Some Advances in Neurosymbolic AI Related to Ontologies and Knowledge Graphs

Pascal Hitzler

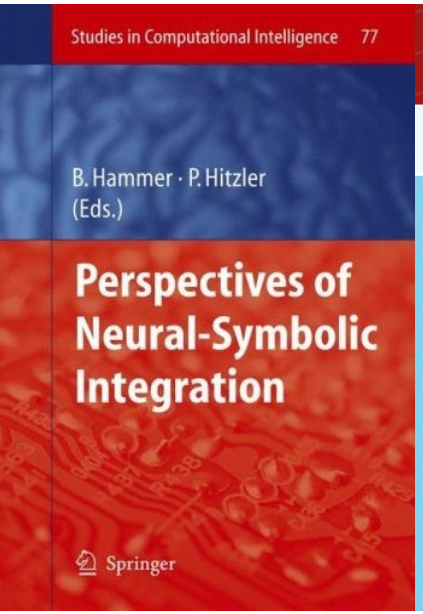
Data Semantics Laboratory (DaSe Lab)

Kansas State University

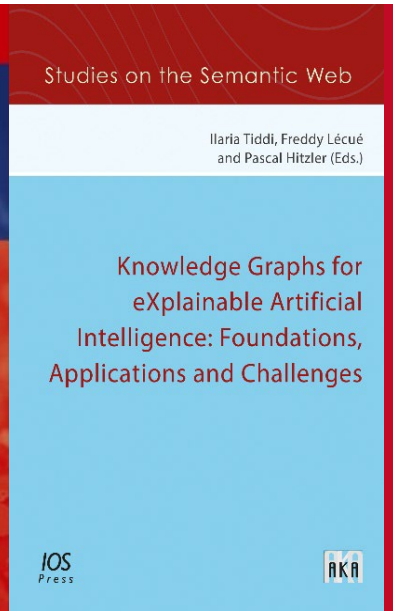
http://www.daselab.org

Neurosymbolic AI

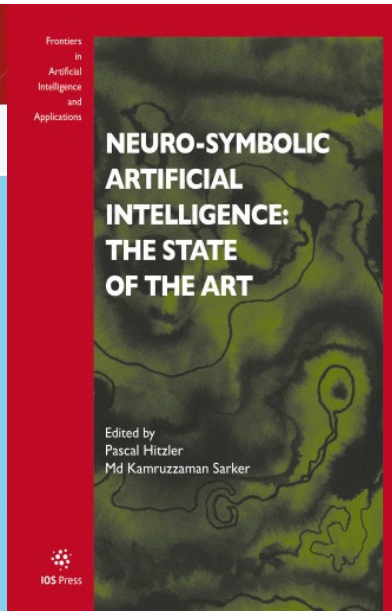
Workshop Series on Neural-Symbolic Learning and Reasoning, since 2005.
Joint with Artur d'Avila Garcez. Conference since 2024.



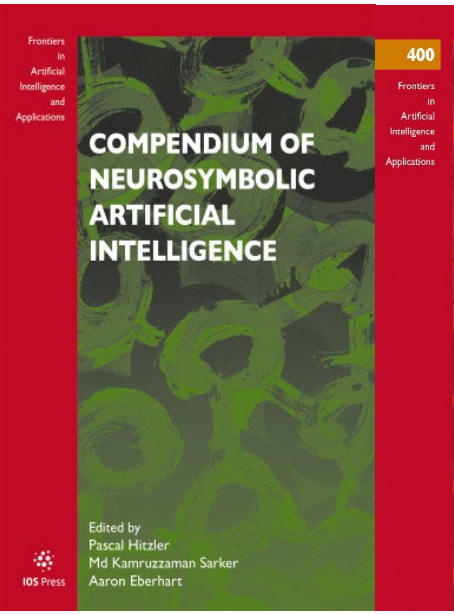
2007



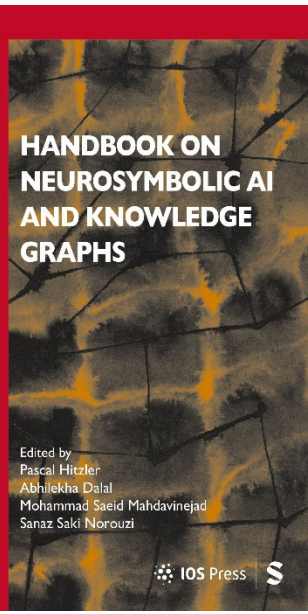
2020



2022



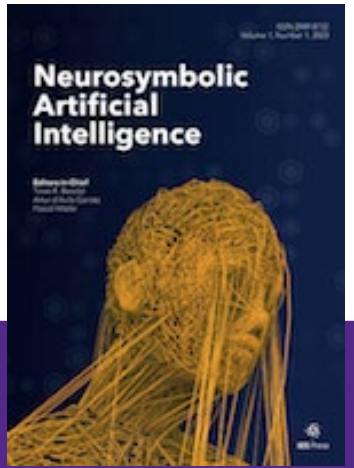
2023



2025

Neurosymbolic AI slack with over 1,300 members
(email hitzler@ksu.edu to join).

Neurosymbolic AI journal (since 2024)



TOC

- **Deep Deductive Reasoning**
  - Monireh Ebrahimi, Aaron Eberhart, Federico Bianchi, Pascal Hitzler, Towards Bridging the Neuro-Symbolic Gap: Deep Deductive Reasoners. Applied Intelligence 51 (9), 6326-6348, 2021. [positive results]
  - Pascal Hitzler, Rushrukh Rayan, Joseph Zalewski, Sanaz Saki Norouzi, Aaron Eberhart, Eugene Y. Vasserman, Deep Deductive Reasoning is a Hard Deep Learning Problem. Neurosymbolic Artificial Intelligence 1, 2025. [negative results]
  - Adrita Barua, Pascal Hitzler, Description Logic Concept Learning using Large Language Models. In: NeSy 2025, 160-178. [new positive results]
  - Pascal Hitzler, Frank van Harmelen, A reasonable Semantic Web. Semantic Web 1 (1-2), 39-44, 2010. [partially in a similar spirit]

- **Ontology-based Hidden Neuron Activation Analysis**
- **LLM-based Knowledge Graph and Ontology Engineering**
- **Some musings on human-AI neurosymbolic agentic whatever**

Deep Deductive Reasoners

- We trained deep learning systems to do deductive reasoning.

- Why is this interesting?
  - For dealing with noisy data (where symbolic reasoners do very poorly).
  - For speed, as symbolic algorithms are of very high complexity.
  - Out of principle because we want to learn about the capabilities of deep learning for complicated cognitive tasks.

  - To perhaps begin to understand how our (neural) brains can learn to do highly symbolic tasks like formal logical reasoning, or in more generality, mathematics.
    A fundamental quest in Cognitive Science.

Reasoning as Classification

- Given a set of logical formulas (a theory).
- Any formula expressible over the same language is either
  - a logical consequence or
  - not a logical consequence.
- This can be understood as a classification problem for machine learning.
- It turns out to be a really hard machine learning problem
  - If we expect

Knowledge Materialization

- Given a set of logical formulas (a theory).
- Produce all logical consequences under certain constraints.
- Without the qualifier this is in general not possible as the set of all logical consequences is infinite.
- So we have to constrain to consequences of, e.g., a certain syntactic form. For relatively simple logics, this is often reasonably possible.

Published deep deductive reasoning work

paper	logic	transfer	generative	scale	performance
[12]	RDFS	yes	no	moderate	high
[25]	RDFS	no	yes	low	high
[10]	$\mathcal{EL}^+$	no	yes	moderate	low
[20]	OWL RL	no*	no	low	high
[6]	FOL	no	yes	very low	high
(new)	RDFS	yes	yes	moderate	???
(new)	EL+	yes	yes	moderate	???

[12]: Ebrahimi, Sarker, Bianchi, Xie, Eberhart, Doran, Kim, Hitzler, AAAI-MAKE 2021

[25]: Makni, Hendler, SWJ 2019

[10]: Eberhart, Ebrahimi, Zhou, Shimizu, Hitzler, AAAI-MAKE 2020

[20]: Hohenecker, Lukasiewicz, JAIR 2020

[6]: Bianchi, Hitzler, AAAI-MAKE 2019

(new): Ebrahimi, Eberhart, Hitzler, June 2021: results were good token-based but triples could not be recovered!

Towards Neurosymbolic DDR

DaSe Lab

Sometimes in logic, finding solutions is high complexity (hard), checking solutions is low complexity (easy).

Use deep learning / generative AI to come up with potential solutions.

Use conventional reasoning algorithms to check correctness.

LLM reasoning issues (Barua & Hitzler 2025)

Asking LLM to infer a formal (logical) definition from data, e.g.

Dora is female, hasChild Rosanna
Rosanna hasChild Valentina

Luigi is male, hasChild Dino
Dino hasChild Francesco

Maria is female, hasChild Serena
(etc; several examples)

Dora is a Grandmother, Luigi and Maria are not Grandmothers.

We would ask for the generation of a definition of Grandmother:

Female and (hasChild some (hasChild some Person)) .

This is an old benchmark from ILP carried over do Description Logics / Concept Induction.

LLM reasoning issues (Barua et al.)

	original	gender flipped
GPT-4o correct	9	5
incorrect	0	4

A model made for reasoning does somewhat better, but takes much more time (and costs more):

o3-mini correct	9	6
incorrect	0	3

“gender flipped” correct answer for “GrandMother”:
Male and (hasChild some (hasChild some Person))

LLM reasoning issues (Barua et al.)

If we introduce feedback and fact-checking, we're doing much better:

	original	gender flipped
GPT-4o correct	9	5 → 7
incorrect	0	4 → 2
o3-mini correct	9	6 → 9
incorrect	0	3 → 0

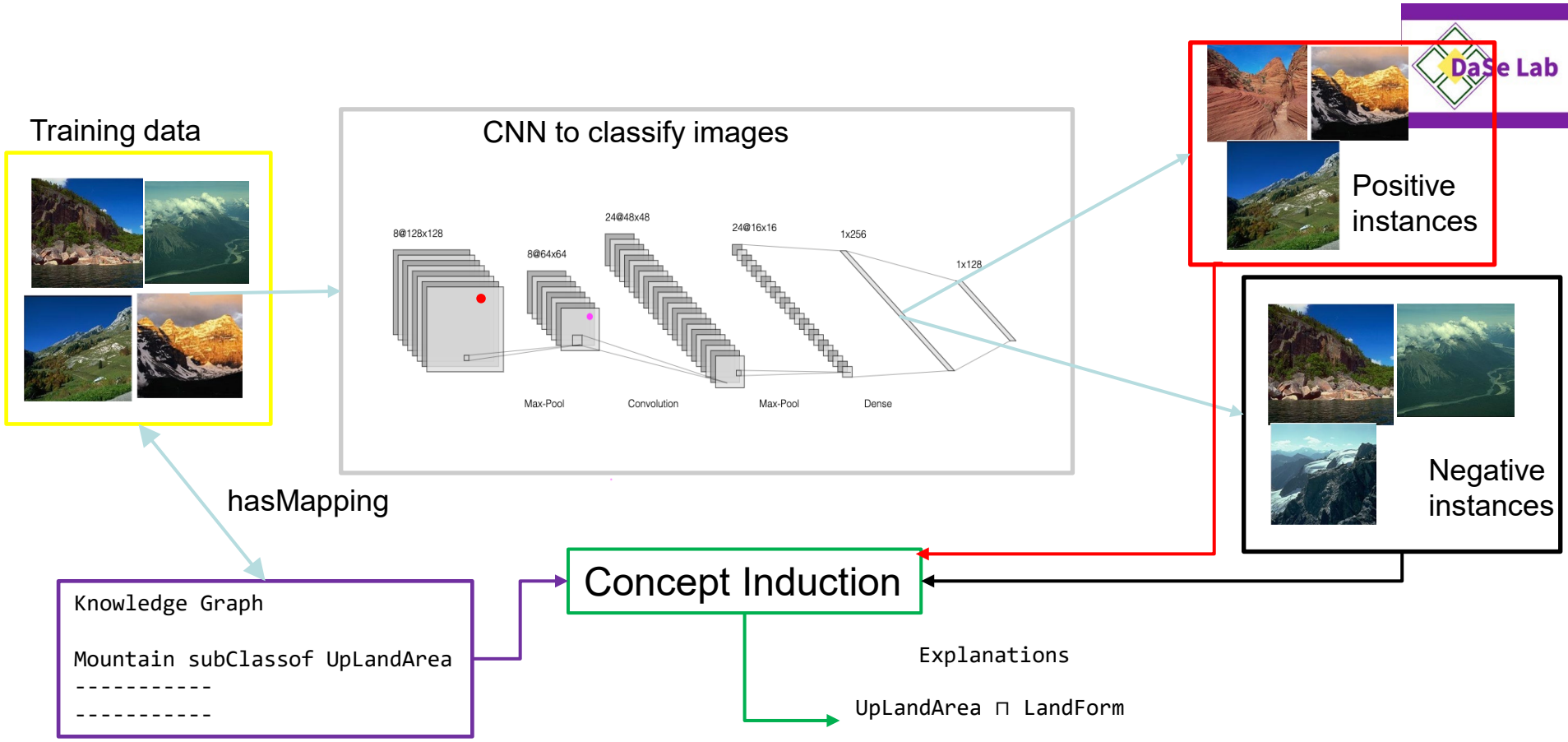
Key insights:

- **Counterfactuals are problematic**
- **Checking helps. Note that checking is much cheaper than deduction.**

TOC

- Deep Deductive Reasoning
- Ontology-based Hidden Neuron Activation Analysis
- LLM-based Knowledge Graph and Ontology Engineering
- Some musings on human-AI neurosymbolic agentic whatever

Idea



Abhilekha Dalal, Rushrukh Rayan, Adrita Barua, Samatha Ereshi Akkamahadevi, Avishek Das, Cara Widmer, Eugene Y. Vasserman, Md Kamruzzaman Sarker, Pascal Hitzler, Towards a Neurosymbolic Understanding of Hidden Neuron Activations. Neurosymbolic Artificial Intelligence. To appear.

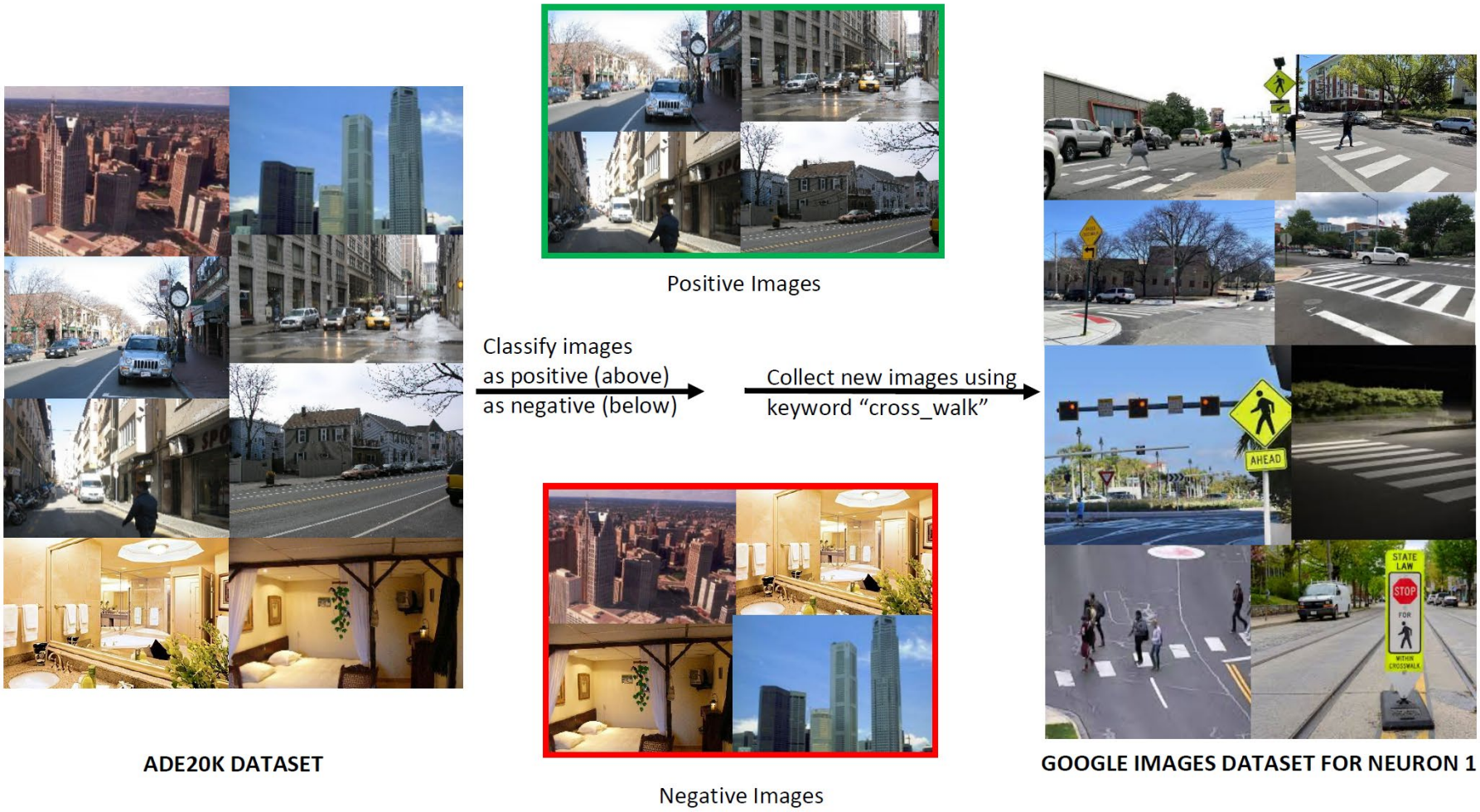


Figure 1: Example of images that were used for generating and confirming the label hypothesis for neuron 1

workflow: label hypothesis generation and confirmation of label hypothesis with new images from Google images

Neuron #	Obtained Label(s)	Images	Coverage	Target %	Non-Target %
0	**building**	**164**	**0.997**	**89.024**	**72.328**
1	**cross_walk**	**186**	**0.994**	**88.710**	**28.923**
3	**night_table**	**157**	**0.987**	**90.446**	**56.714**
6	dishcloth, toaster	106	0.999	16.038	39.078
7	toothbrush, Pipage	112	0.991	75.893	59.436
8	**shower_stall, cistern**	**136**	**0.995**	**100.000**	**53.186**
11	river_water	157	0.995	31.847	22.309
12	baseboard, dish_rag	108	0.993	75.926	48.248
14	rocking_horse, rocker	86	0.985	54.651	47.816
16	**mountain, bushes**	**108**	**0.995**	**87.037**	**24.969**
17	stem	133	0.993	30.827	31.800
18	**slope**	**139**	**0.983**	**92.086**	**69.919**
19	**wardrobe, air_conditioning**	**110**	**0.999**	**89.091**	**65.034**
20	fire_hydrant	158	0.990	5.696	13.233
22	**skyscraper**	**156**	**0.992**	**99.359**	**54.893**
23	fire_escape	162	0.996	61.111	18.311
25	spatula, nuts	126	0.999	2.381	0.883
26	skyscraper, river	112	0.995	77.679	35.489
27	manhole, left_arm	85	0.996	35.294	26.640
28	flooring, fluorescent_tube	115	1.000	38.261	33.198
29	**lid, soap_dispenser**	**131**	**0.998**	**99.237**	**78.571**
30	**teapot, saucepan**	**108**	**0.998**	**81.481**	**47.984**
31	fire_escape	162	0.961	77.160	63.147
33	tanklid, slipper	81	0.987	41.975	30.214
34	left_foot, mouth	110	0.994	20.909	49.216

Neuron #	Obtained Label(s)	Images	Coverage	Target %	Non-Target %
35	utensils_canister, body	111	0.999	7.207	11.223
36	**tap, crapper**	**92**	**0.997**	**89.130**	**70.606**
37	cistern, doorcase	101	0.999	21.782	24.147
38	letter_box, go_cart	125	0.999	28.000	31.314
39	side_rail	148	0.980	35.811	34.687
40	sculpture, side_rail	119	0.995	25.210	21.224
41	**open_fireplace, coffee_table**	**122**	**0.992**	**88.525**	**16.381**
42	pillar, stretcher	117	0.998	52.137	42.169
43	**central_reservation**	**157**	**0.986**	**95.541**	**84.973**
44	saucepan, dishrack	120	0.997	69.167	36.157
46	Casserole	157	0.999	45.223	36.394
48	**road**	**167**	**0.984**	**100.000**	**73.932**
49	**footboard, chain**	**126**	**0.982**	**88.889**	**66.702**
50	night_table	157	0.972	65.605	62.735
51	**road, car**	**84**	**0.999**	**98.810**	**48.571**
53	pylon, posters	104	0.985	11.538	17.332
54	**skyscraper**	**156**	**0.987**	**98.718**	**70.432**
56	**flusher, soap_dish**	**212**	**0.997**	**90.094**	**63.552**
57	**shower_stall, screen_door**	**133**	**0.999**	**98.496**	**31.747**
58	plank, casserole	80	0.998	3.750	3.925
59	manhole, left_arm	85	0.994	35.294	21.589
60	paper_towels, jar	87	0.999	0.000	1.246
61	ornament, saucepan	102	0.995	43.137	17.274
62	sideboard	100	0.991	21.000	29.734
63	**edifice, skyscraper**	**178**	**0.999**	**92.135**	**48.761**

Evaluation

- Each row of the table is a hypothesis, e.g. “neuron 1 activates more strongly on cross_walk images (retrieved from Google images using keyword “cross_walk”) than on other images.”
- Null hypothesis: There is no difference in activations.
- There is no reason to assume a normal distribution,
- hence using Mann-Whitney U test for assessment.

Evaluation results

Neuron #	Label(s)	Images	# Activations (%)		Mean		Median		z-score	p-value
			targ	non-t	targ	non-t	targ	non-t		
0	building	42	80.95	73.40	2.08	1.81	2.00	1.50	-1.28	0.0995
1	cross_walk	47	91.49	28.94	4.17	0.67	4.13	0.00	-8.92	<.00001
3	night_table	40	100.00	55.71	2.52	1.05	2.50	0.35	-6.84	<.00001
8	shower_stall, cistern	35	100.00	54.40	5.26	1.35	5.34	0.32	-8.30	<.00001
16	mountain, bushes	27	100.00	25.42	2.33	0.67	2.17	0.00	-6.72	<.00001
18	slope	35	91.43	68.85	1.59	1.37	1.44	1.00	-2.03	0.0209
19	wardrobe, air_conditioning	28	89.29	65.81	2.30	1.28	2.30	0.84	-4.00	<.00001
22	skyscraper	39	97.44	56.16	3.97	1.28	4.42	0.33	-7.74	<.00001
29	lid, soap_dispenser	33	100.00	80.47	4.38	2.14	4.15	1.74	-5.92	<.00001
30	teapot, saucepan	27	85.19	49.93	2.52	1.05	2.23	0.00	-4.28	<.00001
36	tap, crapper	23	91.30	70.78	3.24	1.75	2.82	1.29	-3.59	<.00001
41	open_fireplace, coffee_table	31	80.65	15.11	2.03	0.14	2.12	0.00	-7.15	<.00001
43	central_reservation	40	97.50	85.42	7.43	3.71	8.08	3.60	-5.94	<.00001
48	road	42	100.00	74.46	6.15	2.68	6.65	2.30	-7.78	<.00001
49	footboard, chain	32	84.38	66.41	2.63	1.67	2.30	1.17	-2.58	0.0049
51	road, car	21	100.00	47.65	5.32	1.52	5.62	0.00	-6.03	<.00001
54	skyscraper	39	100.00	71.78	4.14	1.61	4.08	1.12	-7.60	<.00001
56	flusher, soap_dish	53	92.45	64.29	3.47	1.48	3.08	0.86	-6.47	<.00001
57	shower_stall, screen_door	34	97.06	32.31	2.60	0.61	2.53	0.00	-7.55	<.00001
63	edifice, skyscraper	45	88.89	48.38	2.41	0.83	2.36	0.00	-6.73	<.00001

Table 3: Evaluation details as discussed in Section 4. Images: number of images used for evaluation. # Activations: (targ(et)): Percentage of target images activating the neuron (i.e., activation at least 80% of this neuron's activation maximum); (non-t): Same for all other images used in the evaluation. Mean/Median (targ(et)/non-t(arget)): mean/median activation value for target and non-target images.

-target images not activating neuron 1

Non-target images activating neuron 1

Figure 2: Examples of some Google images used: target images ("cross_walk") that did not activate the neuron; non-target images from labels like "central_reservation," "road and car," and "fire_hydrant" that activated the neuron.

Note: "bushes, bush" is the third-highest concept induction output (coverage 0.993; 48.052% of target images activating the neuron)

Comparison with alternative methods

For each approach we checked the top 64 responses.

Concept Induction:	**18 confirmed neuron labels.**
GPT-4 (no checking):	**12 confirmed neuron labels**
CLIP-Dissect:	**5 confirmed neuron labels**

Tuomas P. Oikarinen, Tsui-Wei Weng: CLIP-Dissect: Automatic Description of Neuron Representations in Deep Vision Networks. ICLR 2023

Replication

In the meantime we replicated these results successfully

- **With other classes from ADE20k**
  - Samatha Ereshi Akkamahadevi, Abhilekha Dalal, Pascal Hitzler, Evaluation of Concept Induction in Explainable AI using Multiple Datasets. In: Ron Petrick and Christopher Gelb (eds.), Proceedings of the AAAI 2025 Spring Symposium Series. AAAI Press, 2025, pp. 317-326.
- **With a completely different benchmark, SUN2012**
  - Moumita Sen Sarma, Samatha Ereshi Akkamahadevi and Pascal Hitzler, A Case Study on Concept Induction for Neuron-Level Interpretability in CNN. K-Cap 2025 (Poster), Dayton, OH, December 2025.
- **For text classification using LSTM (dense layer) (Avishek)**
  - Avishek Das, Abhilekha Dalal, Pascal Hitzler, Hidden Neuron Activation Analysis on Labeled Text Data. K-Cap 2025, Dayton, OH, December 2025.

Meaningfulness for Humans

- **Hypothesis:**
  - **ECII explanations are better than semi-random explanations, but worse than human-generated explanations.**
- **Experimental setting as before.**
- **300 Amazon Mechanical Turk participants**
- **Seven concepts taken from top ECII results.**
- **45 image set pairs, each set corresponding to a category.**

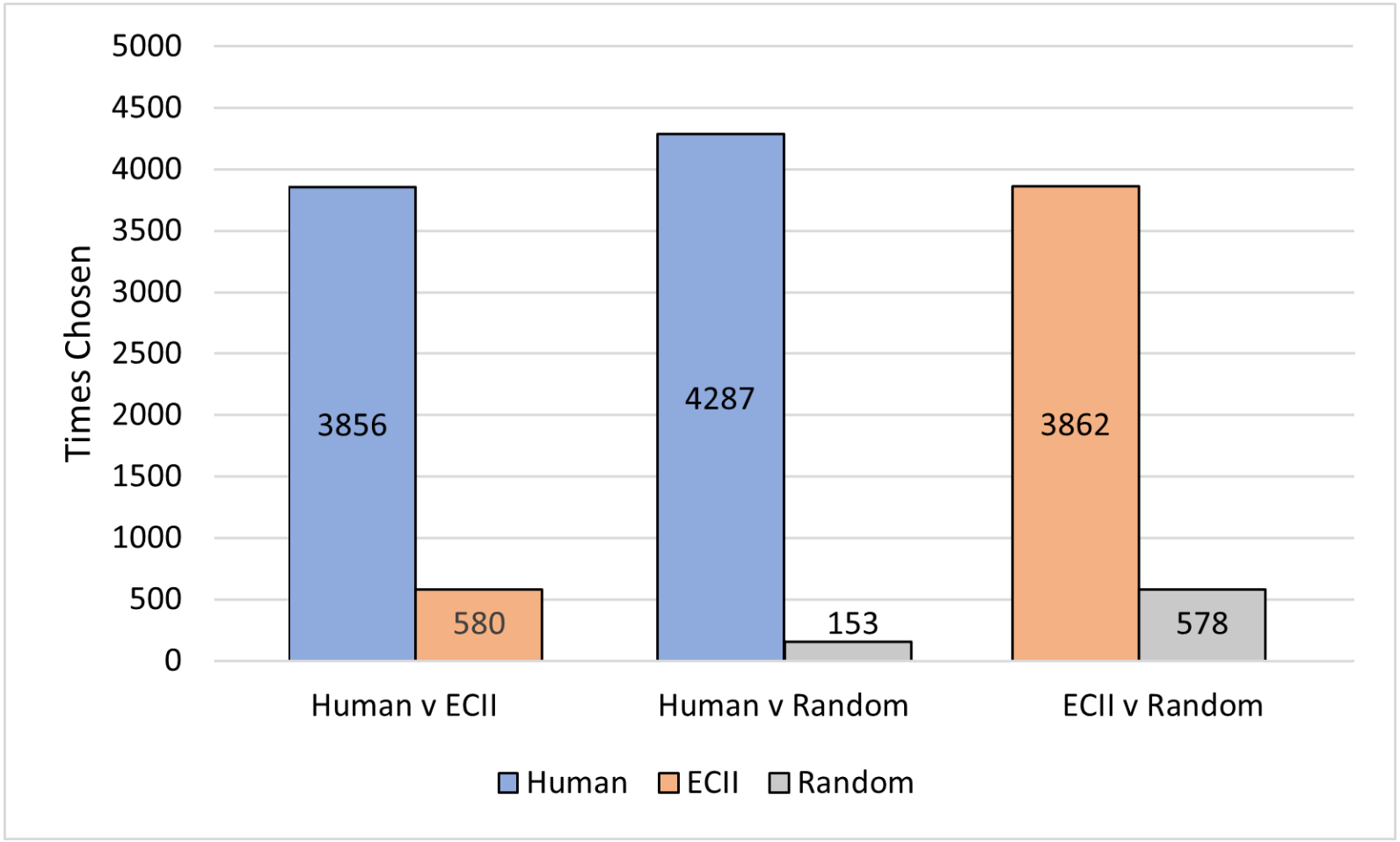
Cara Widmer, Md Kamruzzaman Sarker, Srikanth Nadella, Joshua Fiechter, Ion Juvina, Brandon Minnery, Pascal Hitzler, Joshua Schwartz, Michael Raymer, Towards Human-Compatible XAI: Explaining Data Differentials with Concept Induction over Background Knowledge
Journal of Web Semantics 79:100807, 2023

Are the results human-compatible? Part I

Which of these better represents what the images in group A have that the images in group B do not?

Bake, Bakery, Bread, Indoor, Product, Store, Woman	Basket, Bread, Cake, Ceiling, Floor, Person, Wall

Are the results human-compatible? Part I



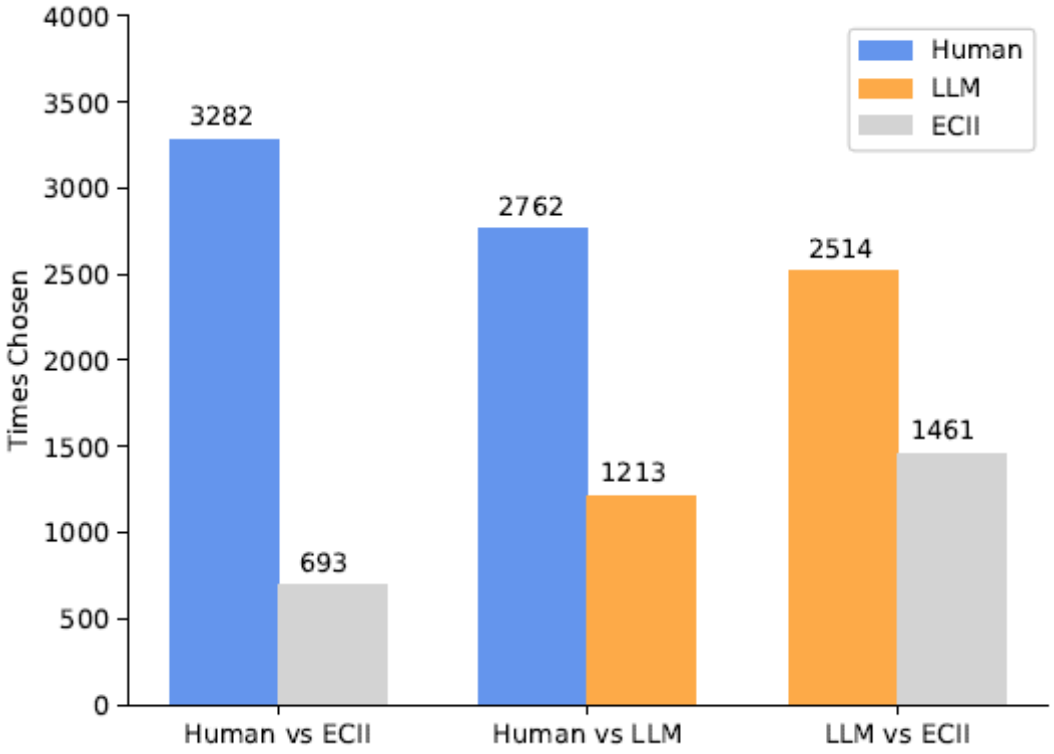
Comparison with GPT-4

HvE and HvL Explanations:

both found statistically significant at $p < 0.0001$

LvE Explanations:

statistically significant at $p < 0.001$



Adrita Barua, Cara Widmer, Pascal Hitzler, Concept Induction using LLMs: a user experiment for assessment. In: Tarek R. Besold, Artur d'Avila Garcez, Ernesto Jimenez-Ruiz, Roberto Confalonieri, Pranava Madhyastha, Benedikt Wagner (eds.), Neural-Symbolic Learning and Reasoning - 18th International Conference, NeSy 2024, Barcelona, Spain, September 9--12, 2024, Proceedings, Part II. Lecture Notes in Computer Science 14980, Springer 2024, pp. 132-148.

TOC

- Deep Deductive Reasoning
- Ontology-based Hidden Neuron Activation Analysis
- LLM-based Knowledge Graph and Ontology Engineering
- Some musings on human-AI neurosymbolic agentic whatever

Main underlying papers

- Cogan Shimizu, Pascal Hitzler, Accelerating Knowledge Graph and Ontology Engineering with Large Language Models. Journal of Web Semantics 85: 100862, 2025.
- Cogan Shimizu, Karl Hammar, Pascal Hitzler, Modular Ontology Modeling. Semantic Web 14 (3), 459-489, 2023.
- Reihaneh Amini, Sanaz Saki Norouzi, Pascal Hitzler, Reza Amini, Towards Complex Ontology Alignment using Large Language Models. KGSWC 2024.
- Adrita Barua, Reza Amini, Sanaz Saki Norouzi, Reihaneh Amini, Pascal Hitzler, Complex Ontology Alignment using LLMs: a case study. In: OM 2025, to appear.

KG and Ontology Engineering (KGOE)

- **KGOE: (vaguely defined) set of tasks central to the lifecycle of ontologies and KGs (as data artifacts).**

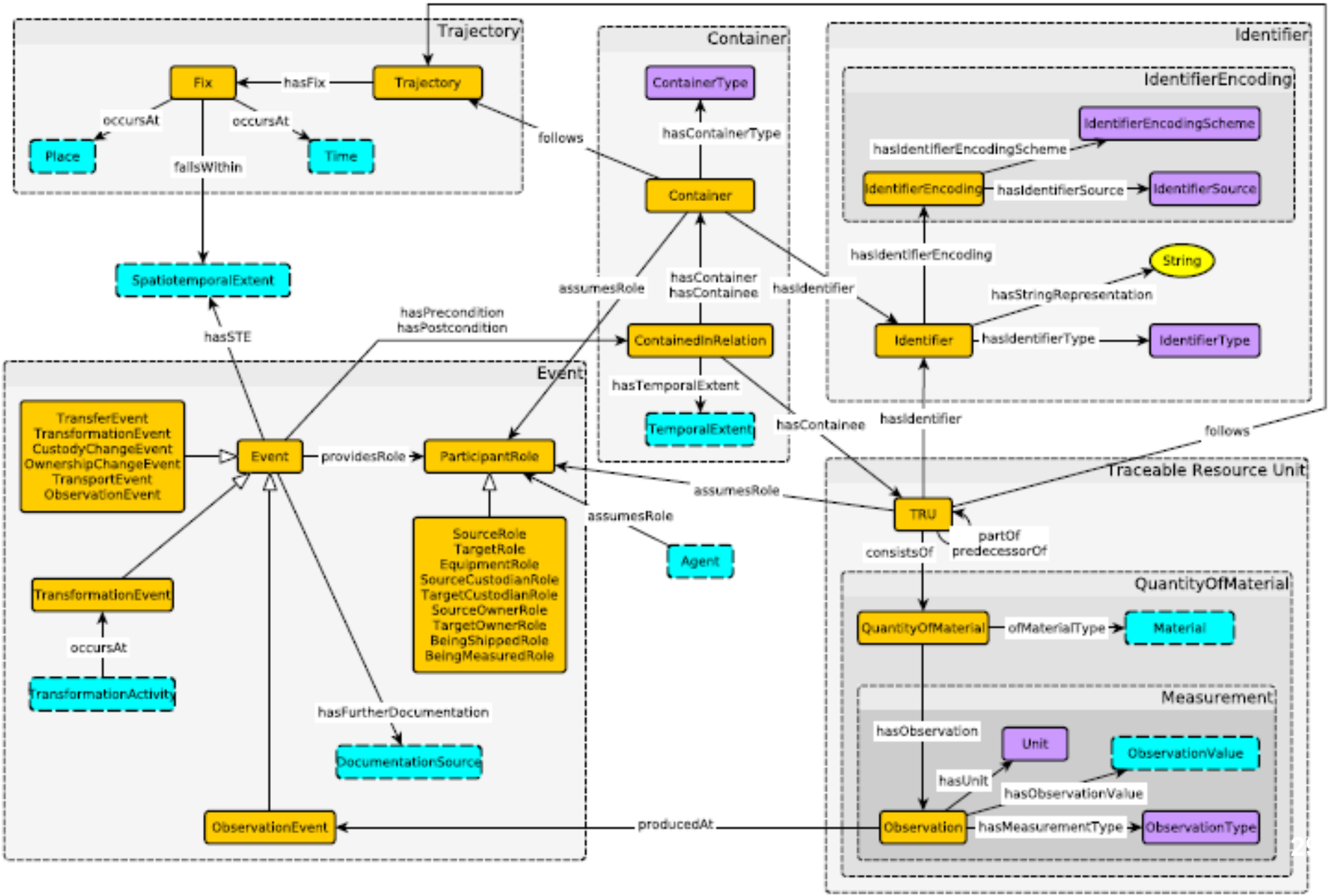
(Some) key tasks are non-trivial and expensive (time, expertise) to obtain high quality outcomes:

- **Ontology modeling (schema construction)**
  - **relatedly, ontology extension and modification**
- **Ontology population (KG construction compliant with schema)**
- **Ontology alignment (schema mapping)**
- **Entity disambiguation / co-reference resolution**

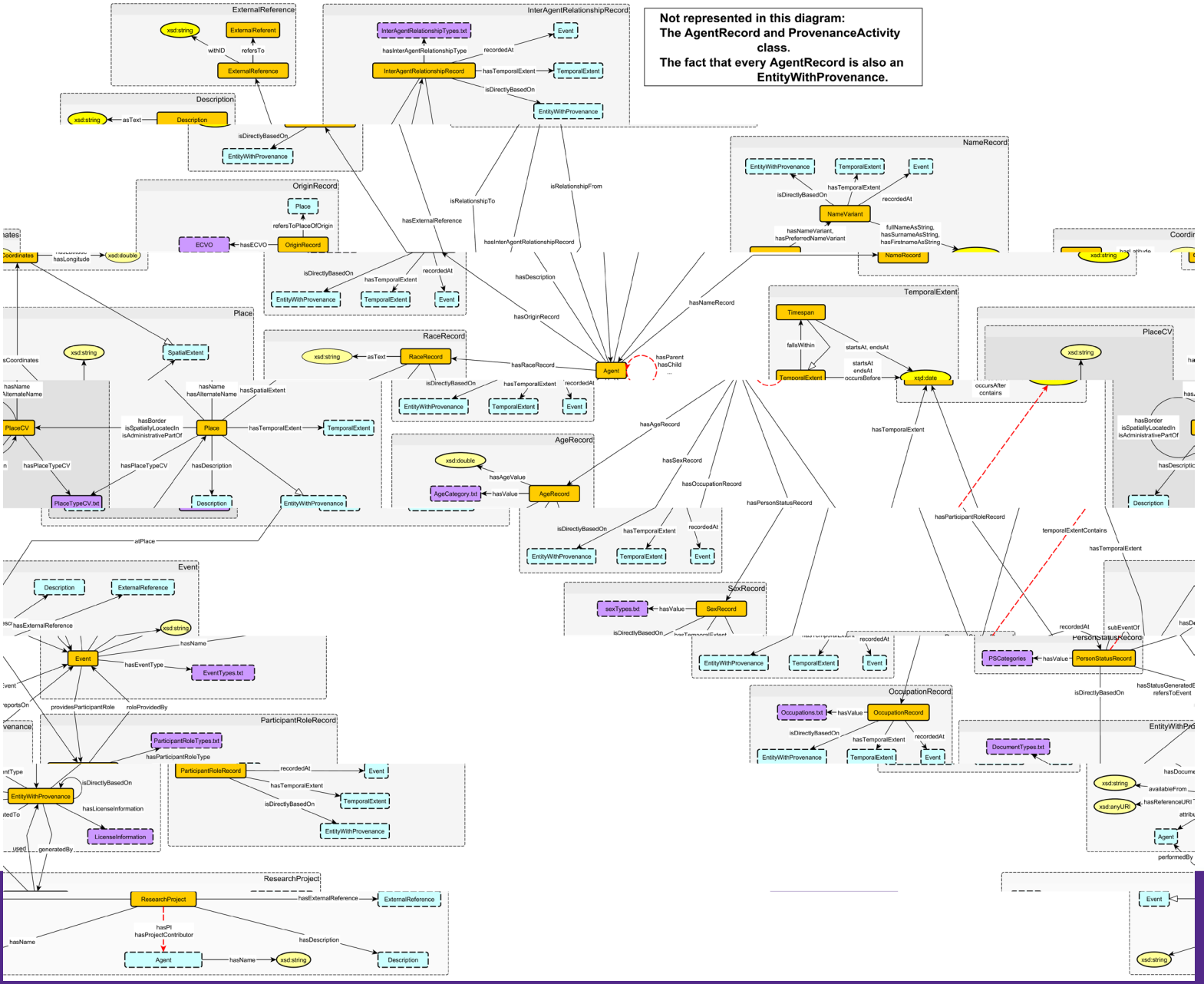
Conjecture:

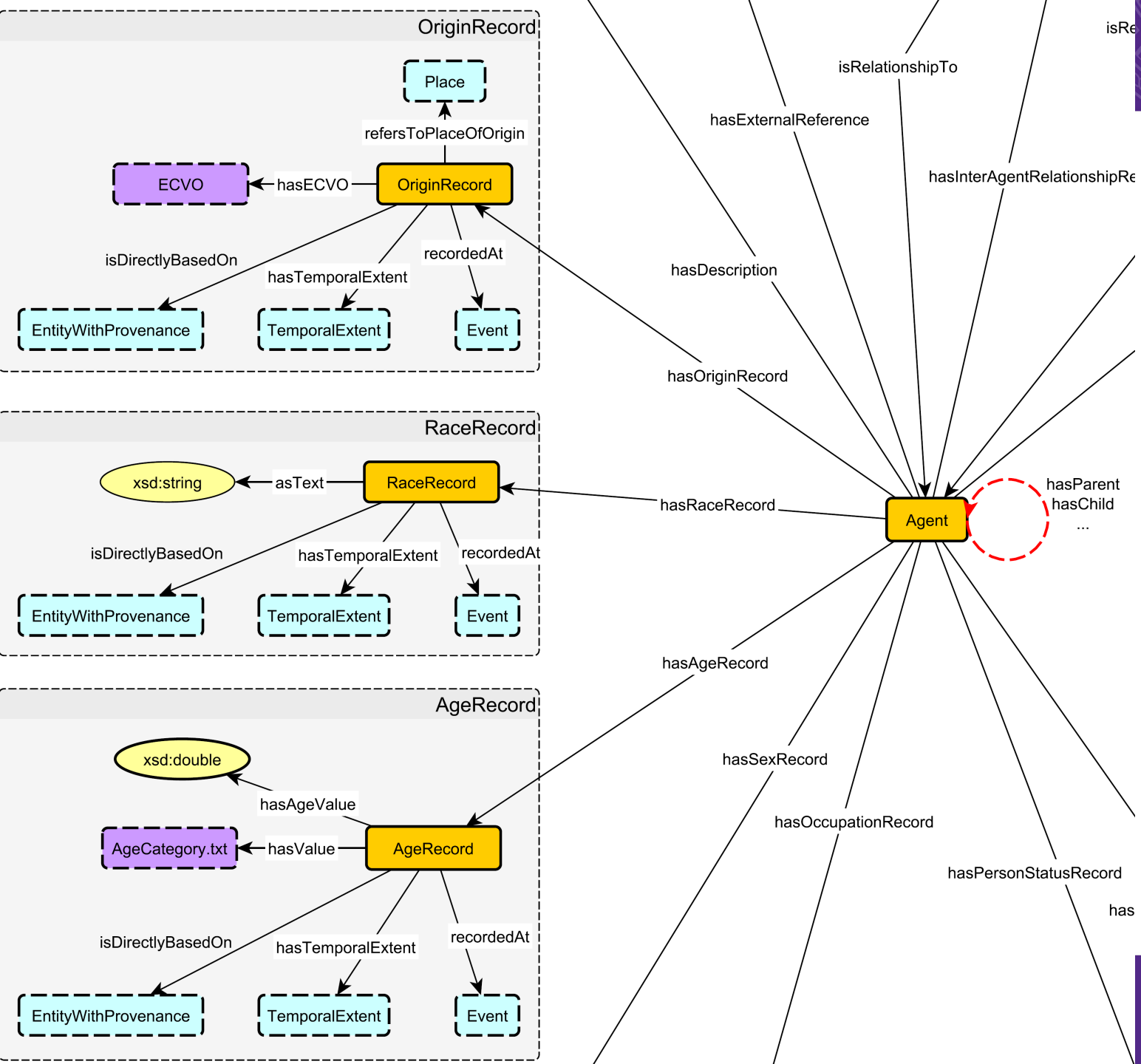
- **Ontology modules will also help with (all) KGOE tasks.**
  - **Both for humans and for automation.**

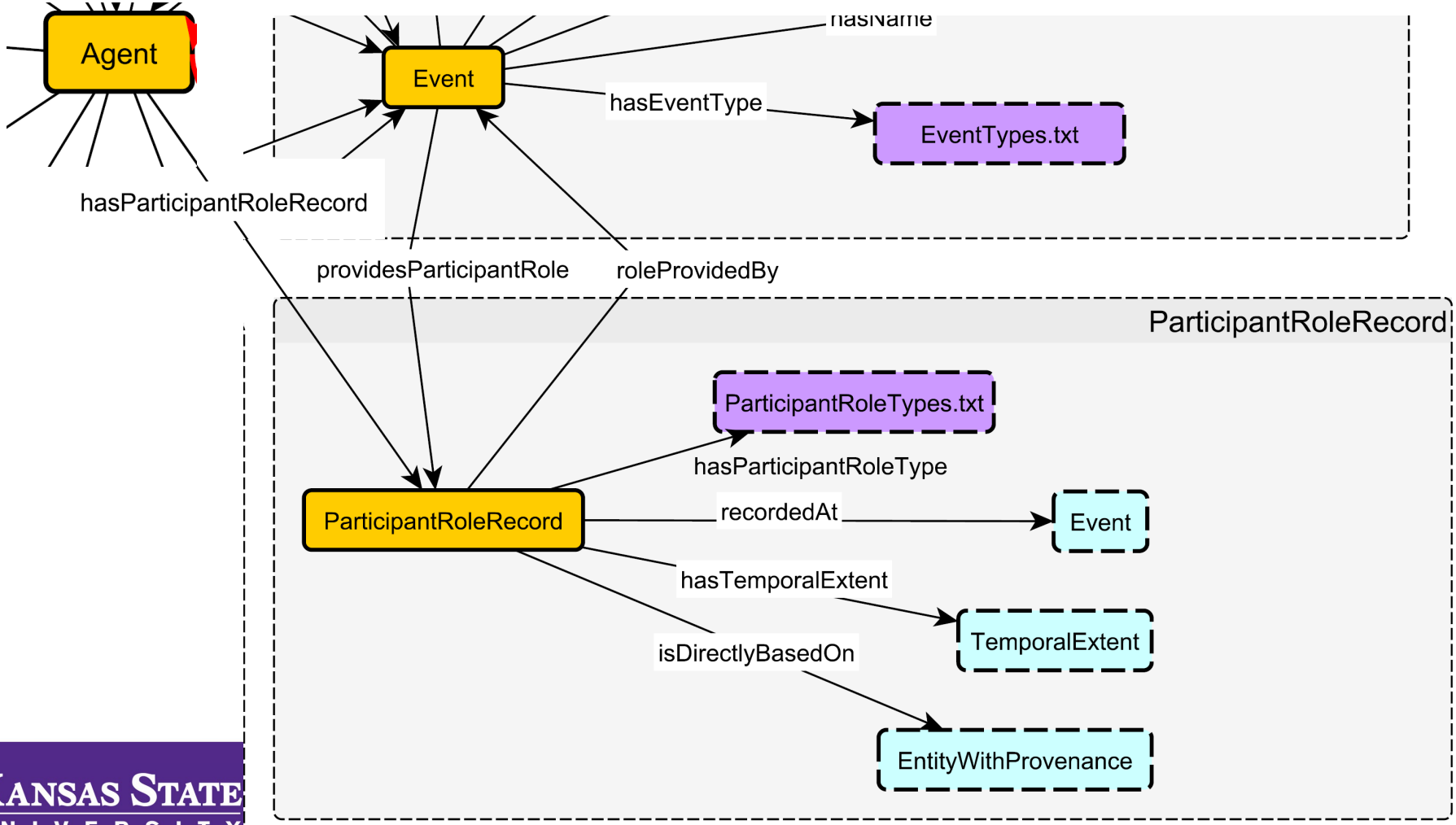
Modular Ontologies (Shimizu et al., 2023)

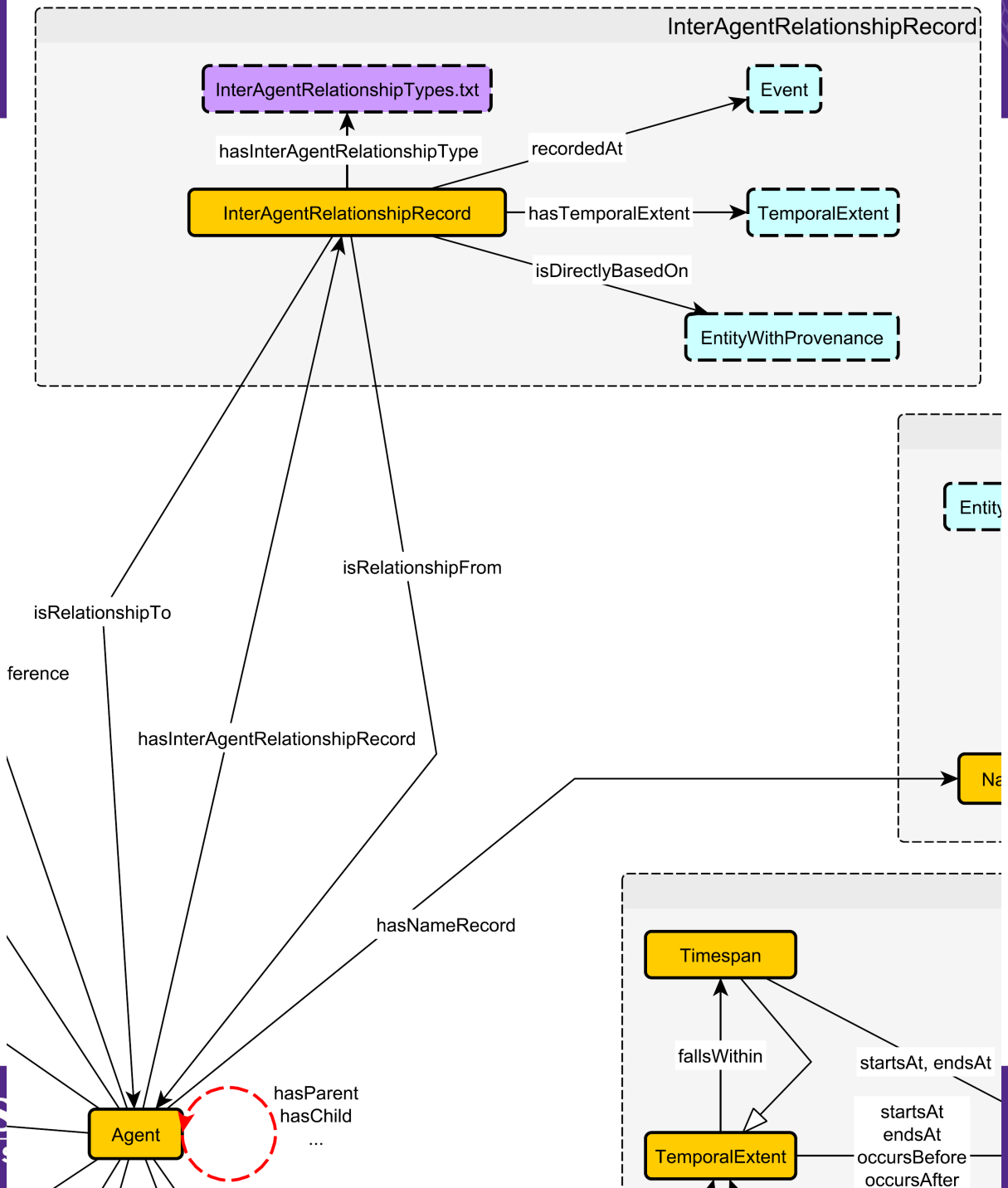


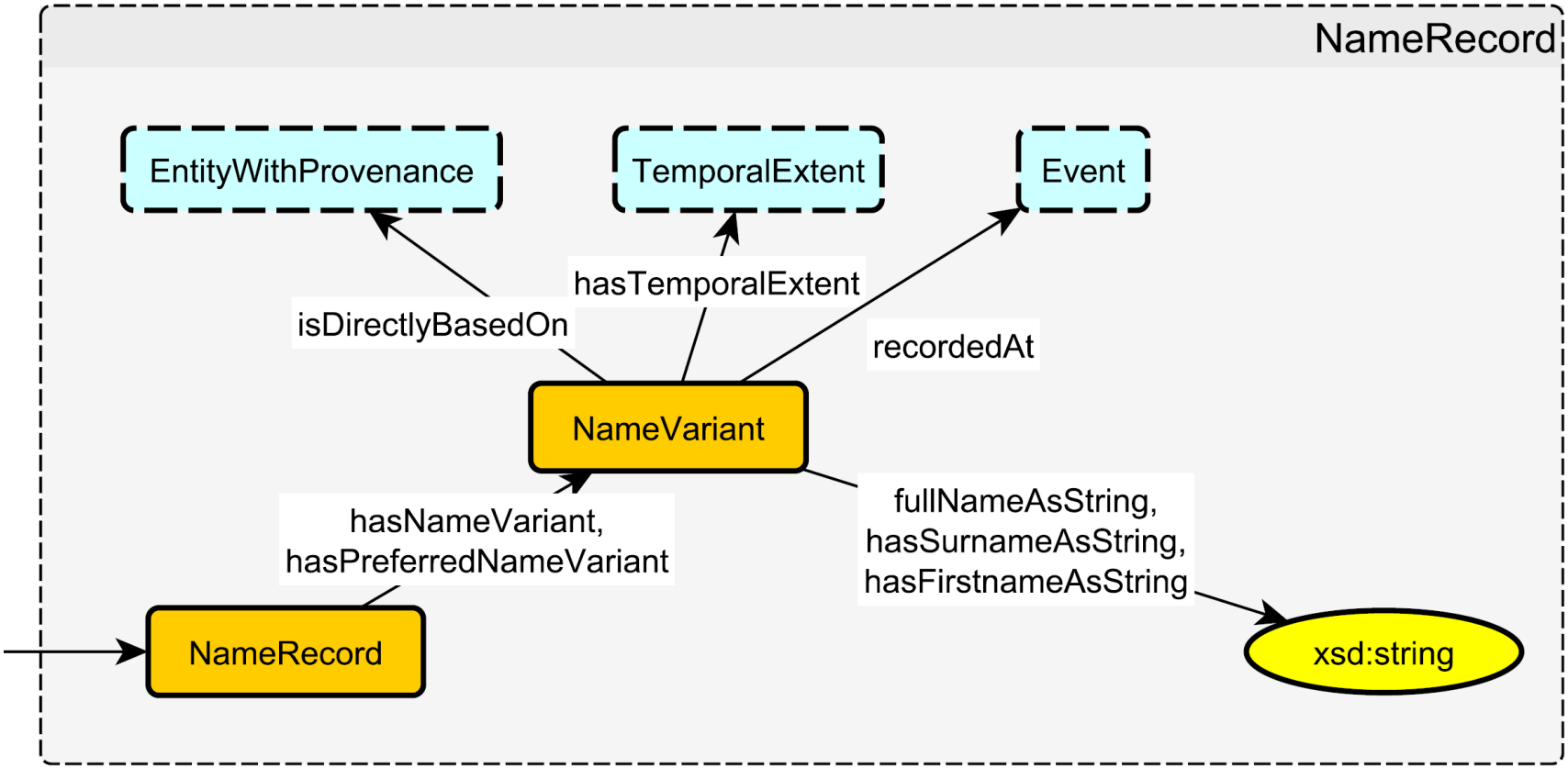
Enslaved Ontology [JWS 2020]




OriginRecord
Place
refersToPlaceOfOrigin
ECVO
hasECVO
OriginRecord
isDirectlyBasedOn
hasTemporalExtent
recordedAt
EntityWithProvenance
TemporalExtent
Event
RaceRecord
xsd:string
asText
RaceRecord
isDirectlyBasedOn
hasTemporalExtent
recordedAt
EntityWithProvenance
TemporalExtent
Event
AgeRecord
xsd:double
hasAgeValue
AgeCategory.txt
hasValue
AgeRecord
isDirectlyBasedOn
hasTemporalExtent
recordedAt
EntityWithProvenance
TemporalExtent
Event
isRelationshipTo
hasExternalReference
hasInterAgentRelationshipRe
hasDescription
hasOriginRecord
hasRaceRecord
Agent
hasParent
hasChild
...
hasAgeRecord
hasSexRecord
hasOccupationRecord
hasPersonStatusRecord
has


Agent
Event
hasName
hasEventType
EventTypes.txt
hasParticipantRoleRecord
providesParticipantRole
roleProvidedBy
ParticipantRoleRecord
ParticipantRoleTypes.txt
hasParticipantRoleType
ParticipantRoleRecord
recordedAt
Event
hasTemporalExtent
TemporalExtent
isDirectlyBasedOn
EntityWithProvenance


InterAgentRelationshipRecord
InterAgentRelationshipTypes.txt
Event
hasInterAgentRelationshipType
recordedAt
InterAgentRelationshipRecord
hasTemporalExtent
TemporalExtent
isDirectlyBasedOn
EntityWithProvenance
isRelationshipFrom
isRelationshipTo
ference
hasInterAgentRelationshipRecord
hasNameRecord
Timespan
fallsWithin
startsAt, endsAt
hasParent
hasChild
...
Agent
TemporalExtent
startsAt
endsAt
occursBefore
occursAfter


NameRecord
EntityWithProvenance
TemporalExtent
Event
hasTemporalExtent
isDirectlyBasedOn
recordedAt
NameVariant
hasNameVariant,
hasPreferredNameVariant
fullNameAsString,
hasSurnameAsString,
hasFirstnameAsString
NameRecord
xsd:string

Modular Ontologies (Shimizu et al., 2023)

Conjectures:

Modularity gives additional content to ontologies.
This content is very helpful for many KGOE tasks.

Modules are easy to define at ontology creation time.

Modularity helps humans in KGOE tasks.

Ergo it should also help LLMs.

State of Ontology Alignment

Ontology Alignment: Automated creation of mappings between ontologies.

OAEI benchmarking competition, since 2004
https://oaei.ontologymatching.org/

- **Almost all work is on simple (1-1) class alignments**
- **Only little work (mid 2010s) on complex (n-m) alignment mappings. Didn't really work without strong assumptions.**

$$\begin{aligned}&\texttt{Award}(x) \wedge \texttt{hasCoPrincipalInvestigator}(x, z) \leftrightarrow \\ &\texttt{FundingAward}(x) \wedge \texttt{providesAgentRole}(x, y) \\ &\wedge \texttt{CoPrincipalInvestigatorRole}(y) \wedge \texttt{performedBy}(y, z)\end{aligned}$$

- **Now there seems to be a reversion to simple alignments, just with LLMs ☹**

- **GeoLink complex alignment benchmark** **(Krisnadhi et al. ISWC 2015) (Zhou et al. Data Intel. 2020)**

$$\texttt{Award}(x) \wedge \texttt{hasCoPrincipalInvestigator}(x, z) \leftrightarrow$$
$$\texttt{FundingAward}(x) \wedge \texttt{providesAgentRole}(x, y)$$
$$\wedge\, \texttt{CoPrincipalInvestigatorRole}(y) \wedge \texttt{performedBy}(y, z)$$

- **Enslaved Ontology complex alignment benchmark** **(Shimizu et al. JWS 2020) (Zhou et al. CIKM 2020)**

$$\texttt{enslaved:Person}(x) \wedge \texttt{enslaved:hasAgeRecord}(x, y) \wedge \texttt{enslaved:AgeRecord}(y)$$
$$\leftrightarrow \texttt{ed:Q410(Person)}(x) \wedge \texttt{ep:P42(hasAge)}(x, y) \wedge \texttt{wikibase:Statement}(y)$$

Ontology into Wikibase: Issues

- The Wikibase realization is similar but not identical to the ontology.
- The RDF export is based on an ontology (graph schema) that is inferred from the Wikibase realization.

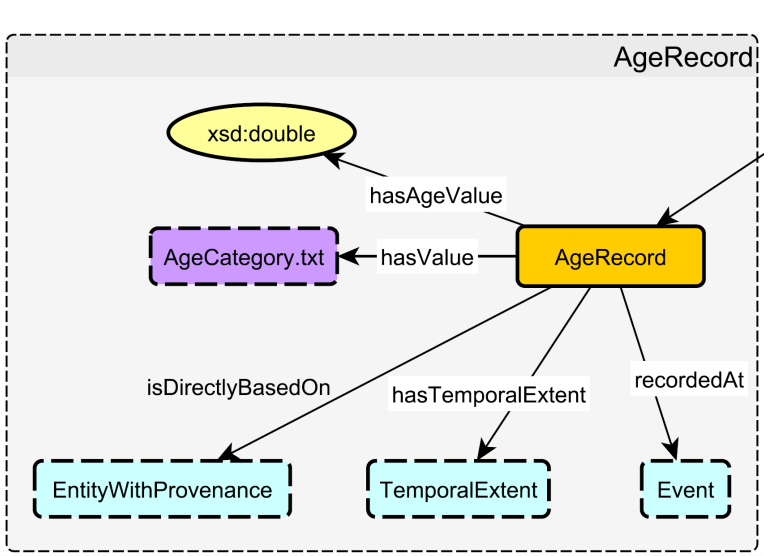
- How difficult is it to map between these two ontologies (and thus, between these to graphs?

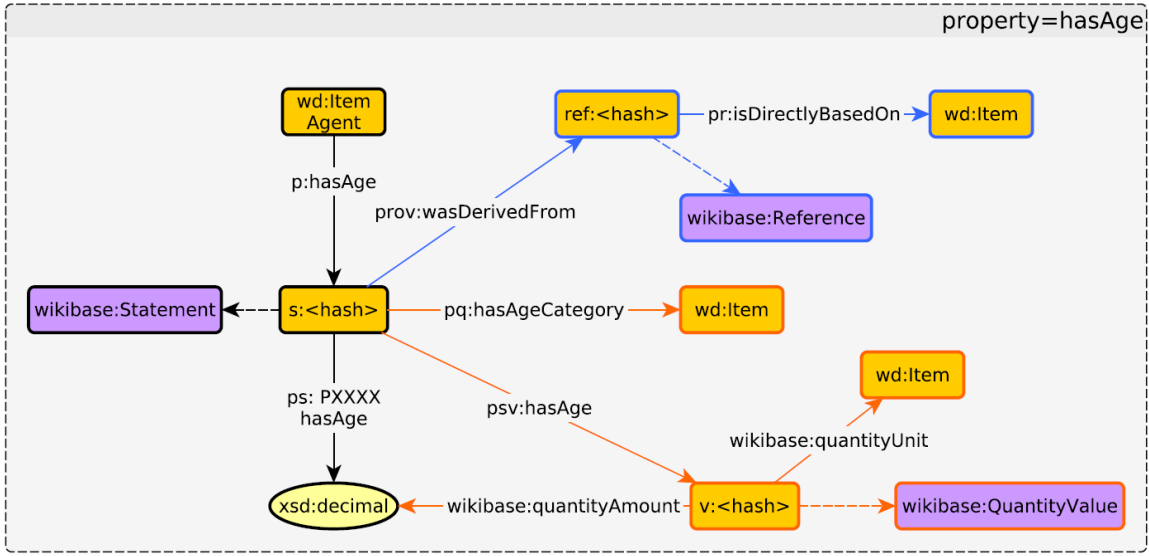
- We used this as basis for an OAEI "complex ontology alignment" benchmark, published at CIKM 2020
  - This realistic benchmark was, in 2020, way beyond capabilities of automated ontology alignment systems.

Let's do it: complex ontology alignment

- **2 ontologies, different structure, same topic.**
- **Merge KGs by mapping between schemas, then use mappings to carry over data.**

$$enslaved : Person(w) \wedge enslaved : hasAgeRecord(w, x) \wedge enslaved : AgeRecord(x) \wedge enslaved : hasAgeValue(x, z) \leftrightarrow$$
$$ed : Q410(Person)(w) \wedge ep : P42(hasAge)(w, x) \wedge wikibase : Statement(x) \wedge eps : P42(hasAge)(x, y) \wedge$$
$$ed : Q424(AgeRecord)(y) \wedge edt : P3(hasAgeValue)(y, z)$$





Enslaved benchmark example mapping rules

$$enslaved : Person(x) \wedge enslaved : hasRaceRecord(x, y) \wedge enslaved : RaceRecord(y) \wedge enslaved : asText(y, z) \leftrightarrow ed : Q410(Person)(x) \wedge ep : P32(hasRaceorColor)(x, y) \wedge wikibase : Statement(y) \wedge eps : P32(hasRaceorColor)(y, z)$$

$$enslaved : Person(w) \wedge enslaved : hasAgeRecord(w, x) \wedge enslaved : AgeRecord(x) \wedge enslaved : hasAgeValue(x, z) \leftrightarrow ed : Q410(Person)(w) \wedge ep : P42(hasAge)(w, x) \wedge wikibase : Statement(x) \wedge eps : P42(hasAge)(x, y) \wedge ed : Q424(AgeRecord)(y) \wedge edt : P3(hasAgeValue)(y, z)$$

$$enslaved : Person(x) \wedge enslaved : hasInterAgentRelationshipRecord(x, y) \wedge enslaved : InterAgentRelationshipRecord(y) \wedge enslaved : hasInterAgentRelationshipType(y, z) \wedge enslaved : InterAgentRelationshipTypes(z) \leftrightarrow ed : Q410(Person)(x) \wedge ep : P39(hasInterAgentRelationshipTypeTo)(x, y) \wedge wikibase : Statement(y) \wedge eps : P39(hasInterAgentRelationshipTypeTo)(y, z) \wedge ed : Q463(InteragentRelationship)(z)$$

$$enslaved : Person(x) \wedge enslaved : hasParticipantRoleRecord(x, y) \wedge enslaved : ParticipantRoleRecord(y) \wedge enslaved : roleProvidedBy(y, z) \wedge enslaved : Event(z) \leftrightarrow ed : Q410(Person)(x) \wedge ep : P17(hasParticipantRole)(x, y) \wedge wikibase : Statement(y) \wedge epq : P19(roleProvidedBy)(y, z) \wedge ed : Q238(Event)(z)$$

$$enslaved : Person(w) \wedge enslaved : hasNameRecord(w, x) \wedge enslaved : NameRecord(x) \wedge enslaved : hasPreferredNameVariant(x, y) \wedge enslaved : NameVariant(y) \wedge enslaved : fullNameAsString(y, z) \leftrightarrow ed : Q410(Person)(w) \wedge ep : P20(hasName)(w, x) \wedge wikibase : Statement(x) \wedge eps : P20(hasName)(x, z)$$

Alignment Task

- Given a left-hand side of a benchmark alignment rule, find all right-hand side predicate names.
  - This doesn't assemble the rule itself, but is extremely helpful to support a human doing the alignment.

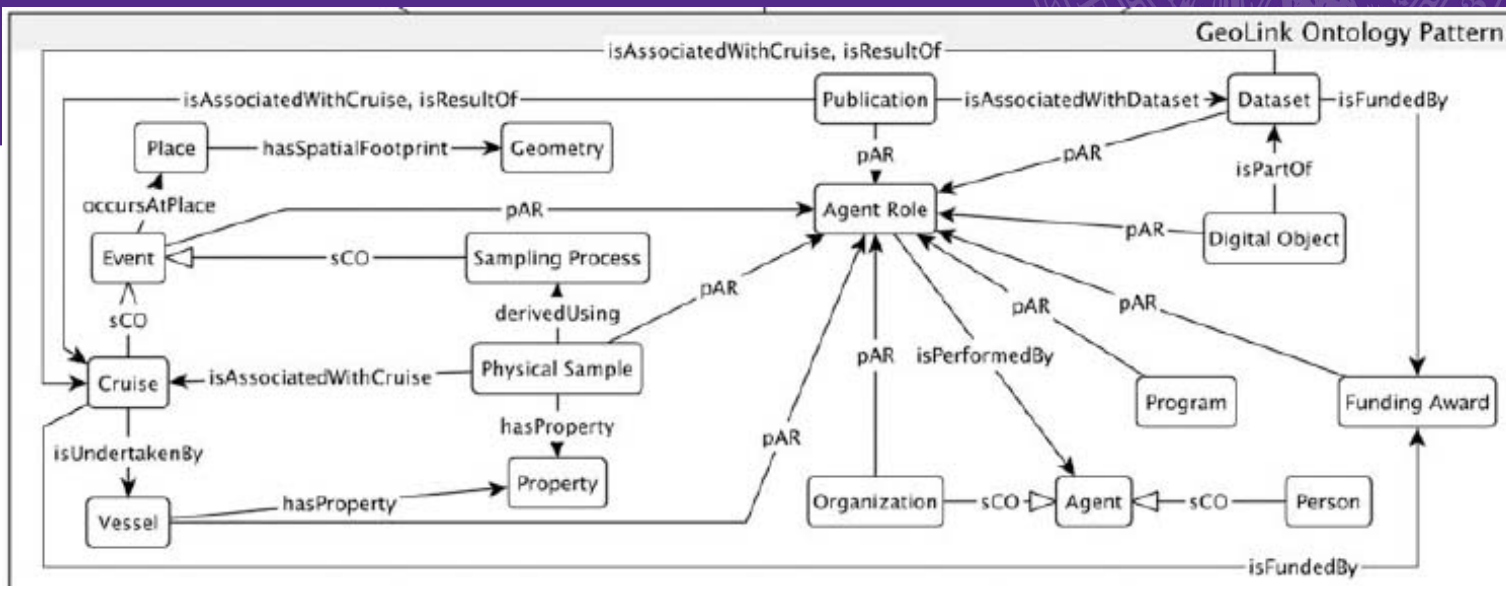
$$\texttt{Award}(x) \wedge \texttt{hasCoPrincipalInvestigator}(x, z) \leftrightarrow$$
$$\texttt{FundingAward}(x) \wedge \texttt{providesAgentRole}(x, y)$$
$$\wedge\, \texttt{CoPrincipalInvestigatorRole}(y) \wedge \texttt{performedBy}(y, z)$$

$$\texttt{enslaved:Person}(x) \wedge \texttt{enslaved:hasAgeRecord}(x, y) \wedge \texttt{enslaved:AgeRecord}(y)$$
$$\leftrightarrow \texttt{ed:Q410(Person)}(x) \wedge \texttt{ep:P42(hasAge)}(x, y) \wedge \texttt{wikibase:Statement}(y)$$

- Direct prompting (full ontologies) didn't work at all.

Modules

GeoLink:



Cogan Shimizu, Karl Hammar, Pascal Hitzler,
Modular Ontology Modeling.
Semantic Web 14 (3), 459-489, 2023.

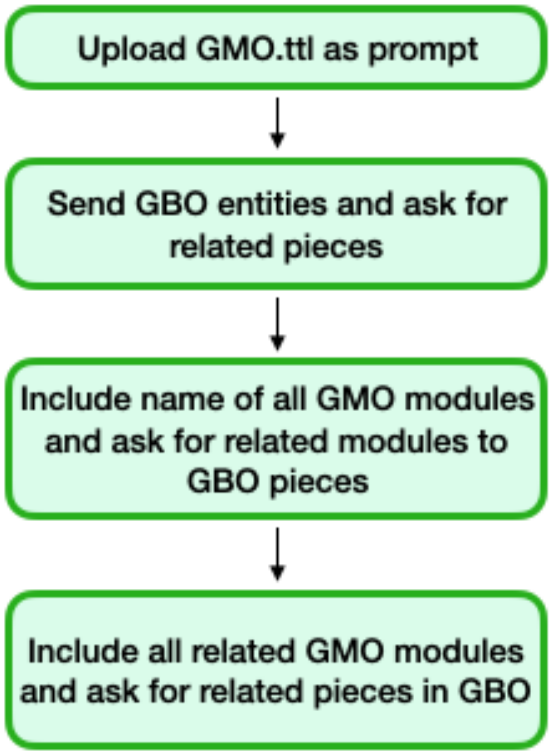
Enslaved modules:

ExternalReference **Description** **OriginRecord**
RaceRecord **AgeRecord** **SexRecord**
PersonStatusRecord **Provenance** **OccupationRecord**
ResearchProject **Event** **ParticipantRoleRecord**
Place **TemporalExtent** **NameRecor**
InterAgentRelationshipRecord

Complex ontology alignment

Reihaneh Amini, Sanaz Saki Norouzi, Pascal Hitzler, Reza Amini, Towards Complex Ontology Alignment using Large Language Models. KGSWC 2024.

Adrita Barua, Reza Amini, Sanaz Saki Norouzi, Reihaneh Amini, Pascal Hitzler, Complex Ontology Alignment using LLMs: a case study. In: OM 2025, to appear.



- **Direct prompting (full ontologies) didn't work at all for GeoLink.**
- **Modularized approach (right) did work like a charm.**

	Recall			Precision		
	≥ 0.5	≥ 0.75	$= 1$	≥ 0.5	≥ 0.75	$= 1$
Enslaved Entities	72.0%	51.0%	37.0%	69.0%	43.0%	33.0%
GMO Entities	73.3%	62.3%	45.0%	69.7%	59.6%	45.8%

Curveballs

(very new results)

- Enslaved, 98 alignment rules.
- With modularization:

recall > 0.5	precision >0.5
72	69
recall > 0.75	precision >0.75
51	43
recall==`1	precision ==1
37	33

without modularization

recall > 0.5	precision >0.5
73	83
recall > 0.75	precision >0.75
54	54
recall==`1	precision ==1
39	47

- Data leakage?
- Was LLM improving between runs?
- Size difference GeoLink/Enslaved.
- Poor performance on GeoLink without modularization was quite some time ago; LLMs may have generally improved?

Modularization

- Finding modules in a monolithic ontology is hard, because of scaling.
- Can we use LLMs?
- [Sen Sarma 2026] preliminary results indicate: we probably can, to some extent
- ChatGPT 5.2, Input: lists of class and property names.
- Ouptput: The following list of modules with reasonable assignment of entities to modules:
  - Core Agents & Roles
  - Agent Records & Identity Attributes
  - Inter-Agent Relationships
  - Event & Event Classification
  - Places, Place Typing, Spatial Hierarchy
  - Geospacial Geometry & Spatial Extent
  - Temporal Extent & Time Spans
  - Provenance & Source Traceability
  - Licensing & Rights Information
  - Descriptions & Free-Text Content
  - Matching & Entity Resolution
  - Research Projects

Other work

- **LLM generated ontology design patterns for ontology construction.**
  - **Andrew Eells, Brandon Dave, Pascal Hitzler, Cogan Shimizu, Commonsense Ontology Micropatterns. In: Proceedings NeSy 2024.**

- **LLM-based ontology population.**
  - **S. S. Norouzi, A. Barua, A. Christou, N. Gautam, A. Eells, P. Hitzler, C. Shimizu. Ontology population using LLMs. In: Handbook on Neurosymbolic AI and Knowledge Graphs. Frontiers of Artificial Intelligence and Applications vol. 400, IOS Press, Amsterdam, 2025, pp. 421-440.**
  - **A. Saini, CD Jaldi, J. Ethier, C. Shimizu, Research Directions for Ontology-Guided Domain-Specific Knowledge Graph Population Using LLMs. In: Proceedings KGSWC 2025.**

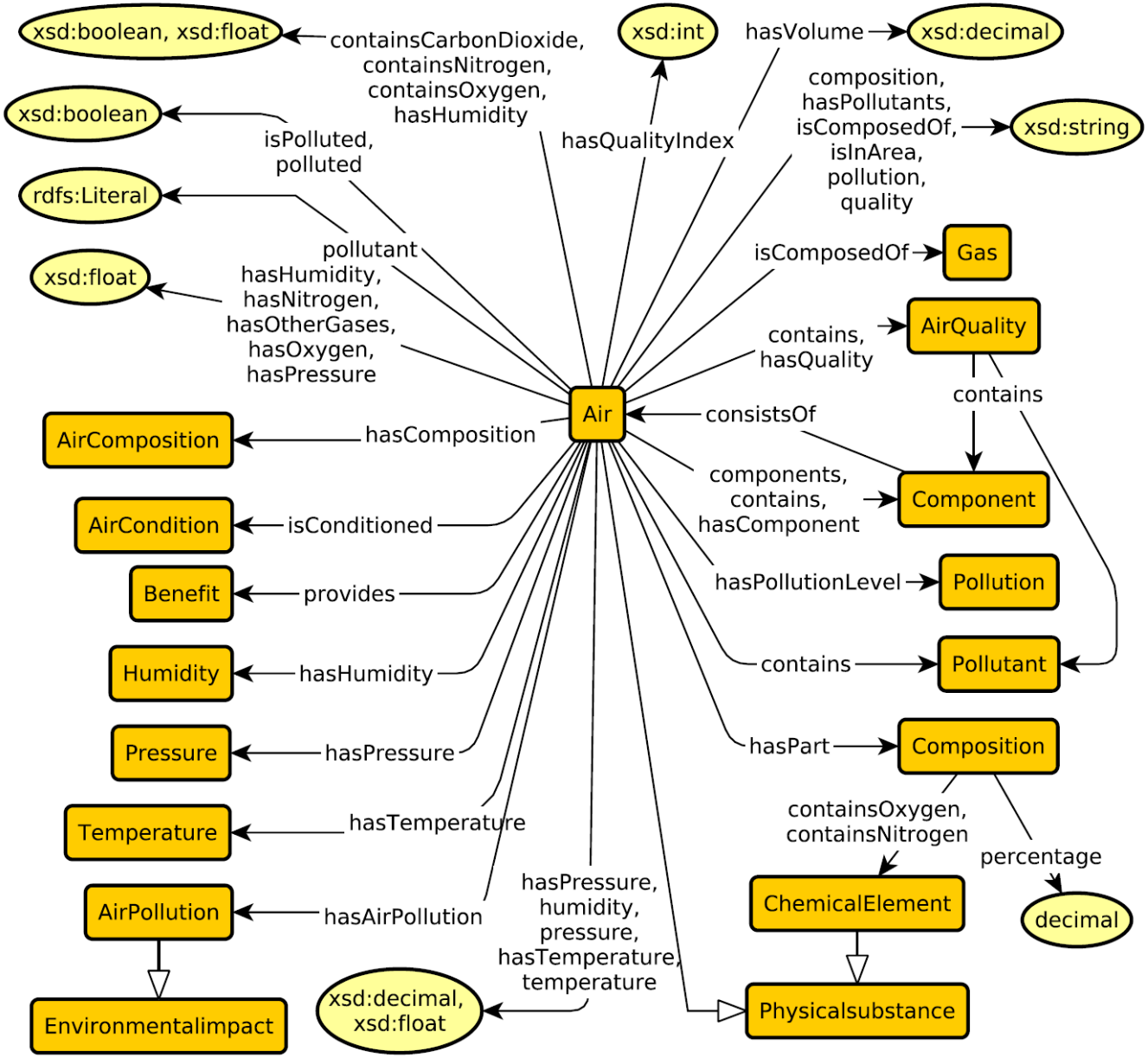
MOMo workflow

1. Describe use cases and gather possible data sources.
2. Gather competency questions.
3. Identify key notions for the domain to be modeled.
4. Identify existing ontology design patterns to be used.
5. Create schema diagrams for modules.
6. Set up documentation and determine axioms for each module.
7. Create ontology schema diagram from the module schema diagrams.
8. Add axioms spanning more than one module.
9. Reflect on entity naming and all axioms.
10. Create OWL file(s).

- **Core steps 3 to 5 (or 2 to 6) amenable to LLM-ing**
- **Requires good ODP libraries!**
  - **Let's make them with LLMs?**

Cogan Shimizu, Karl Hammar, Pascal Hitzler, Modular Ontology Modeling. Semantic Web 14 (3), 459-489, 2023.

LLM based micropatterns



Andrew Eells, Brandon Dave, Pascal Hitzler, Cogan Shimizu, Commonsense Ontology Micropatterns. In: Neural-Symbolic Learning and Reasoning -- 18th International Conference, NeSy 2024. Lecture Notes in Computer Science 14980, Springer 2024.

Ontology Population

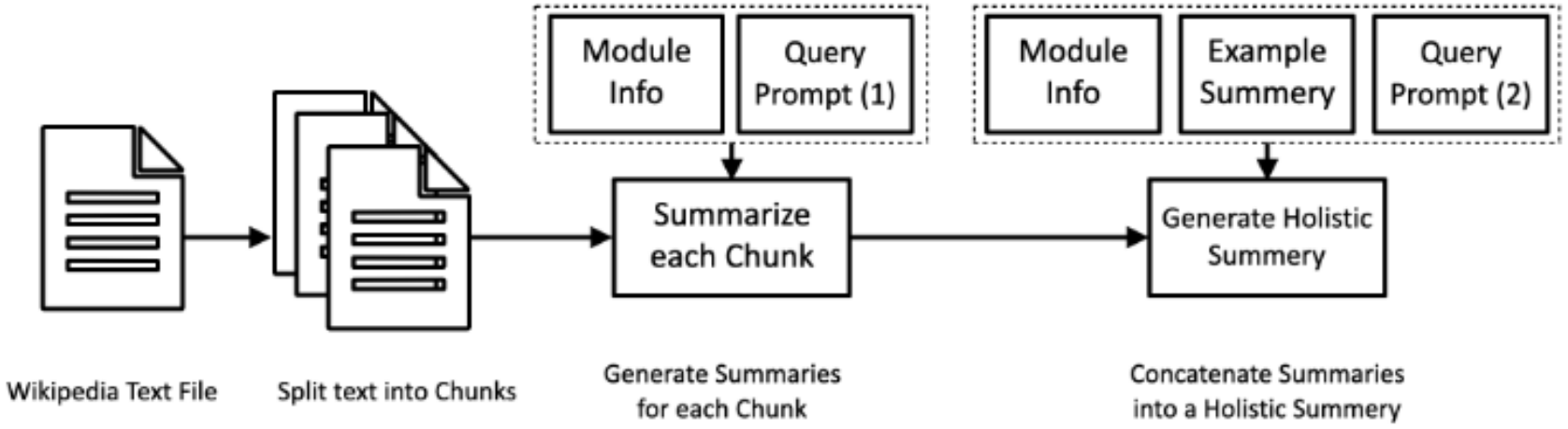
- **Enslaved ontology (both versions)**
- **Input: Unstructured text (from Wikipedia, on persons)**
- **Evaluation: Comparison with triples from persons already on enslaved.org that came from trusted data sources**

Sanaz Saki Norouzi, Adrita Barua, Antrea Christou, Nikita Gautam, Andrew Eells, Pascal Hitzler, Cogan Shimizu, Ontology Population using LLMs. In: Pascal Hitzler, Abhilekha Dalal, Mohammad Saeid Mahdavinejad, Sanaz Saki Norouzi (eds.), Handbook on Neurosymbolic AI and Knowledge Graphs. Frontiers of Artificial Intelligence and Applications vol. 400, IOS Press, Amsterdam, 2025, pp. 421-440.

Approaches

- **Modules used to focus on specific parts of the ontology**

- **Text summarization used on Wikipedia texts to deal with prompt size**



- **Prompt structure: Instruction + Ontology Module + Raw Text**
- **Looked at different approaches, including RAG.**

Evaluation

LLM Model	Avg %	Ttl %	#F	Avg_A %	Ttl_A %
GPT4_Enslaved_MainAgent	82.30	81.60	14	88.55	88.09
GPT4_Enslaved_notrestrictedToMAgent	87.87	86.82	4	89.11	88.69
GPT4_WB_MainAgent	77.08	76.10	25	88.02	87.63
GPT4_WB_notrestrictedToMAgent	85.03	84.12	7	87.69	87.34
GPT-4_Summarization_Enslaved	81.16	80.86	0	81.16	80.86
GPT-4_Summarization_WB	82.60	82.03	0	82.60	82.03
llama_WB_MainAgent	71.17	70.88	6	73.64	73.20
llama_WB_notrestrictedToMAgent	71.25	70.63	2	71.92	71.41

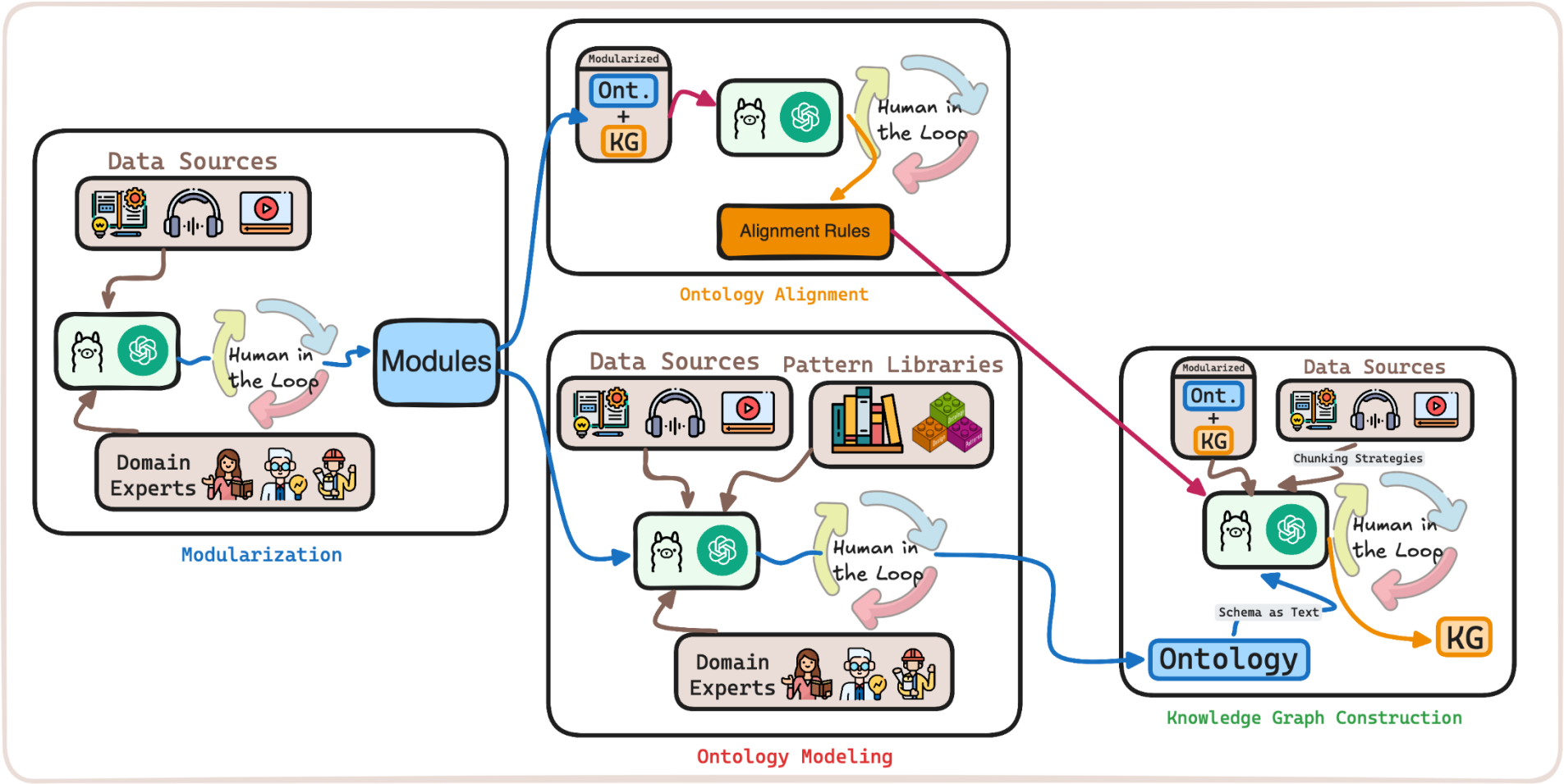
Table 1. Similarity matching between triples extracted by the LLM models plus prompting strategy, reporting the average and total coverage for each module and in aggregate (columns with subscript *A*). The #F column indicates the number of modules for which triple extraction entirely failed (i.e., a coverage of 0).

The extraction of triples is generally effective, approaching 90% coverage.

Conclusions

- LLM-based KGOE is a very promising and exciting direction at the moment.
- We posit that modularity may be a key ingredient. Our preliminary investigations support that.
- Generally, we should spend more time investigating how common ontology and KG paradigms could be *easily improved* to make hard KGOE tasks easier.
- Neurosymbolic slack, over 1,300 researchers, email me at hitzler@ksu.edu to join.

LLM-based KGOE



Concept, Ideas and image credit: Cogan Shimizu, Adrita Barua, Pascal Hitzler (unpublished)

TOC

- Deep Deductive Reasoning
- Ontology-based Hidden Neuron Activation Analysis
- LLM-based Knowledge Graph and Ontology Engineering
- Some musings on human-AI neurosymbolic agentic whatever
  - Vaccaro, Almaatouq, Malone, When combinations of humans and AI are useful: A systematic review and meta-analysis. Nature Human Behaviour 8, 2024, pp. 2293-2303. (NHB)
  - Ilievski et al., Aligning Generalisation Between Humans and Machines. Nature Machine Intelligence 7, 1378-1389, 2025. (NMI)

Human-AI teaming (NHB paper)

- Meta-analysis – comparisons between
  - Human-AI team
  - Human alone
  - AI alone

- Human-AI is better than human alone
- Human-AI is worse than max(Human alone, AI alone)
- Human-AI is worse than AI alone in about 48% of studies

Statistically,

- If the AI is better than the human, then human-AI is worse than AI alone
- If the human is better than the AI, then human-AI is better than human alone and better than AI alone

Human-AI teaming (NHB paper)

- Based on 370 publications between Jan 2020 and June 2023.
  - Apply with caution.

- How to design human-AI teaming to maximize performance? Perhaps
  - Give only subtasks to the AI where the AI is clearly better than the human?
  - Let the more capable partner decide on task allocation?
  - The *process* seems to be decisive.

- Note that study only focused on outcome quality. There are other aspects of practical importance, e.g.
  - Time saved
  - Resources/costs

Particular human strengths compared to ML

- Generalizing from a few examples
- Excel at compositionality
- Robust generalization to
  - noise
  - shifts
  - OOD data

Human-AI teaming aims at each side addressing the other side's weaknesses.

(Nature MI paper)

Hybrid AI

- Neural
  - Generalizes within bounds
  - Quick (after training)
  - Makes mistakes
  - Some flexible control of task allocation possible
- Symbolic
  - Provable correctness
  - Slow
  - Can’t learn
  - Excels on pre-defined workflows
- Human
  - Very slow
  - Superior task allocation in new situations
  - Excellent at OOD tasks
  - Can deal with shifts, exceptions
  - Expert correctness checking and intervention when needed

Thanks!