: 31-45. Best paper award

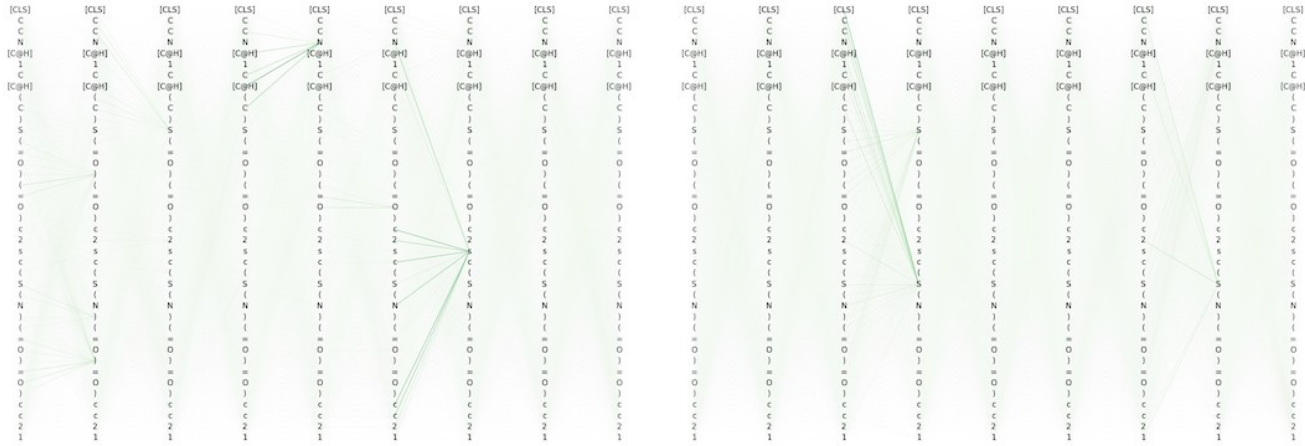
Our model outperformed the state of the art without feature augmentation

Dataset	Tox 21 (MoleculeNet)					Tox21 (Challenge)			
Metric	F1		ROC-AUC			F1		ROC-AUC	
Model	Our Model				SSL-GCN	Our Model			
Ontology Pre-training	yes	no	yes	no	-	yes	no	yes	no
NR-AR	0.41	0.52	0.82	0.76	0.80	0.1	0.14	0.63	0.62
NR-AR-LBD	0.51	0.5	0.85	0.77	0.76	0.05	0.1	0.69	0.67
NR-AhR	0.53	0.45	0.81	0.82	0.83	0.23	0.05	0.8	0.69
NR-Aromatase	0.33	0.15	0.84	0.8	0.73	0.25	0.04	0.75	0.69
NR-ER	0.44	0.4	0.74	0.71	0.72	0.16	0.09	0.64	0.62
NR-ER-LBD	0.37	0.3	0.84	0.76	0.69	0.14	0.12	0.66	0.63
NR-PPAR-gamma	0.29	-	0.84	0.83	0.76	0.14	-	0.67	0.66
SR-ARE	0.48	0.53	0.8	0.84	0.73	0.37	0.23	0.71	0.69
SR-ATAD5	0.14	0.19	0.75	0.74	0.72	0.16	-	0.65	0.65
SR-HSE	0.24	0.22	0.82	0.82	0.78	0.13	0.09	0.76	0.68
SR-MMP	0.62	0.53	0.9	0.88	0.81	0.48	0.21	0.86	0.82
SR-p53	0.39	0.35	0.83	0.8	0.75	0.3	-	0.82	0.78

Table 2: Class-wise scores on the test set on both Tox21 datasets. Bold values denote the best value for a particular combination of dataset and metric. NR - nuclear receptor; AR - androgen receptor; LBD - luciferase; AhR - aryl hydrocarbon receptor; ER - estrogen receptor; PPAR - peroxisome proliferator-activated receptor; SR - stress response; ARE - nuclear factor antioxidant response; ATAD5 - genotoxicity; HSE - heat shock factor response; MMP - mitochondrial response; p53 - DNA damage response.

Martin Glauer, Fabian Neuhaus, Till Mossakowski, Janna Hastings: Ontology Pre-training for Poison Prediction. KI 2023: 31-45. Best paper award

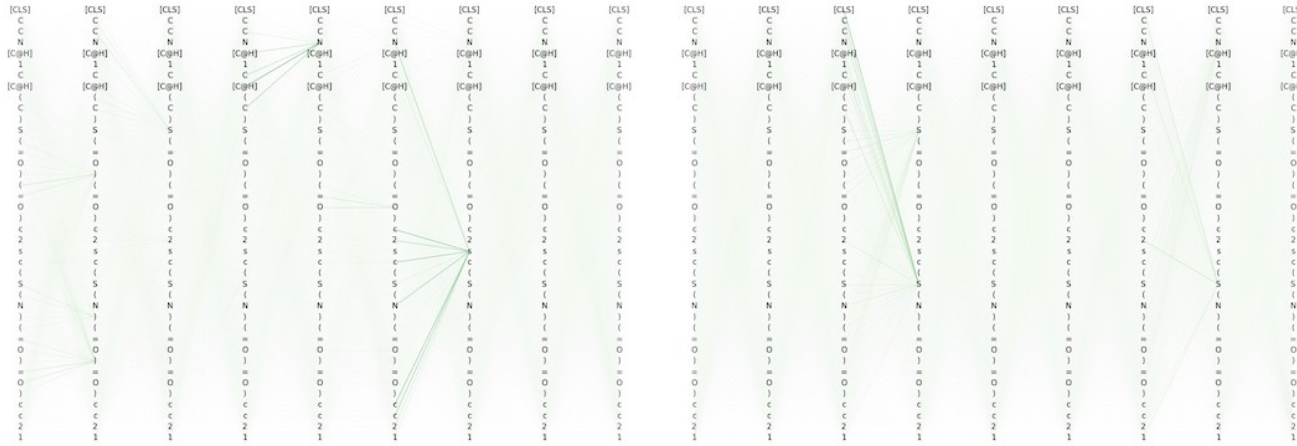
Pre-training improved the interpretability of attention weights



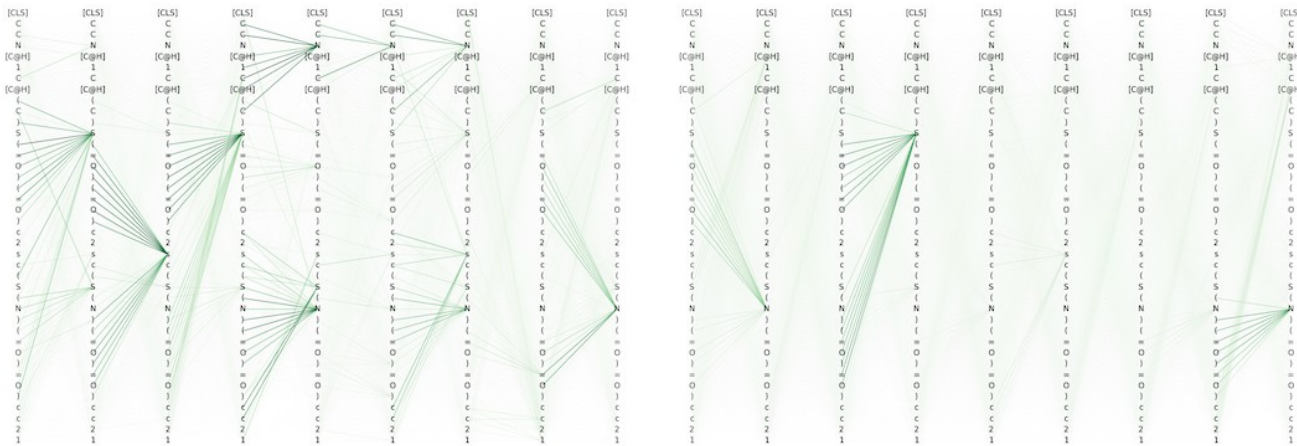
attention without ontology pre-training

Martin Glauer, Fabian Neuhaus, Till Mossakowski, Janna Hastings:
Ontology Pre-training for Poison Prediction. KI 2023: 31-45. Best paper award

Pre-training improved the interpretability of attention weights



attention without ontology pre-training

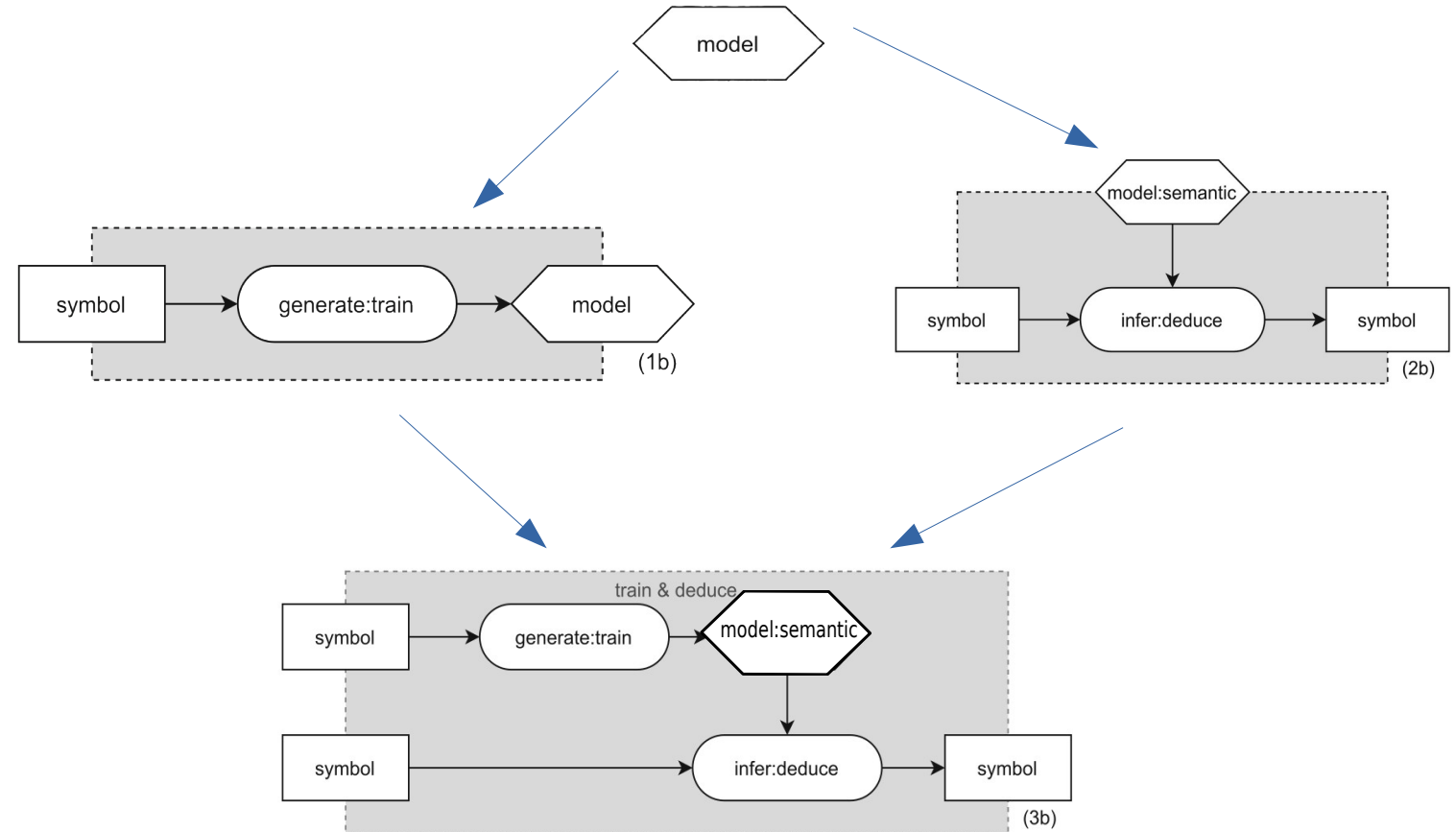
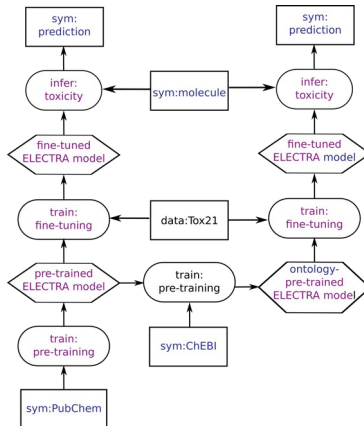


attention with ontology pre-training

Martin Glauer, Fabian Neuhaus, Till Mossakowski, Janna Hastings:
Ontology Pre-training for Poison Prediction. KI 2023: 31-45. Best paper award

Neuro-symbolic design patterns

- NeSy design patterns by Bekkum et al.
- Goal: provide language for refining and combining patterns
- Use DOL's refinement and combination



Till Mossakowski. Modular design patterns for neural-symbolic integration: refinement and combination.
In: 16th International Workshop on Neural-Symbolic Learning and Reasoning 2022,
CEUR Workshop Proceedings 3213, 192–201, 2022.

Overview

1 Symbolic knowledge representation

2 Neuro-symbolic extension of ontologies

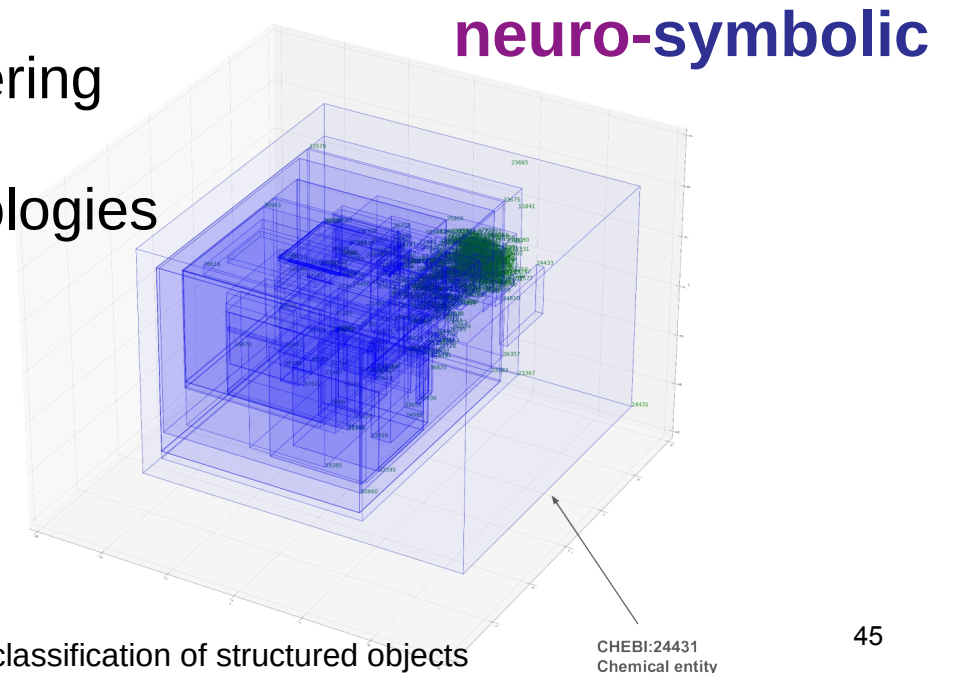
3 Transformers for more performance and interpretability

4 Ontology pre-training can improve Transformer performance

5 Outlook

Conclusion

- Neural networks can extend ontologies of structured objects **neuro** → **symbolic**
- Ontology pre-training can improve transformer performance **symbolic** → **neuro**
- We can beat purely symbolic and purely neural baselines
- Design patterns as systematic building blocks
=> towards a theory of neuro-symbolic engineering
- Future work: Novel neural embeddings for ontologies



Acknowledgements

- Prof. Dr. Janna Hastings, Zürich and St. Gallen
- MSc. Martin Glauer. Magdeburg
- Dr. Fabian Neuhaus, Magdeburg
- MSc Adel Memariani, Magdeburg