NEURO-SYMBOLIC AI

A SYNERGISTIC INTEGRATION OF KNOWLEDGE REPRESENTATION AND MACHINE LEARNING

ALESSANDRO OLTRAMARI
BOSCH RESEARCH AND TECHNOLOGY CENTER
PITTSBURGH, USA

BOSCH

Introduction
About Me

- Senior Research Scientist at Bosch Center for Artificial Intelligence in Pittsburgh, USA
- PhD in Cognitive Science from University of Trento (Italy)

- Postdoc at Carnegie Mellon University (USA)
- Visiting Scientist at Princeton (USA)
- Associated Researcher at the Italian National Research Council (Italy)

- Core expertise: KRR, Computational Linguistics, Cognitive Architectures, Neuro-symbolic AI

Introduction
About Bosch Research in North America

North America

Research and Technology Center

Technology scouting in America and research in the areas of

- Human-Machine Intelligence
- Modeling, Design and Control of Energy Systems and Materials
- Secure and Intelligent IoT[1]
- Intelligent and Connected Sensors and Systems

[1]Internet of Things (IoT)

Introduction
About Bosch Research in Pittsburgh

Creating solutions ensuring the right people have the right access to the right data at the right time

- → **Secure + Private Cloud Computing**
- → **Connected Vehicle Security**

Applications and infrastructure that deliver innovative UX reliably and at cloud-scale

- → **Intelligent Audio Analytics**
- → **IoT App Ecosystems**

Leveraging social network technology to create productive links between communities and machines

- → **AI Knowledge + Training Platform**
- → **“Common Sense” Machine Intelligence**

BOSCH

Agenda

1. Making the Case for Neuro-Symbolic AI
2. Knowledge Graphs
3. Neuro-symbolic Commonsense Question Answering

MAKING THE CASE FOR NEURO-SYMBOLIC AI

Making the case for Neuro-Symbolic AI
Myth: "AI will make humans obsolete"

What Are the Risks of Using AI in Finance?

At the moment, most organizations cannot afford premium AI applications. The high-end technology is too expensive for the majority of fintech businesses out there, at least for now.

Yet, we are on the verge of entering a new decade, and things could rapidly change. Nonetheless, many renowned experts have issued warnings about the dangerous nature of artificial intelligence.

Of course, the primary axiom of these claims is that AI will make humans obsolete. Once computers develop their own intelligence, they will be unstoppable.

Nonetheless, the situation out in the trenches shows that this is nothing more than a myth. No matter how complex the algorithms are, they cannot copy our common sense. In other words, human intuition remains an elusive ingredient that makes the difference between robots and the human race.

https://fpa-trends.com/article/pros-and-cons-using-ai-applications-finance-sector

Making the case for Neuro-Symbolic AI
Reality: Idiot Savants?

- Human:

I want some birthday cake. I should…

- GPT-3

..go to a more dangerous place and get arrested

BOSCH

MIT Technology Review

175B parameters

Wikipedia (6M articles) = 1% of training set

Artificial intelligence / Machine learning

OpenAI's new language generator GPT-3 is shockingly good—and completely mindless

The AI is the largest language model ever created and can generate amazing human-like text on demand but won't bring us closer to true intelligence.

by **Will Douglas Heaven** July 20, 2020

Internal | CR/AEU2 | 2019-06-03

BOSCH

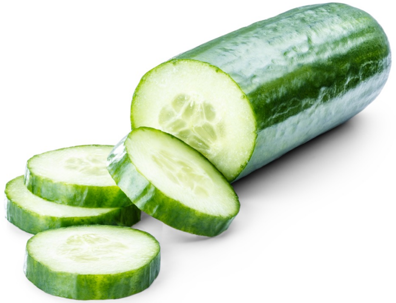
"One vegetable that is about as big as your head is…."

corn

gooseberries

lettuce

cucumber



carrot

spinach

pumpkin

cauliflower

beet

BOSCH

Making the Case for Neuro-Symbolic AI
Common Analogic Reasoning

ProtoQA (2020)

- The example is from ProtoQA task (inspired by the popular American TV game “Family Feud”)
- Humans are quite consistent in picking candidates like “pumpkin”, “cauliflower”, “cabbage”
- Neural language models…not so much
  - **GPT3** [“broccoli”, “carrot”, “cabbage”, “cauliflower”, “carrots”, “lettuce”, “spinach”, “cucumber”, “potato”, “corn”]
  - **BERT** [“cucumber”, “broccoli”, “beet”, “carrot”, “lettuce”, “pumpkin”, “spinach”, “cabbage”, “cauliflower”, “celery”]
- Models learn some properties of vegetables from training data, but not to compare “size”
  - Lack of analogical reasoning (Ushio 2021)
  - How about other forms of reasoning?
    - Ettinger (2020) diagnoses the BERT model, finding that it struggles with complex inference, role-based event prediction, and grasping the contextual impacts of negation.

- Ushio, Asahi, Luis Espinosa-Anke, Steven Schockaert, and Jose Camacho-Collados. "BERT is to NLP what AlexNet is to CV: Can Pre-Trained Language Models Identify Analogies?." *arXiv:2105.04949* (2021).
- *Ettinger, Allyson. "What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models." Transactions of the Association for Computational Linguistics 8 (2020): 34-48.*

BOSCH

Making the Case for Neuro-Symbolic AI
Common Temporal Reasoning

bAbI (2015)

- Bill went back to the cinema yesterday.
- Julie went to the school this morning.
- Fred went to the park yesterday.
- Yesterday Julie went to the office.

  Q: Where was Julie before the school?

  A: Office

Weston, Jason, Antoine Bordes, Sumit Chopra, Alexander M. Rush, Bart van Merriënboer, Armand Joulin, and Tomas Mikolov. "Towards ai-complete question answering: A set of prerequisite toy tasks." *arXiv preprint arXiv:1502.05698* (2015).

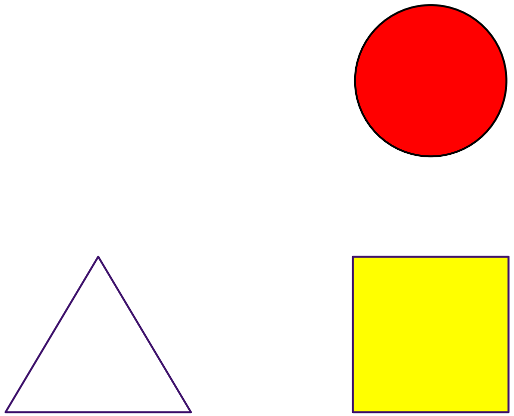
BOSCH

Making the Case for Neuro-Symbolic AI
Common Spatial Reasoning

bAbI
(2015)

- **The red sphere is above the yellow square.**
- **The triangle is to the left of the yellow square.**

Q	Is the triangle above the red sphere?	no
Q	Is the triangle to the left of the red sphere?	yes
Q	Is the red sphere above the triangle?	yes
Q	Is the red sphere to the left of the triangle?	no
Q	Is the triangle below the red sphere?	yes
Q	Is the triangle to the right of the red sphere?	no



Weston, Jason, Antoine Bordes, Sumit Chopra, Alexander M. Rush, Bart van Merriënboer, Armand Joulin, and Tomas Mikolov. "Towards ai-complete question answering: A set of prerequisite toy tasks." *arXiv preprint arXiv:1502.05698* (2015).

Making the Case for Neuro-Symbolic AI
Common Reasoning with Negation

ProtoQA (2020)

- “One item of clothing that you would **not** lend to one person is”
  - "correct": ["shoes", "underwear", "jeans"], "wrong": ["socks", "sweater"," "hat", "pants"]}
- One activity that’s **often** depicted in movies though people **seldom** do it in real life is",
  - "correct": ["fighting"],
  - "wrong": ["dancing", "swimming", "running", "wrestling", "bowling", "sports", "boxing", "surfing"}

Boratko, Michael, Xiang Lorraine Li, Rajarshi Das, Tim O'Gorman, Dan Le, and Andrew McCallum. "ProtoQA: A Question Answering Dataset for Prototypical Common-Sense Reasoning." arXiv:2005.00771 (2020).

BOSCH

Making the Case for Neuro-Symbolic AI
Common Social Reasoning

SocialIQA (2019)

Reasoning about Motivation

Tracy had accidentally pressed upon Austin in the small elevator and it was awkward.

Q Why did Tracy do this?

A (a) get very close to Austin
(b) squeeze into the elevator ✓
(c) get flirty with Austin

Reasoning about what happens next

Alex spilled the food she just prepared all over the floor and it made a huge mess.

Q What will Alex want to do next?

A (a) taste the food
(b) mop up ✓
(c) run around in the mess

Reasoning about emotional reactions

In the school play, Robin played a hero in the struggle to the death with the angry villain.

Q How would others feel afterwards?

A (a) sorry for the villain
(b) hopeful that Robin will succeed ✓
(c) like Robin should lose

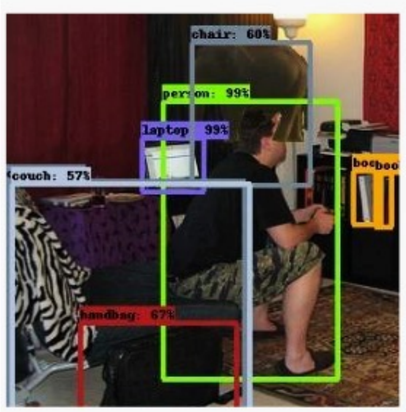
Sap, Maarten, Hannah Rashkin, Derek Chen, Ronan LeBras, and Yejin Choi. "Socialiqa: Commonsense reasoning about social interactions." *arXiv preprint arXiv:1904.09728* (2019).

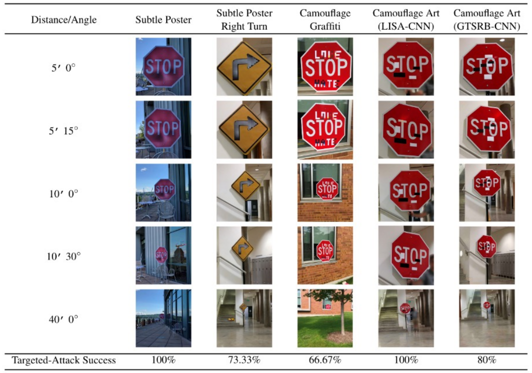
Making the Case for Neuro-Symbolic AI
Beyond Language: Perception and Decision

- Learning algorithms are not sufficient to enable robust AI

Perception

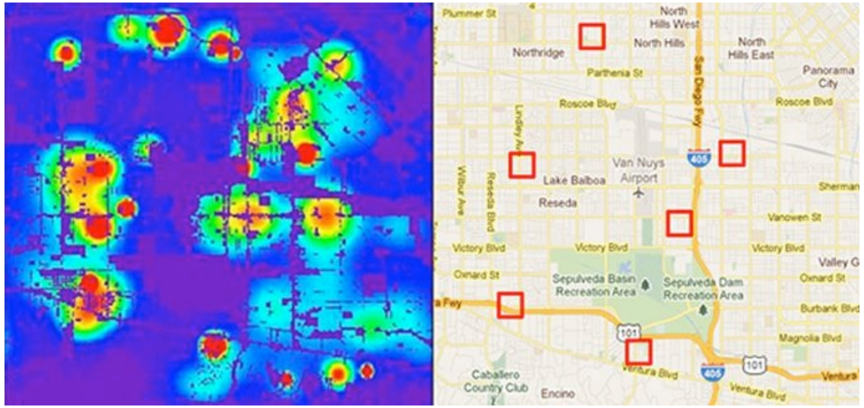
- Visual classification fails by modifying image sub-regions
  - *Rosenfeld, A., Zemel, R., & Tsotsos, J. K. (2018). The elephant in the room. arXiv preprint arXiv:1808.03305.* [*arXiv:1808.03305*](arXiv:1808.03305)
  - *Evtimov, I., Eykholt, K., Fernandes, E., Kohno, T., Li, B., Prakash, A., & Song, D. (2017). Robust physical-world attacks on deep learning models.* *arXiv preprint arXiv:1707.08945 1 (2017):* *1.*





Decision

- Prediction models are vulnerable to biased datasets
  - *Eubanks, V. (2018). A child abuse prediction model fails poor families.* *Wired Magazine* (01.15.18)
  - *Verghese, A., Shah, N. H., & Harrington, R. A. (2018). What this computer needs is a physician: humanism and artificial intelligence. Jama, 319(1), 19-20.*
  - *Tashea, J. (2017). Courts are using AI to sentence criminals. That must stop now.* *Wired Magazine*.
  - Deshmukh, S., & Annappa, B. (2019). Prediction of Crime Hot Spots Using Spatiotemporal Ordinary Kriging. In *Integrated Intelligent Computing,*



Internal | CR/AEU2 | 2019-06-03

Fact:
Neural models lack of common forms of human reasoning

General Research Question:
How do we endow neural models with human-like common (sense) knowledge and reasoning?

Specific Research Question:
How do we endow neural models applied to finance with relevant commonsense knowledge

Fact:
Neural models lack of common forms of human reasoning

General Research Question:
How do we endow neural models with human-like common (sense) knowledge and reasoning?

Specific Research Question:
How do we endow neural models applied to finance with relevant commonsense knowledge

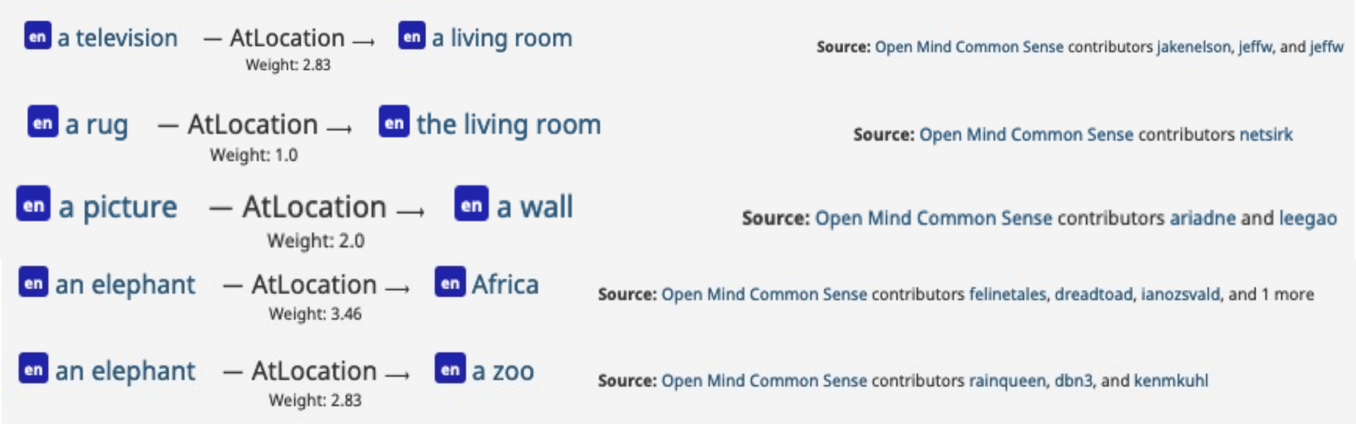
Making the Case for Neuro-Symbolic AI
Machine Common Sense **Knowledge**

- **Human Commonsense**
  - Massive body of tacit knowledge derived from experience
  - Basis for effective and efficient communication

- **Machine Commonsense**
  - Represented in knowledge bases and knowledge graphs
    - Challenge: completeness
      - ..but: domain-specific KGs are completable

Elephants don't inhabit living rooms

`AtLocation(Elephant,Living Room) = FALSE`[ConceptNet]



Alessandro Oltramari (CR/RTC3.1-NA) 2020-10-20

BOSCH

Making the Case for Neuro-Symbolic AI
Machine Common sense **Reasoning**

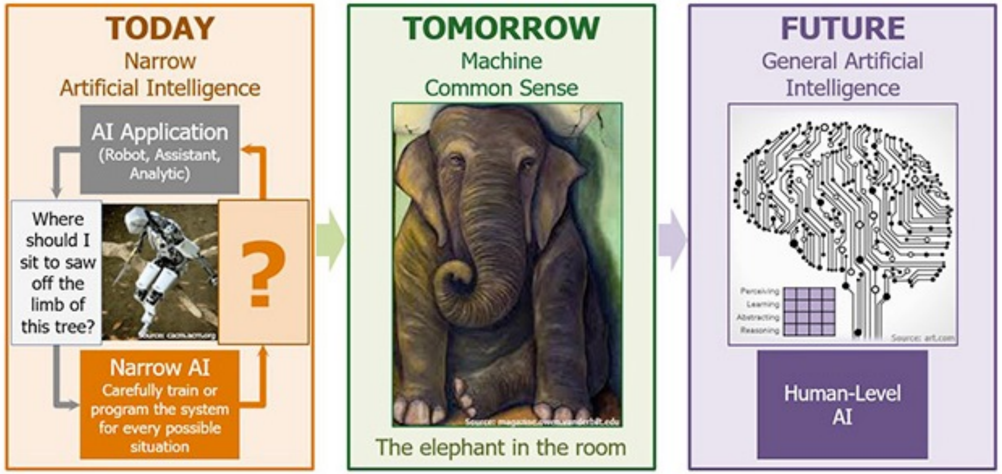
- "MCS can enable AI systems to understand situations, perform reasonable actions, communicate more effectively with people, transfer learning to new domains"

- *At the machine level*, CS has been traditionally articulated and encoded by using symbolic languages (e.g., Cyc project, MIT's ConceptNet)
  - Pro: expressive and suitable for automatic reasoning
  - Con: neither scalable nor comprehensive
  - New approach: hybrid AI, ***neuro-symbolic*** architectures that leverage deep learning methods and symbolic representations and reasoning

Teaching Machines Common Sense Reasoning

DARPA program seeks to articulate and encode humans' basic background knowledge for intelligent systems

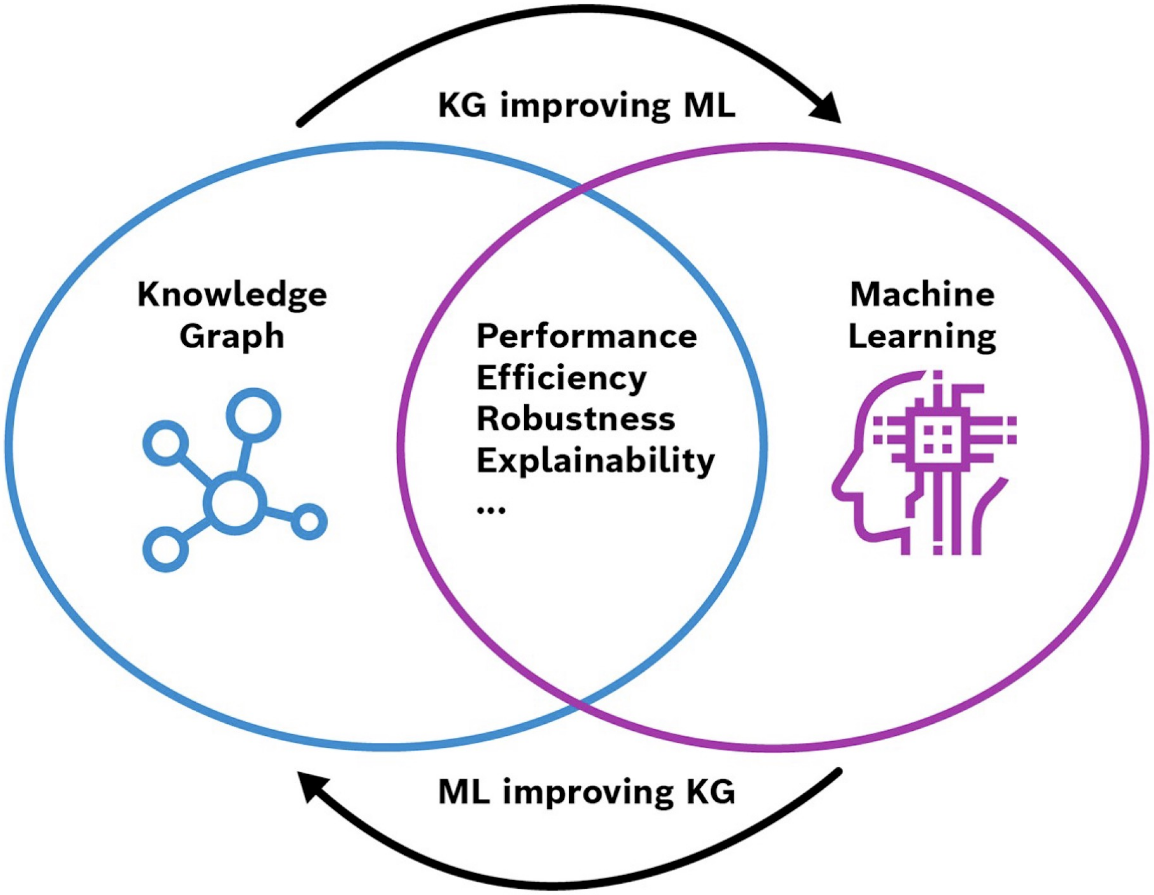
OUTREACH@DARPA.MIL
10/11/2018



Introduction
Overcoming AI Limitations with Neuro-Symbolic Methods

- Neuro-symbolic (NS) AI is a synergistic integration of knowledge representation and reasoning (KRR) and machine learning (ML) leading to improvements in scalability, efficiency, and explainability.
- NS methods stem from a pragmatic approach, which can be distilled into 3 core observations:
  1. In real-world applications, it is often impractical and inefficient to learn all relevant facts and data patterns from scratch, especially when prior knowledge is available.
  2. KRR systems utilize well-formed axioms and rules, which guarantees explainability both in terms of asserted and inferred knowledge (a hard-to-satisfy requirement for neural systems).
  3. Neural systems make data-driven algorithms scalable and automatic knowledge construction viable.
- Over the last few years, NS-AI has been flourishing thanks to new methods for integrating deep learning and knowledge graph technologies.
- A variety of downstream tasks, including question answering, robot navigation, etc., have been instrumental in benchmarking NS methods and evaluating datasets in a rigorous and reproducible way.

Introduction
Synergy between KRR and ML



Introduction
Neuro-symbolic AI projects at Bosch Research

BLOG POST | Research

Neuro-symbolic AI for scene understanding >

https://www.bosch.com/stories/neuro-symbolic-ai-for-scene-understanding

BLOG POST | Research

Assisting the technical workforce with Neuro-symbolic AI >

https://www.bosch.com/stories/assisting-the-technical-workforce-with-neuro-symbolic-ai

BLOG POST | Research

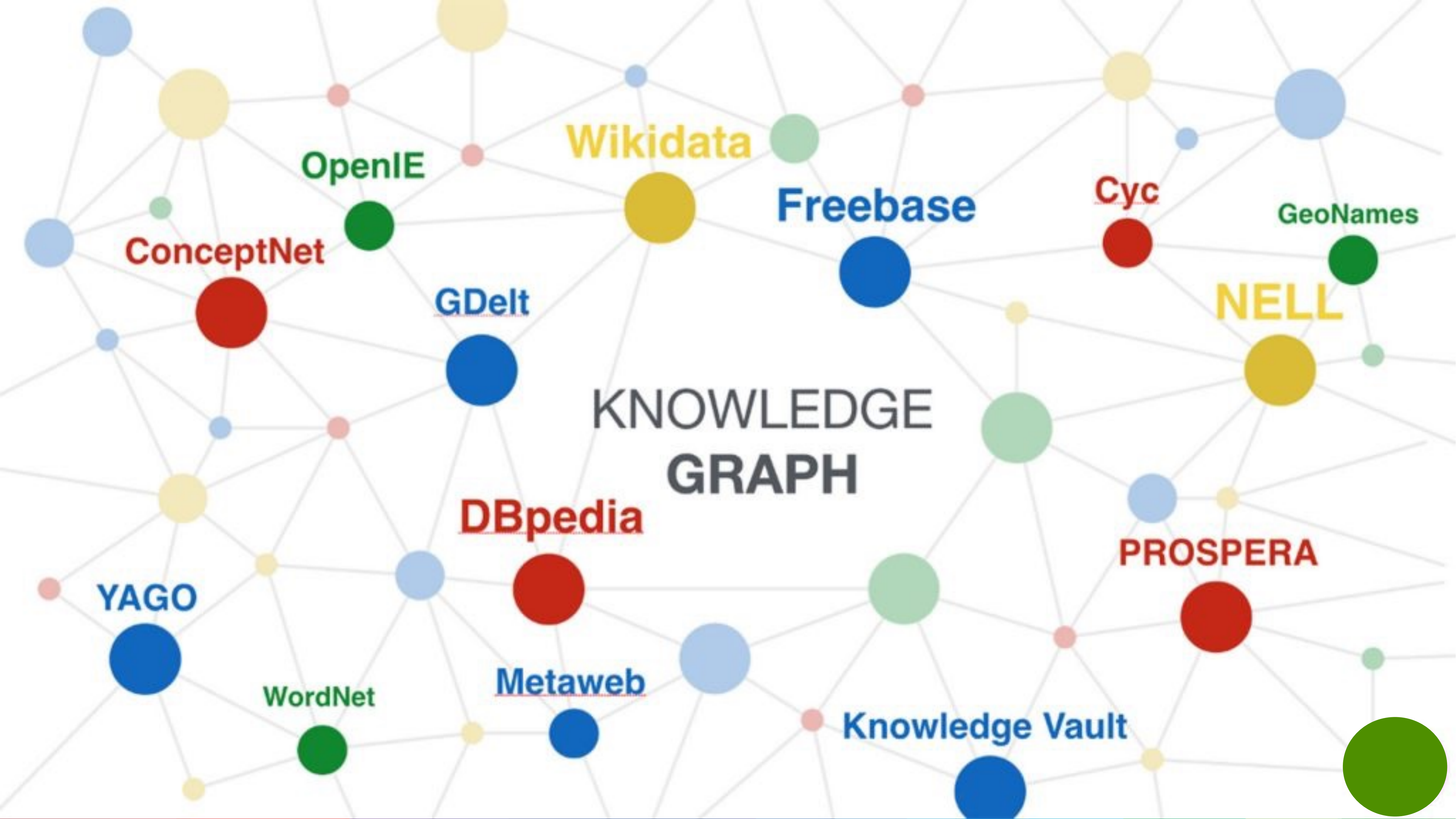
First workshop on Common Sense Knowledge Graphs >

https://www.bosch.com/stories/first-workshop-on-common-sense-knowledge-graphs

BOSCH

KNOWLEDGE GRAPHS

(THE –“SYMBOLIC” IN “NEURO-SYMBOLIC”)

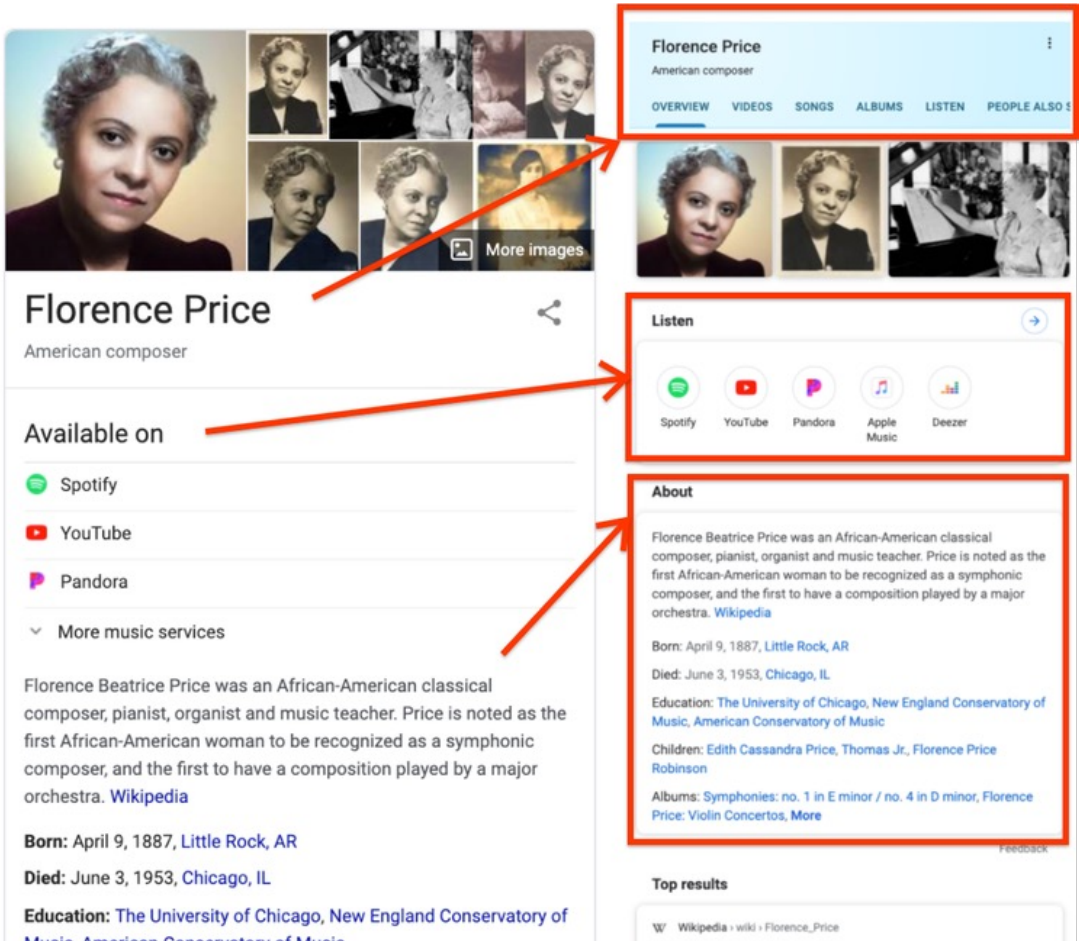


Wikidata
OpenIE
Freebase
Cyc
GeoNames
ConceptNet
GDelt
NELL
KNOWLEDGE
GRAPH
DBpedia
PROSPERA
YAGO
WordNet
Metaweb
Knowledge Vault

Knowledge Bases and Knowledge Graphs
Google Knowledge Graph

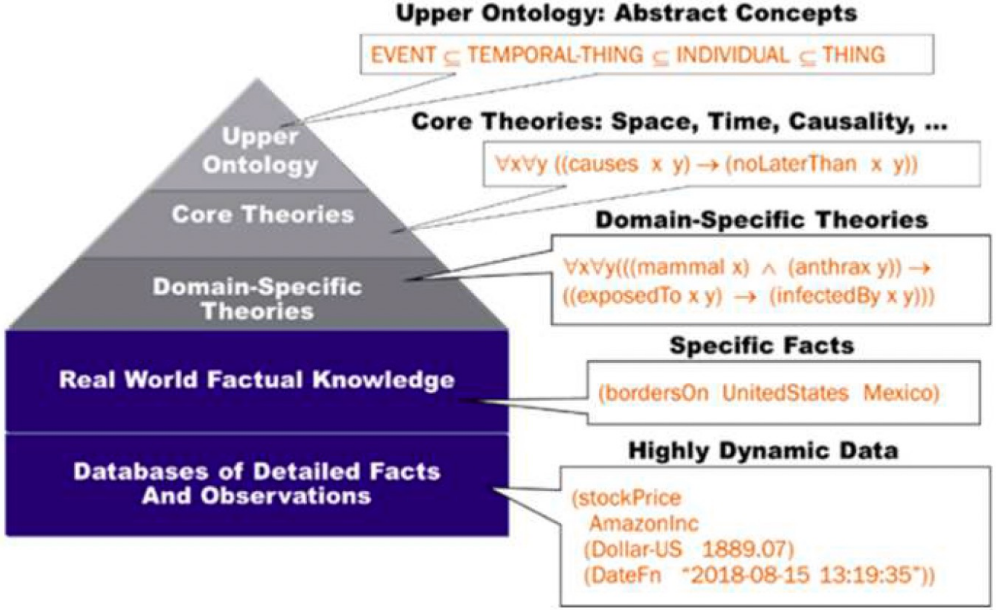
- **570 million entities and 18 billion facts***

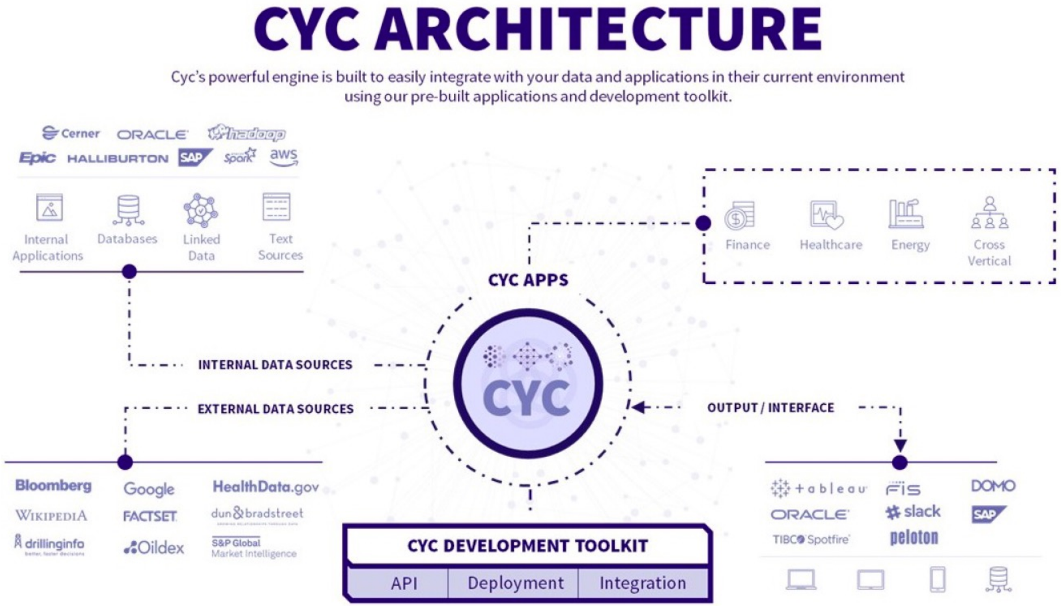
*Source: Wikipedia (no specific publications about the G-KG, only about using the G-KG for specific applications



BOSCH

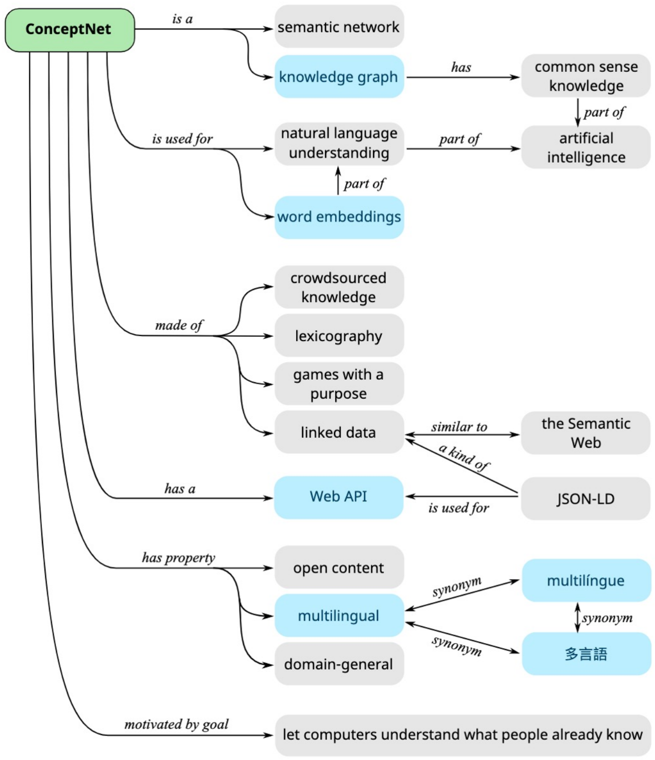
Knowledge Graphs
Cyc (old-school KB, not KG)





Matuszek, Cynthia, Michael Witbrock, Robert C. Kahlert, John Cabral, Dave Schneider, Purvesh Shah, and Doug Lenat. "Searching for common sense: Populating cyc from the web." UMBC Computer Science and Electrical Engineering Department Collection (2005).

Knowledge Bases and Knowledge Graphs
ConceptNet and ATOMIC



Declarative VS Procedural knowledge

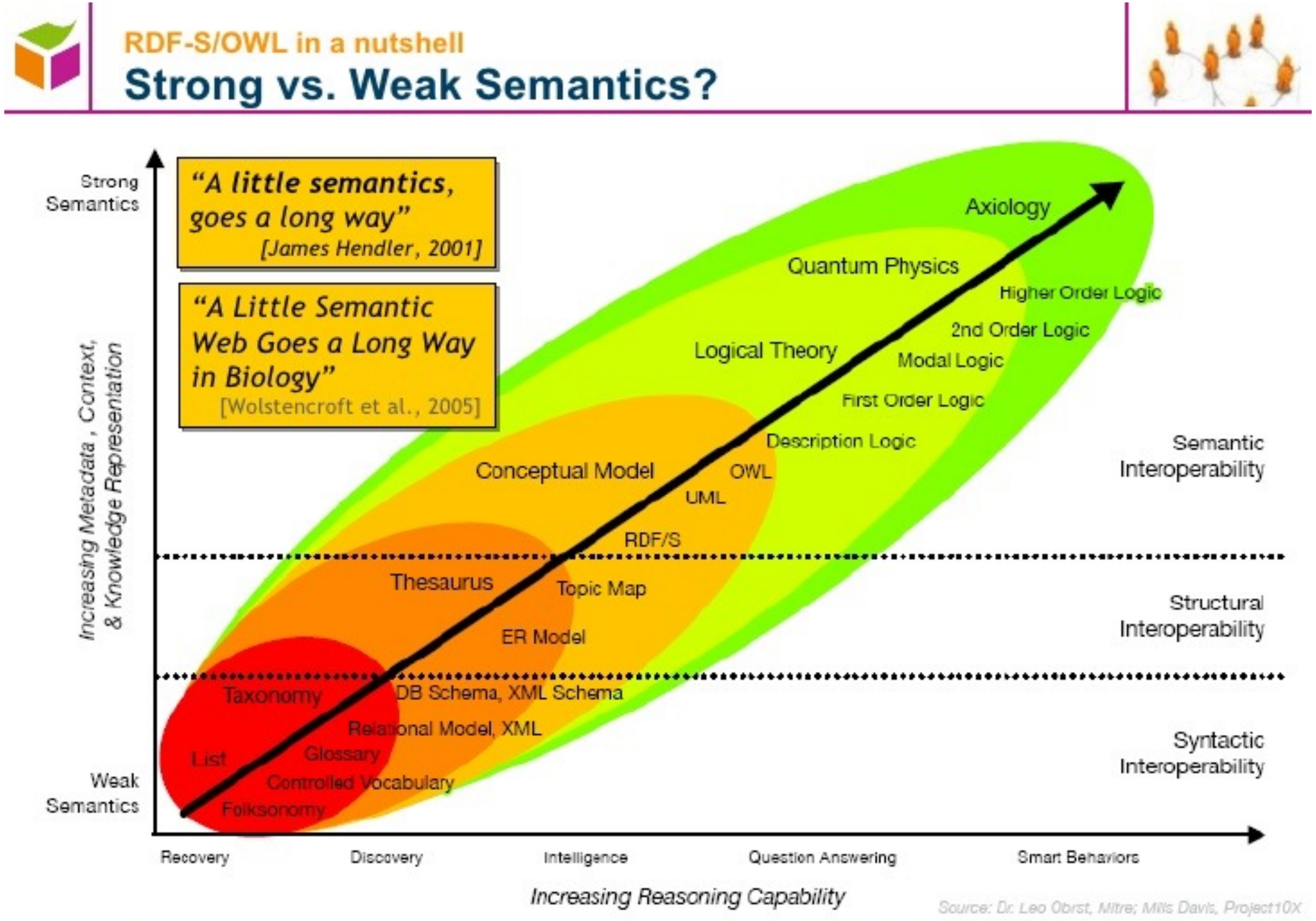
Triples form VS natural language

Generic vs Social

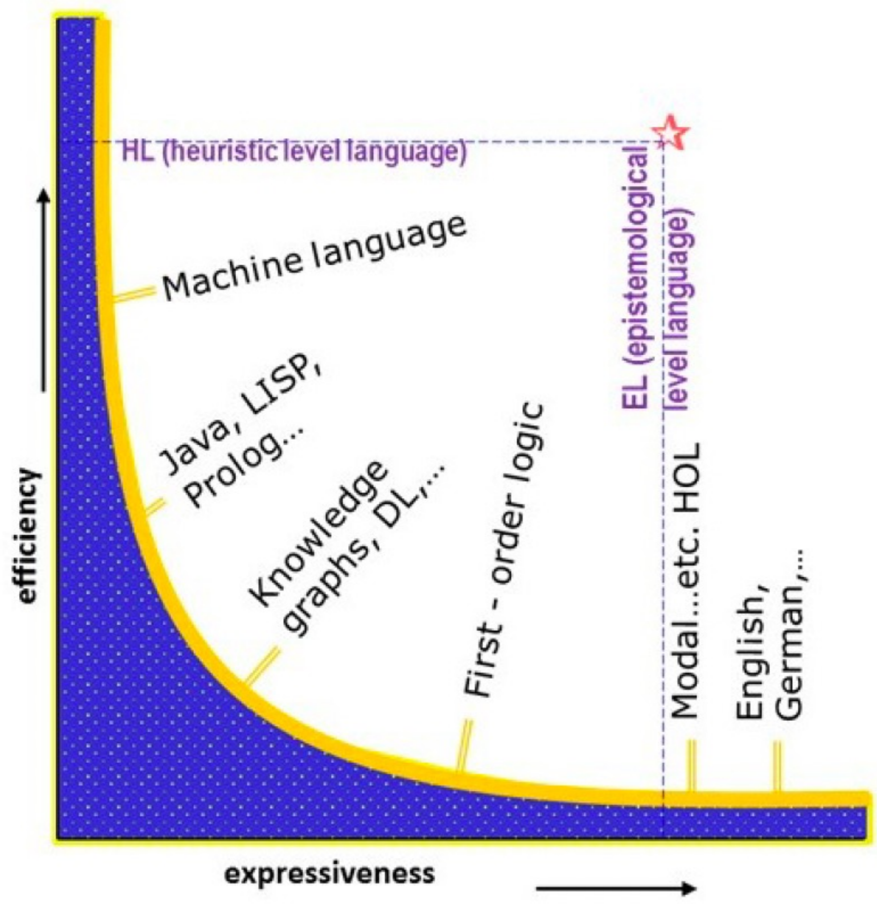
Speer, Robyn, Joshua Chin, and Catherine Havasi. "Conceptnet 5.5: An open multilingual graph of general knowledge." In Thirty-first AAAI conference on artificial intelligence. 2017.

Sap, Maarten, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A. Smith, and Yejin Choi. "Atomic: An atlas of machine commonsense for if-then reasoning." In *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, pp. 3027-3035. 2019.

Knowledge Bases and Knowledge Graphs
Expressivity and Reasoning



Knowledge Bases and Knowledge Graphs
Expressivity and Efficiency

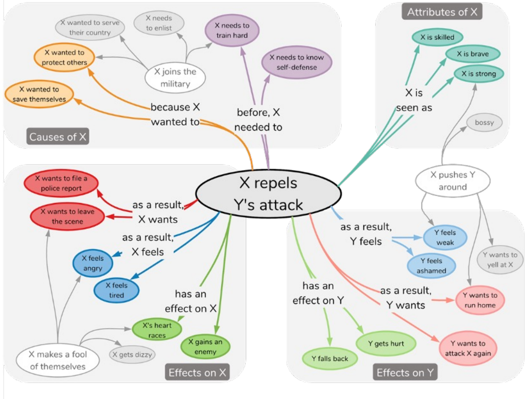


BOSCH

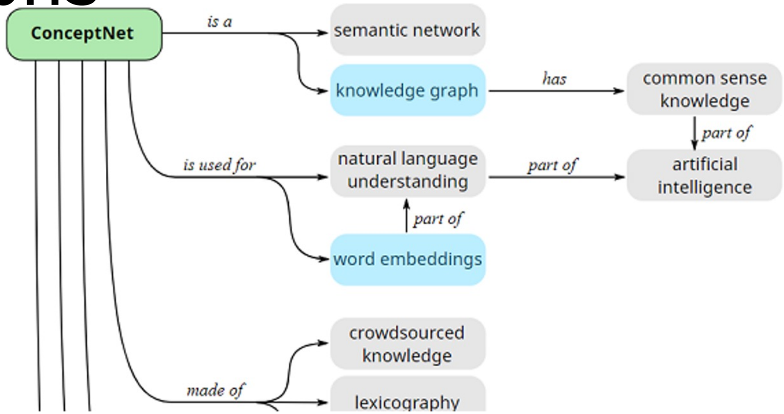
Knowledge Bases and Knowledge Graphs
Breadth & Depth

- Individual KGs only cover portions of Commonsense
  - ConceptNet, WordNet, ATOMIC, Visual Genome, ...
- Recent effort to align individual Commonsense KGs: USC's "CSKG"
- Still, CSKG enforces a "syntax-based" alignment
  - typically interpreted as task of expanding the conceptual coverage through mappings
    - REMINDER: Completeness is a challenge
  - Limited efforts focused on vertical expansion, i.e.:
    - Domain-specific commonsense knowledge dimensions
    - Specific forms of **reasoning**
- For real-world use cases, vertical KRR augmentation is key
  - Real-time monitoring (e.g., of traffic, vital signals, market trends) requires solid spatiotemporal reasoning

Knowledge Bases and Knowledge Graphs
Consolidated Knowledge Graphs

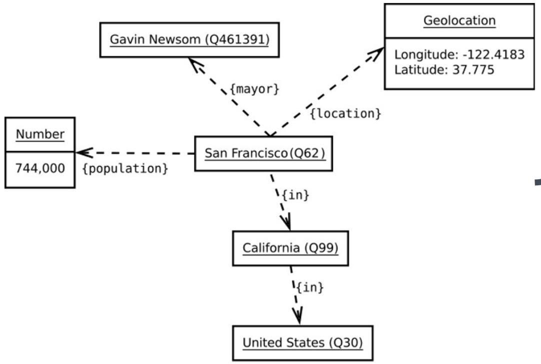


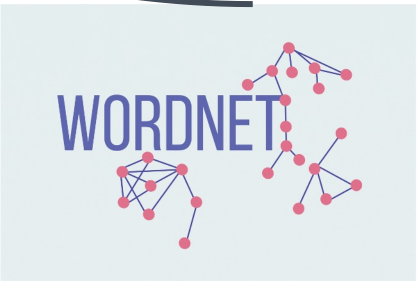
ATOMIC (Sap et al. 2019)



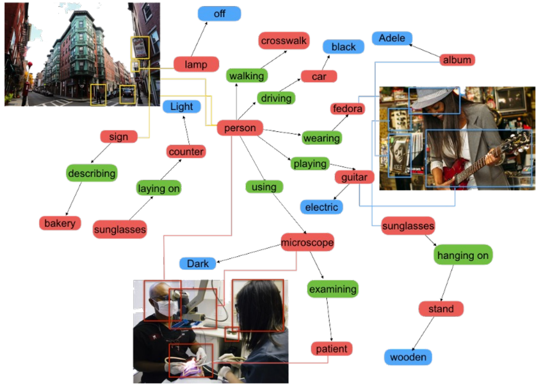
ConceptNet (Speer, Chin and Havasi 2017)

CSKG (Ilievski et al. 2020)





WordNet (Miller 1995)



Visual Genome (Krishna et al. 2017)

BOSCH

NEURO-SYMBOLIC COMMONSENSE QUESTION-ANSWERING

Deep Dive

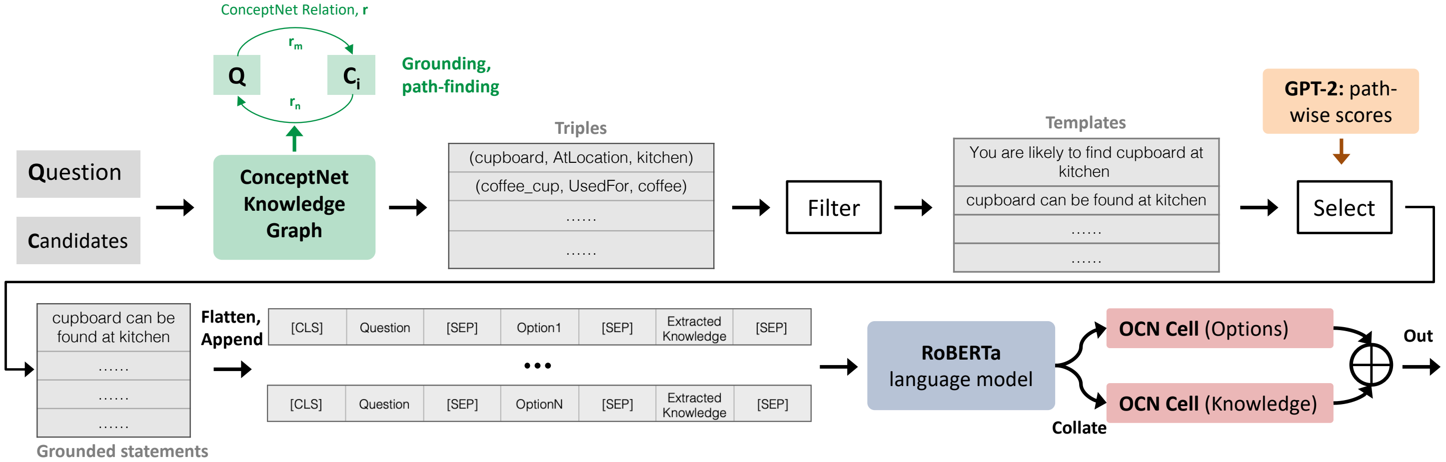
Generalizable Neuro-Symbolic Architectures for Commonsense QA

Question Answering.
*Given a natural language question **Q**, and **n** possible answers* $\{A_1,\ldots,A_n\}$*, find the most probable single answer* A_i*.*

- Identify the most appropriate commonsense knowledge type and knowledge resource (e.g., ConceptNet, ATOMIC), for each task
- Perform grounding, path-finding, and lexicalisation to obtain knowledge-driven representations
- Select the best knowledge+neural integration mechanism (e.g., attention, pre-training)
- Evaluate on two datasets (e.g., DREAM, CommonsenseQA); particularly interested in tasks that require common sense to solve
- *Learned:* knowledge-task alignment is crucial to downstream performance

An example from the DREAM dataset, which assesses models' abilities to perform commonsense reasoning. The asterisk (*) denotes the correct answer.

Dialogue:
M: I hear you drive a long way to work every day.
W: Oh, yes. it's about sixty miles. but it doesn't seem that far, the road is not bad, and there's not much traffic.
Question:
How does the woman feel about driving to work?
Answer choices:
A. She doesn't mind it as the road conditions are good.*
B. She is unhappy to drive such a long way everyday.
C. She is tired of driving in heavy traffic.



Alignment between knowledge, task structure, and data is crucial for downstream performance

Deep Dive
Generalizable Neuro-Symbolic Architectures for Commonsense QA

Experimental Design

- Knowledge resources: ConceptNet **versus** ATOMIC
- Methodology: knowledge distillation (pre-training) **versus** attention-based knowledge injection
    - ConceptNet:
        - Pretrained BERT on Open Mind Common Sense (OMCS) with masked language modelling (MLM) objective
        - e.g., [CLS] pen is [MASK] the [MASK] size as pencil [SEP]
    - ATOMIC:
        - Created special tokens for relations: xIntent, xReact…. And only mask out tail tokens
        - e.g., [CLS] person ##x waits a <blank> to get <xReact> [MASK] [MASK] [SEP]
- Datasets: DREAM, CommonsenseQA

An example from the DREAM dataset, which assesses models' abilities to perform commonsense reasoning. The asterisk (*) denotes the correct answer.

Dialogue:
M: I hear you drive a long way to work every day.
W: Oh, yes. it's about sixty miles. but it doesn't seem that far, the road is not bad, and there's not much traffic.
Question:
How does the woman feel about driving to work?
Answer choices:
A. She doesn't mind it as the road conditions are good.*
B. She is unhappy to drive such a long way everyday.
C. She is tired of driving in heavy traffic.

An example from the CommonsenseQA dataset, which assesses models' abilities to perform commonsense reasoning. The asterisk (*) denotes the correct answer.

Question:
A revolving door is convenient for two direction travel, but it also serves as a security measure at a what?
Answer choices:
A. Bank*; B. Library; C. Department Store;
D. Mall; E. New York

Deep Dive

Generalizable Neuro-Symbolic Architectures for Commonsense QA

Model performances on **DREAM** dataset (*) denotes results from leaderboard. “AT” is ATOMIC, “CN” ConceptNet, and ”OMCS” refers to the ‘Open Mind Common Sense” corpus, on which CN is based.

Models	Dev Acc	Test Acc
BERT Large(*)	66.0	66.8
XLNet(*)	-	**72.0**
OCN	70.0	69.8
OCN + CN injection	**70.5**	69.6
OCN + AT injection	69.6	**70.1**
OCN + OMCS pre-train	64.0	62.6
OCN + ATOMIC pre-train	60.3	58.8

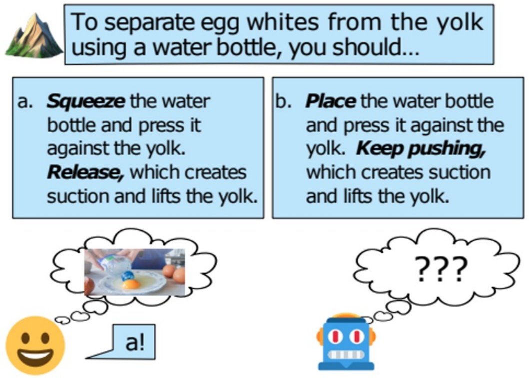
Model performance on **CommonsenseQA** dataset. Asterisk (*) denotes results taken from task leaderboard. “CSPT” refers also to commonsense pre-training on OMCS.

Models	Dev Acc
BERT + OMCS pre-train(*)	68.8
RoBERTa + CSPT(*)	**76.2**
OCN	64.1
OCN + CN injection	67.3
OCN + OMCS pre-train	65.2
OCN + ATOMIC pre-train	61.2
OCN + OMCS pre-train + CN inject	**69.0**

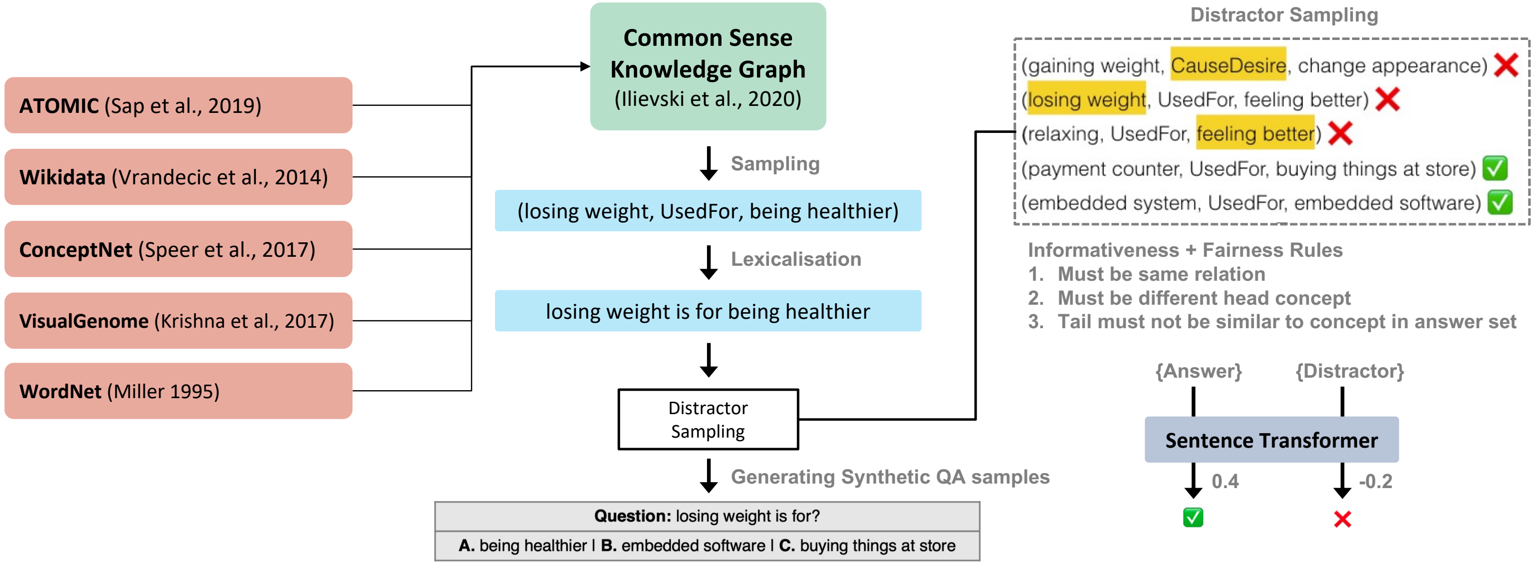
Deep Dive
Knowledge-driven Data Construction for Zero-shot Evaluation

Our framework takes existing **knowledge graphs** and **transforms** them into datasets of Question-Answer-Distractor samples, in order to enable optimisation of **language models** through **pre-training** and to solve a diverse set of commonsense QA **tasks**, under zero-shot evaluation.

- Knowledge graphs – ATOMIC, ConceptNet, WordNet, Wikidata, Visual Genome
- Transformation – Random sampling, adversarial sampling, adversarial filtering
- Language model (classes) – RoBERTa, GPT-2
- Pretraining – Marginal ranking (MR), language modeling (MLM)
- Tasks – PIQA, aNLI, CSQA, SocialIQA, WG



Ex: **PIQA task** (Bisk et al. 2020)



Models with commonsense pre-training have higher zero-shot evaluation performance

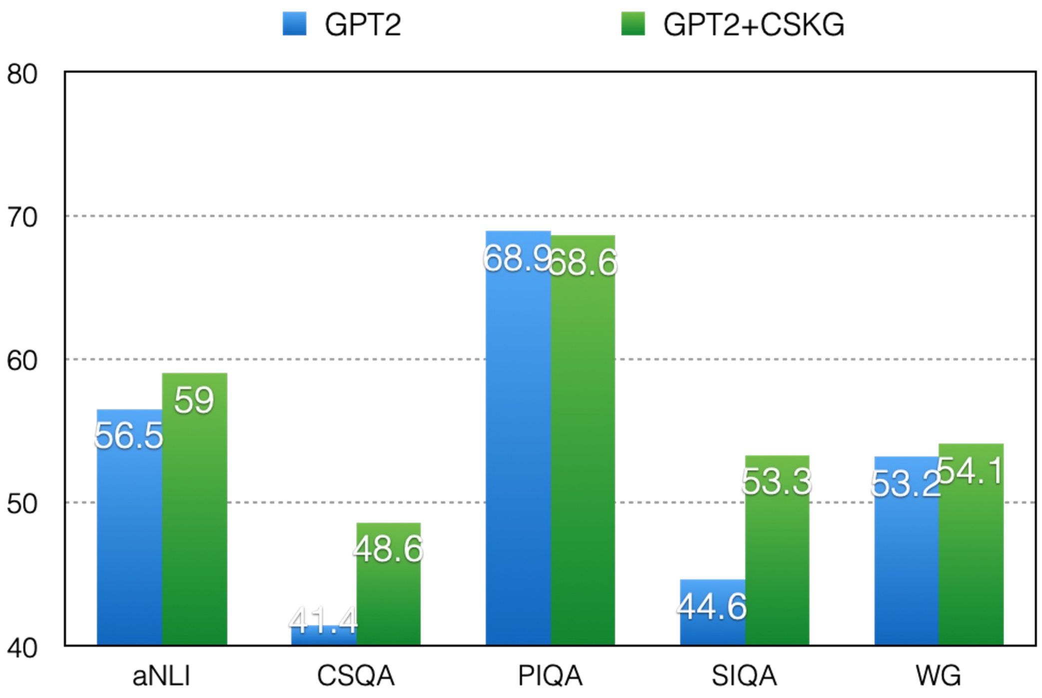
Deep Dive

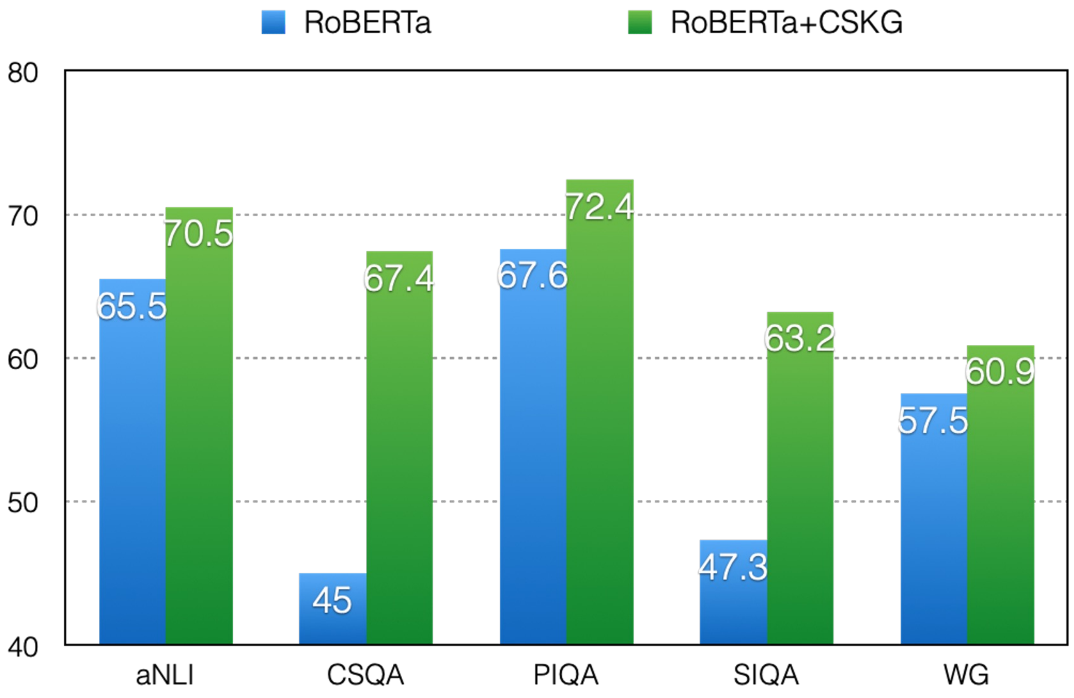
Knowledge-driven Data Construction for Zero-shot Evaluation

Comparison of model classes

Takeaways:

- Pretraining on artificial QA sets helps accuracy.
- RoBERTa (MLM class) outperforms GPT2 (autoregressive class).





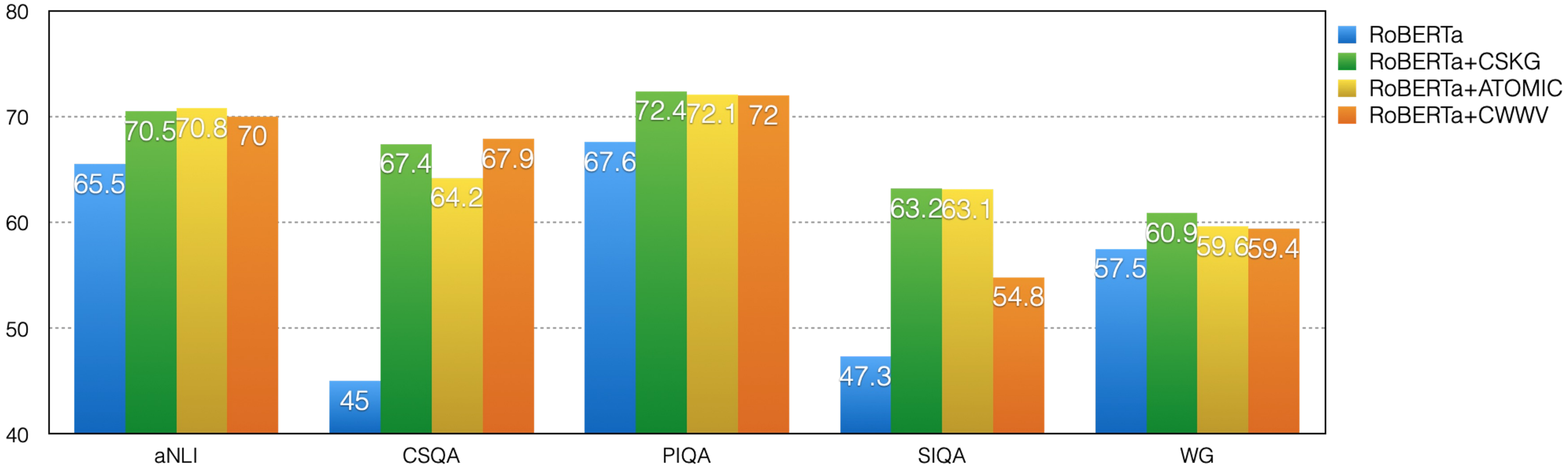
Deep Dive

Knowledge-driven Data Construction for Zero-shot Evaluation

Comparison of commonsense knowledge types.

Takeaways:

- The impact of knowledge depends on KG-task alignment.
- Adding the right knowledge improves accuracy.



Deep Dive
Neuro-Symbolic Commonsense Question Answering

Summary

- We believe that it is important to identify the most appropriate knowledge type required for specific tasks. Once the knowledge type is identified, we can then select the appropriate knowledge-base(s) and the suitable neural integration mechanisms.

- Naïve attempts to fuse commonsense knowledge with task context actually lead to reductions in performance

- Understanding how to represent and incorporate symbolic commonsense knowledge in learning-based systems is crucial for their improved optimisation and generalisability.

Approaches discussed

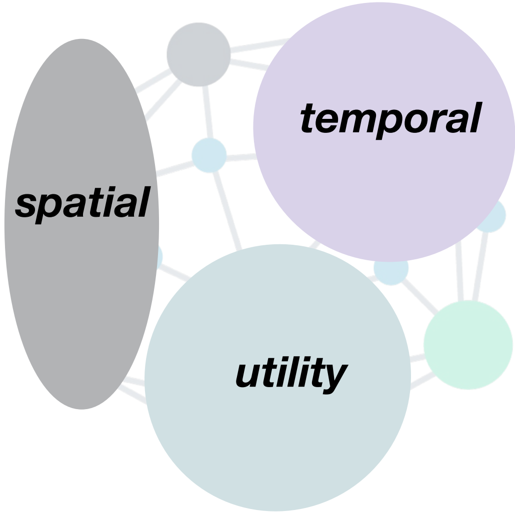
- Generalizable Neuro-Symbolic Architectures for Commonsense QA *(commonsense knowledge extraction + injection)*
- Knowledge-driven Data Construction for Zero-shot Evaluation *(pre-training, logical rules, task structures)*

COMMONSENSE DIMENSIONS FOR "VERTICAL" REASONING

BOSCH

What is the role of different types of (commonsense) knowledge?

+



Ilievski, Filip, Alessandro Oltramari, Kaixin Ma, Bin Zhang, Deborah L. McGuinness, and Pedro Szekely. "Dimensions of commonsense knowledge." *Knowledge-Based Systems* 229 (2021): 107347.

Commonsense knowledge about food in 21 sources

Source	Subject	Relation	Object
ConceptNet*	food	capable of	go rotten
ATOMIC	person X bakes bread	xEffect	eat food
GLUCOSE	someone makes something that is food	Causes/Enables	someone eats something
WebChild	restaurant food	quality	expensive
Quasimodo	pressure cooker	cook faster	food
SenticNet	coldfood	polarity	negative
HasPartKB	dairy food	has part	vitamin
Probase	apple	is a	food
Isacore	snack food	is a	food
Wikidata	food	has quality	mouthfeel
YAGO4	banana chip	type	food
SUMO*	food	hyponym	food product
WordNet	food	hyponym	comfort food
Roget	dish	synonym	food
FrameNet	cooking	has frame element	creation
MetaNet	Food	has role	food
VerbNet	feed.v.01	Arg1	PPT
Visual Genome	food	on	plate
Flickr30k	a food buffet	corefers with	a food counter
LMs	Aardvarks search for food		
GPT	Food causes a person to be hungry and a person to eat		

BOSCH

Category	Source	# Relations	Subject	Relation	Object
Commonsense KG	ConceptNet*	34	food	capable of	go rotten
	ATOMIC	9	person X bakes bread	xEffect	eat food
	GLUCOSE	10	someone makes something that is food	Causes/Enables	someone eats something
	WebChild	4 groups	restaurant food	quality	expensive
	Quasimodo	**78,000**	pressure cooker	cook faster	food
	SenticNet	1	coldfood	polarity	negative
	HasPartKB	1	dairy food	has part	vitamin
	Probase	1	apple	is a	food
	Isacore	1	snack food	is a	food
General KG	Wikidata	**6,700**	food	has quality	mouthfeel
	YAGO4	116	banana chip	type	food
	SUMO*	**1,600**	food	hyponym	food product
Lexical resource	WordNet	10	food	hyponym	comfort food
	Roget	2	dish	synonym	food
	FrameNet	8	cooking	has frame element	creation
	MetaNet	14	Food	has role	food
	VerbNet	36	feed.v.01	Arg1	PPT
Visual source	Visual Genome	**42,000**	food	on	plate
	Flickr30k	1	a food buffet	corefers with	a food counter
Corpora and LMs	LMs		Aardvarks search for food		
	GPT		Food causes a person to be hungry and a person to eat		

BOSCH

Research questions

- Can we consolidate the relations of commonsense knowledge?
- Is the consolidated knowledge useful in downstream tasks such as Q/A, text generation, vision-language understanding, etc.
  - *Can it boost downstream tasks in specific domains (e.g. traffic monitoring)*

13 dimensions of commonsense relations	overlap, coverage and gaps of commonsense sources	impact of dimensions on downstream tasks

Dimensions of Commonsense

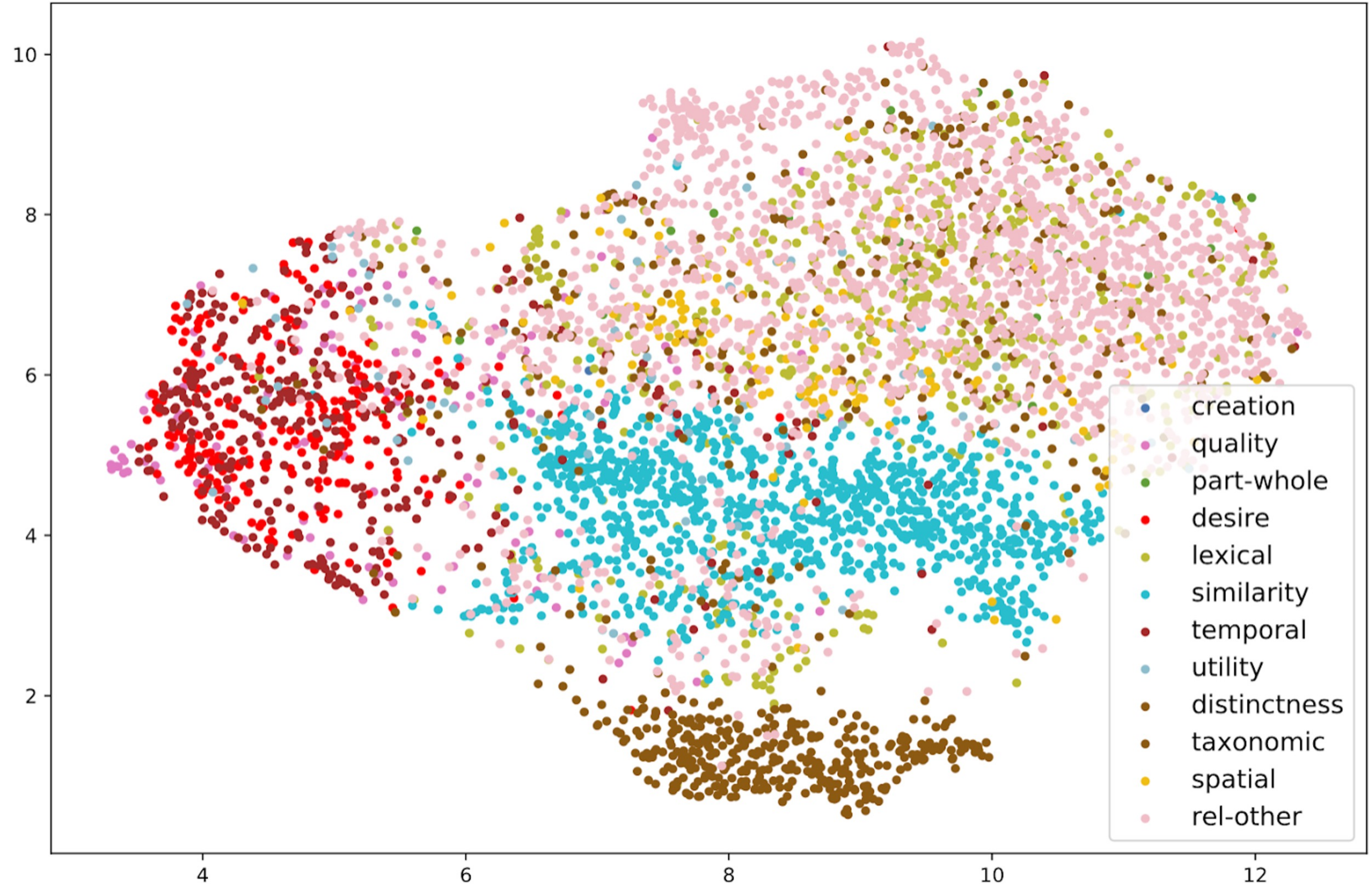
Dimension	Example (food)	Source
lexical	**derivationaly related form:** nutrient	WordNet
similarity	**synonym:** dish	ROGET
distinctiveness	**opposite of:** non-food item	Wikidata
taxonomic	**hypernym:** substance	WordNet
part-whole	**things with food:** minibar	ConceptNet
spatial	**located near:** plate	Visual Genome
creator	**is created by:** cook	COMET
utility	**used for:** pleasure	ConceptNet
motivational	**xWant:** watch movie together - get some food	ATOMIC
quality	**has the property:** tasty	COMET
comparative	**healthier:** home cooking - fast food	WebChild
temporal	**has effect:** food allergy	Wikidata
relational-other	**related to:** refrigerator	ConceptNet

Dimensions of Commonsense

Dimension	ATOMIC	ConceptNet	WebChild	Roget	WikidataCS	WordNet	FrameNet
lexical		704			0.5	207	14
similarity		255	343	1,023	1	152	0.4
distinctiveness		22		381	7	4	
taxonomic		244	783		73	89	23
part-whole		19	5,752		8	22	
spatial		28	660		0.5		
creation		0.3			0.2		
utility		69	2,843		2		1
motivational	244	20					
quality	143	9	6,510		1		11
comparative			813				
temporal	346	71	2,135		3		0.6
relational-other		1,969	291		6		0.7

Dimensions of Commonsense

RoBERTa embeddings of lexicalized triples: little agreement with dimensions



BOSCH

Dimension-aware LM adaptation

dimension

CSKG

+

Question: losing weight is for?
A: being healthier **B**. embedded software **C.** buying things at store

=

losing weight is used for...

- en adjusting body proportions →
- en becoming phyiscally fit →
- en being healthier →
- en better agility →
- en better health →
- en buying more clothes that fit →
- en competing in the olympics →
- en the decadent →
- en decreasing waist size →
- en feeling better →
- en feeling better about yourself →
- en finding clothes that fit →
- en fitting into clothes →
- en fitting into your old clothes →
- en gaining better health →
- en getting healthier →
- en getting in shape →
- en getting rid of excess fat →
- en health →
- en heavyweights →

More »

SocialIQA
test data

CSQA
test data

PhysicalIQA
test data

Winogrande
test data

aNLI
test data

Dimensions of Commonsense

Training data

commonsense triples to questions

pretrain language model, by dimension

evaluate on CommonsenseQA, SocialIQA

Dimension	Train	Dev
part-whole	87,765	4,620
taxonomic	340,609	17,927
lexical	107,861	5,677
distinctness	20,286	1,068
similarity	166,575	8,768
quality	116,593	12,492
utility	63,862	3,362
creation	304	17
temporal	312,628	31,587
relational-other	242,759	12,777
spatial	21,726	1,144
desire/goal	194,906	20,912

BOSCH

CSQA = Commonsense QA
SIQA = SocialIQA

Pre-training language models with dimensions

Dimensions	CSQA	SIQA
Baseline	45.0	47.3
+part-whole	63.0(±1.4)	52.6(±1.9)
+taxonomic	62.6(±1.4)	52.2(±1.6)
+lexical	49.9(±2.9)	49.0(±0.4)
+distinctness	57.2(±0.5)	50.2(±1.5)
+similarity	61.4(±0.8)	53.5(±0.6)
+quality	65.7(±0.5)	60.0(±0.7)
+utility	**67.4**(±**1.0**)	54.8(±0.7)
+creation	49.9(±1.1)	47.8(±0.2)
+temporal	67.3(±0.3)	**62.6**(±**0.9**)
+relational-other	58.2(±1.7)	51.3(±1.7)
+spatial	63.3(±0.2)	53.1(±0.3)
+desire/goal	65.0(±1.8)	60.0(±0.6)
+all	66.2(±1.4)	61.0(±0.7)

Novelty per dimension

Can ‘vanilla’ RoBERTa answer the questions without pretraining?

Dimensions	Dev
part-whole	67.5
taxonomic	57.0
lexical	90.1
distinctness	77.3
similarity	65.6
quality	45.5
utility	67.9
creation	82.4
temporal	47.2
relational-other	37.6
spatial	56.9
desire/goal	48.0

Neuro-Symbolic Commonsense Question Answering

Our most recent work

- Discriminative tasks
  - Knowledge is only helpful when there is alignment of data (questions) and knowledge-base types
  - Attention based injection is preferable to pretraining on knowledge bases
  - A model trained on a rich set of questions synthetized from CSKG, and covering a wide spectrum of knowledge types, keeps performance stable across tasks
- Generative tasks
  - Attention-based injection of relevant triples improve quality of generated answers
  - Task with similar data structure transfer better
- Dimensions
  - The temporal and intentional dimensions are very beneficial for reasoning on current downstream tasks, while distinctness and lexical knowledge have little impact.

EMNLP-COIN 2019: Ma, Kaixin, Jonathan Francis, Quanyang Lu, Eric Nyberg, and Alessandro Oltramari. "Towards generalizable neuro-symbolic systems for commonsense question answering." *arXiv preprint arXiv:1910.14087* (2019).

AAAI 2021: Ma, Kaixin, Filip Ilievski, Jonathan Francis, Yonatan Bisk, Eric Nyberg, and Alessandro Oltramari. "Knowledge-driven Data Construction for Zero-shot Evaluation in Commonsense Question Answering." arXiv preprint arXiv:2011.03863 (2020).

AAAI-CSKG 2021: Li, Yikang, Pulkit Goel, Varsha Kuppur Rajendra, Har Simrat Singh, Jonathan Francis, Kaixin Ma, Eric Nyberg, and Alessandro Oltramari. "Lexically-constrained Text Generation through Commonsense Knowledge Extraction and Injection." *arXiv preprint arXiv:2012.10813* (2020).

ACL-IJCNLP 2021: Chen, Xi, Lin, Faner, Zhou, Yeju, Ma, Kaxin, Jonathan, Francis, Eric Nyberg, Alessandro Oltramari. "Building Goal-oriented Document-grounded Dialogue Systems" (*submitted*)

KBS Journal 2021: Ilievski, Filip, Alessandro Oltramari, Kaixin Ma, Bin Zhang, Deborah L. McGuinness, and Pedro Szekely. "Dimensions of commonsense knowledge." *arXiv preprint arXiv:2101.04640* (2021). (*submitted*)

Neuro-Symbolic Commonsense Question Answering
A retrospective of our most recent Work

294

Neuro-Symbolic Artificial Intelligence: The State of the Art
P. Hitzler and M.K. Sarker (Eds.)
IOS Press, 2022
© 2022 The authors and IOS Press. All rights reserved.
doi:10.3233/FAIA210360

Chapter 13

Generalizable Neuro-Symbolic Systems for Commonsense Question Answering

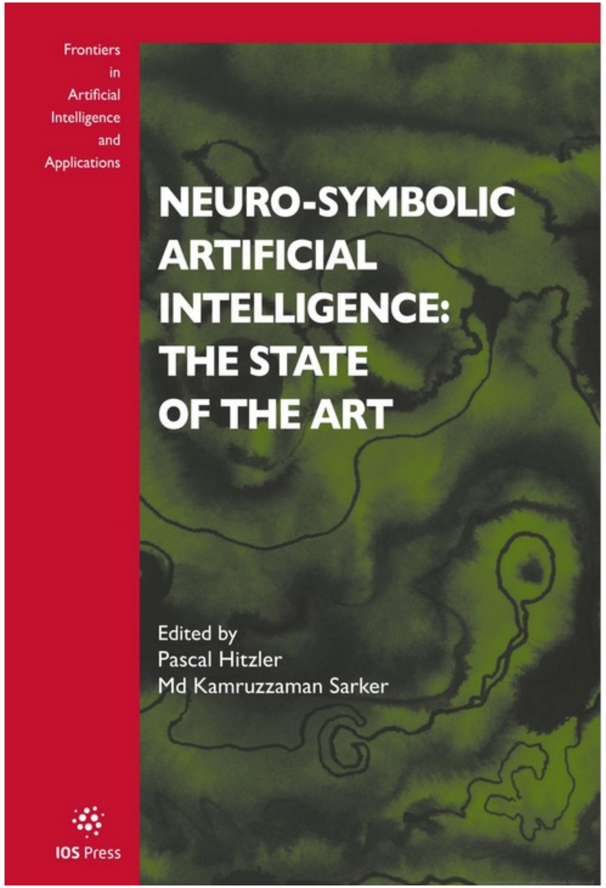
Alessandro Oltramari, Bosch Research (Pittsburgh, PA USA)

Jonathan Francis, Bosch Research (Pittsburgh, PA USA); School of Computer Science, Carnegie Mellon University (Pittsburgh, PA USA)

Filip Ilievski, Information Sciences Institute, Viterbi School of Engineering, University of Southern California (Marina del Rey, CA USA)

Kaixin Ma, School of Computer Science, Carnegie Mellon University (Pittsburgh, PA USA)

Roshanak Mirzaee, Department of Computer Science and Engineering, Michigan State University (East Lansing, MI USA)



BOSCH

GLANCING OVER OTHER USE CASES

Traffic Monitoring
The "Present"

Bosch camera

Intelligent Video Analytics

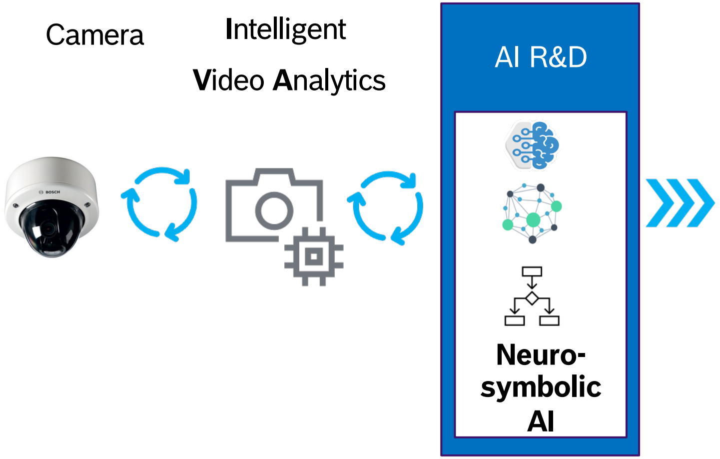
LOW LEVEL FEATURES

- **Object Recognition**
  - Car
  - Truck
  - Pedestrian
  - "Other"
- **Object Attributes**
  - Units
  - Speed
  - Trajectory
- **"Volume" and position**
  - Normal traffic in eastbound lane
- **Basic behavior**
  - Car above speed limit
  - Collision

BOSCH

Traffic Monitoring
The “Future”?

Traffic Monitoring
The "Future"!

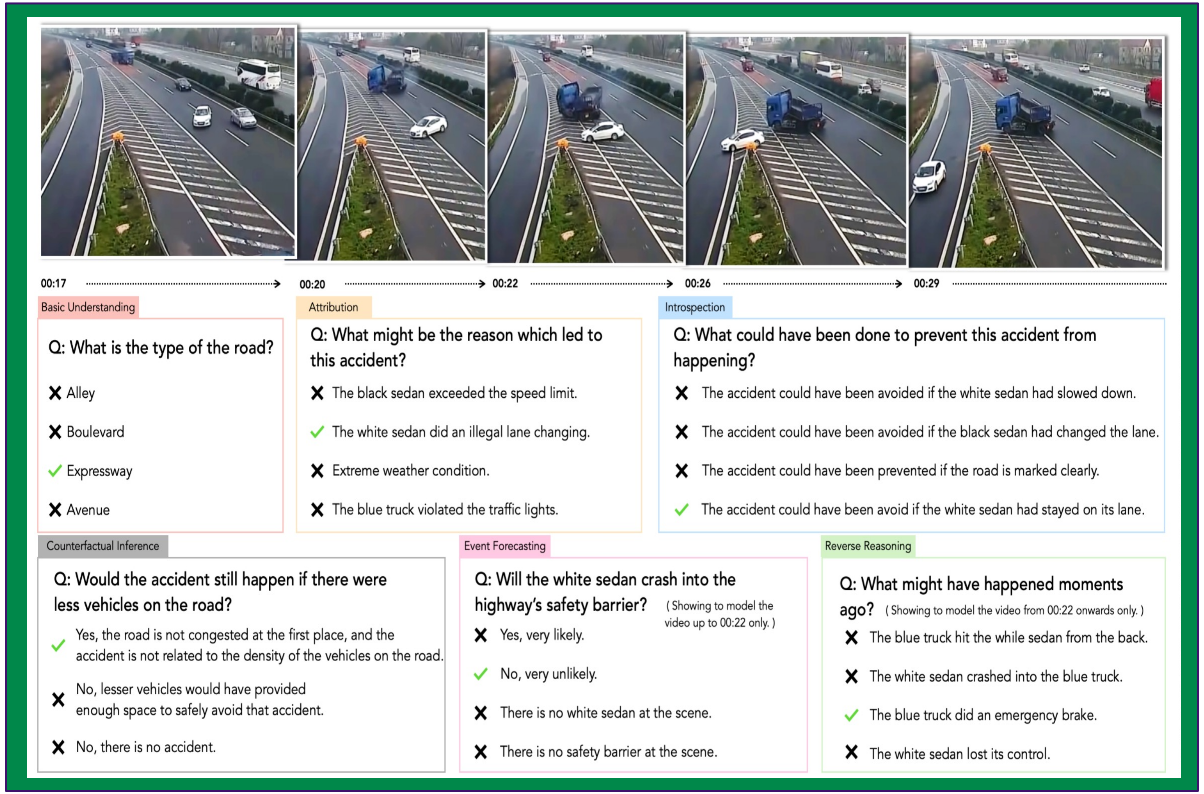


HIGH LEVEL CONTEXT

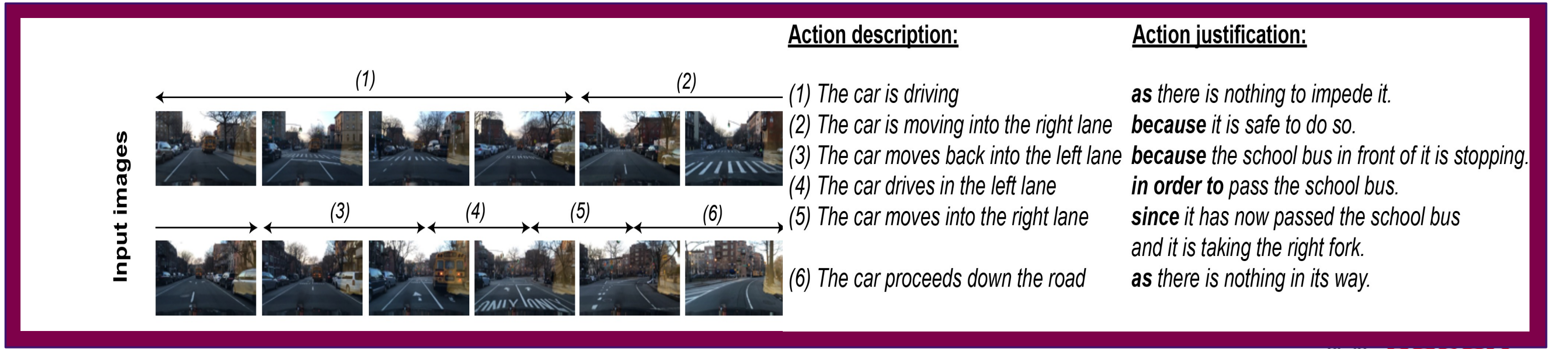
- **Vehicle behavior**
  - Garbage truck
  - Light-duty plow truck
  - Parking violation
  - Near miss
- **Human behavior**
  - Jaywalking
  - Illegal dumping
  - Criminal activity
    - auto theft
    - reckless driving
- **Forensic Search**
  - Re-evaluate metadata with new criteria (e.g., for alarms)

Traffic Monitoring
Multimodal QA

- Ongoing investigation of open multimodal datasets
  - **SUTD-TrafficQA** containing manually annotated 62,535 QA pairs
  - **BDD-X dataset** containing 26,228 annotations for 6,984 videos.



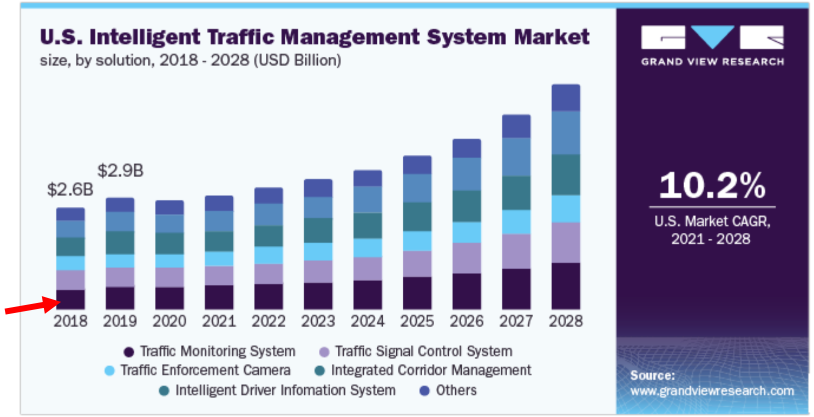
Xu, Li, He Huang, and Jun Liu. "**Sutd-trafficqa: A question answering benchmark and an efficient network for video reasoning over traffic events.**" In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9878-9888. 2021.



Kim, Jinkyu, Anna Rohrbach, Trevor Darrell, John Canny, and Zeynep Akata. "**Textual explanations for self-driving vehicles.**" In *Proceedings of the European conference on computer vision (ECCV)*, pp. 563-578. 2018.

Traffic Monitoring
Market figures

▶ Intelligent Traffic Management & Monitoring



Source: Grandview Research, September 2021

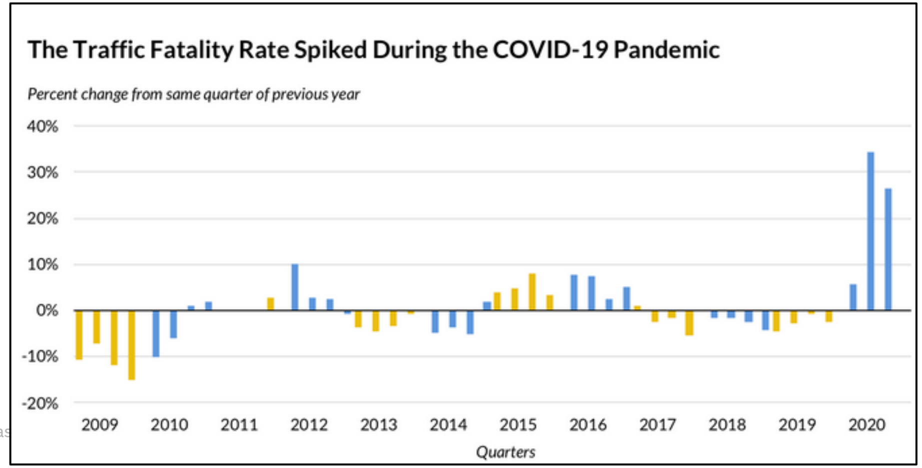
- “Safe streets for All” – U.S. Infrastructure bill
  - *$11 billion for transportation safety, including a program to help states and localities reduce crashes and fatalities, especially of cyclists and pedestrians, according to the White House. It would direct funding for safety efforts involving highways, trucks, and pipeline and hazardous materials.*

Source: CNN, Sunday Nov 7 2021

▶ Road Safety and Security

	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
Road safety & security	895	995	1104	1227	1379	1550	1742	1960	2205	2483
YoY Growth Rate (%)	*10.7*	*11.2*	*11.0*	*11.1*	*12.4*	*12.4*	*12.4*	*12.5*	*12.5*	*12.6*

Source: BT-VS/XSW-ITS



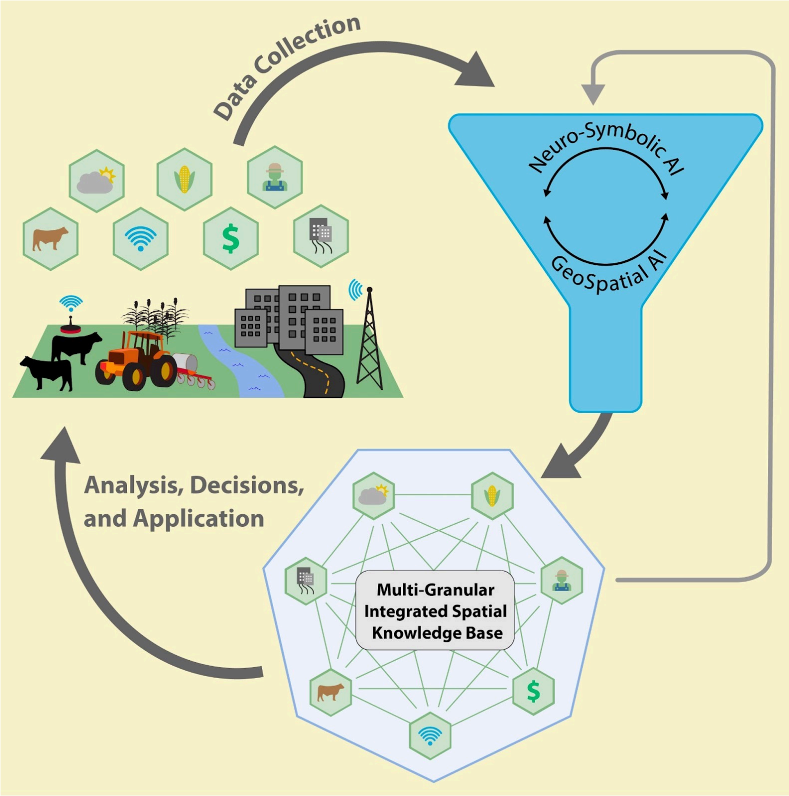
Source:
National Highway Traffic Safety Administration (2020) estimates.

Note:
Alternating colors show quarters for distinct years. Traffic fatality rate is the number of deaths from motor vehicle crashes per miles traveled

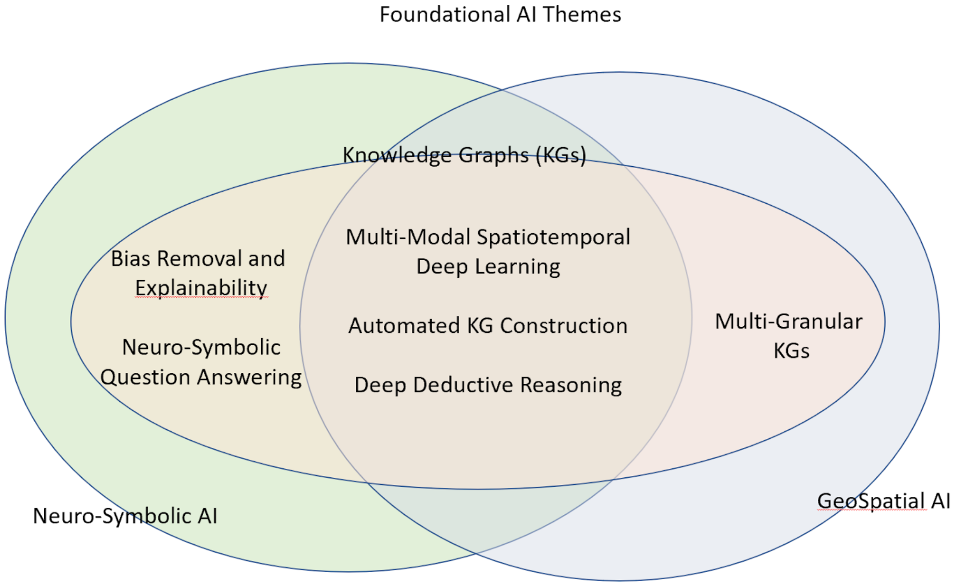
BOSCH

Glancing Over Other Use Cases

Knowledge-Centric AI for Climate Adaptive Agriculture in the Great Plains



Key components



Key topics

BOSCH

Glancing over other use cases
Intelligence, Surveillance, Reconnaissance

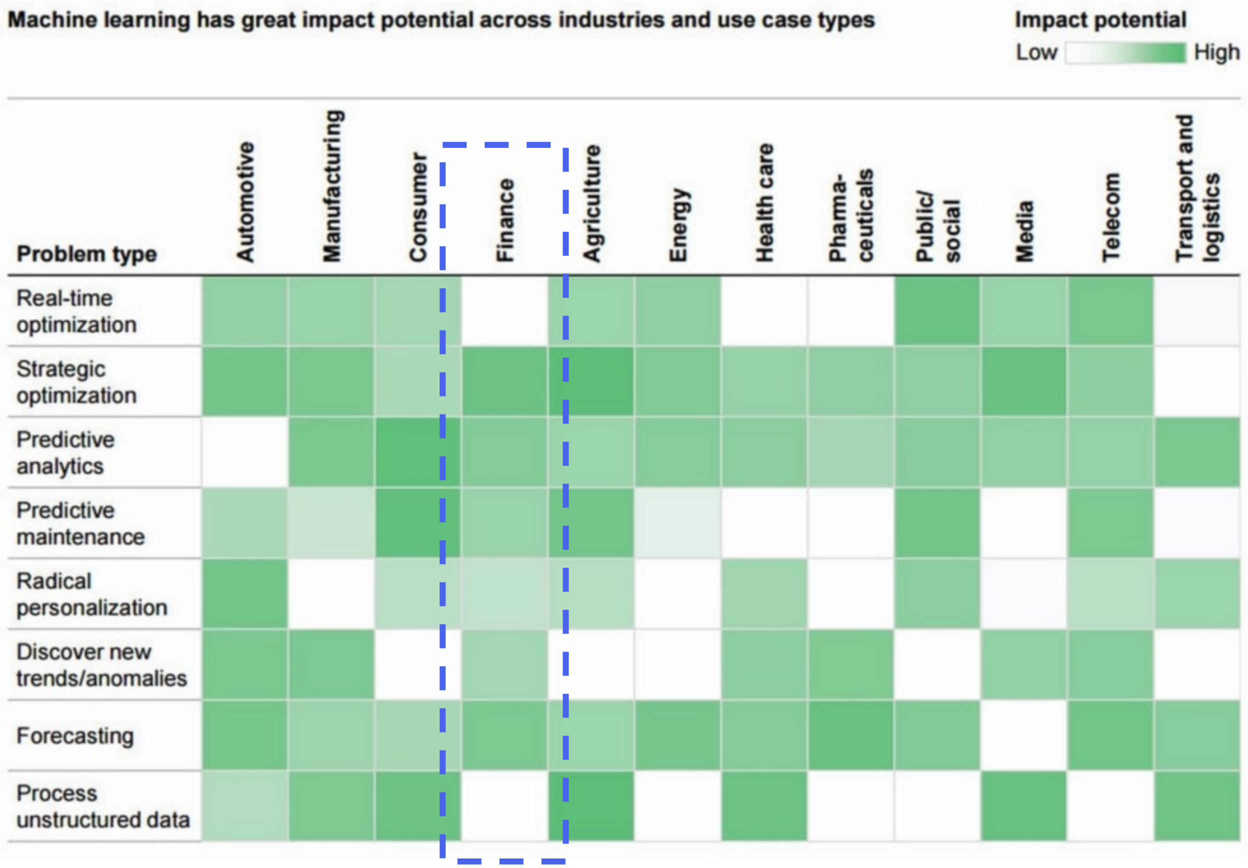
DARPA's ANSR to Improving Trustworthy AI

DARPA seeks proposals for Assured Neuro Symbolic Learning and Reasoning (ANSR) program

OUTREACH@DARPA.MIL
6/3/2022

Glancing over other use cases

NSAI as a booster for ML methods applied to Finance



Source:

McKinsey's 2016 Analytics Study on The Future Of Machine Learning

Internal | CR/AEU2 | 2019-06-03

Neuro-Symbolic AI
Conclusions

“We are building systems that govern healthcare, finance, and mediate our civic dialogue.

We influence elections.

I would like to live in a society whose systems are built on top of verifiable, rigorous, thorough knowledge, and not on ***alchemy****”*

— Ali Rahimi, recipient of Test-of-time award @NIPS 2017

BOSCH

THANKS!

QUESTIONS?