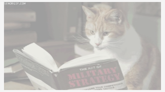
Agenda

Part-I

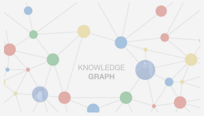


Extract triples and organize (from multimodal input)

Train to fill missing word etc.

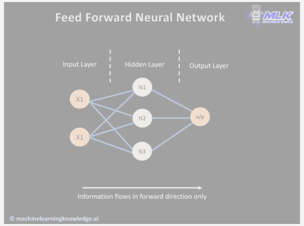
Part-II

symbolic representation



Neural helps symbolic

symbolic helps neural



neural representation of encoded knowledge in the hidden layers

Part-III

Evaluate KG

symbolic query, matching reqd

knowledge triple(s) retrieved

Task

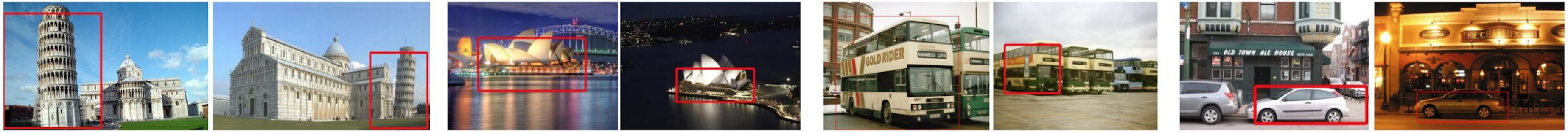
decode query relevant knowledge

embed symbolic query, auto matching

Multimodal KGs: NEIL KB

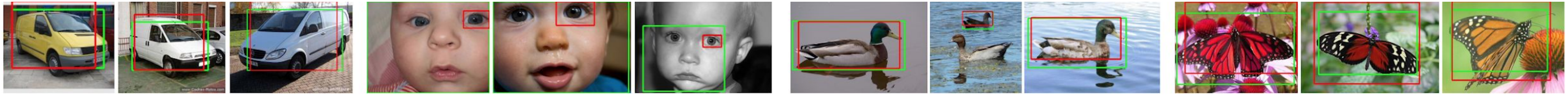
Scene-object relationships mined

Helicopter is found in **Airfield** | **Zebra** is found in **Savanna** | **Ferris wheel** is found in **Amusement park** | **Throne** is found in **Throne room**



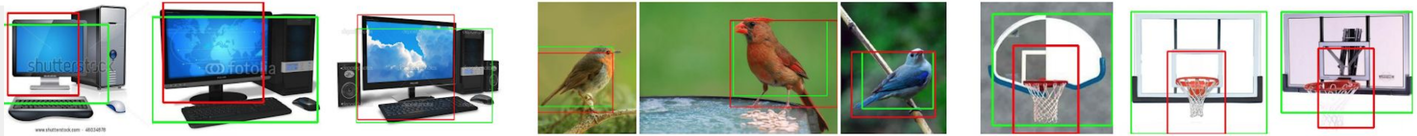
Leaning tower is found in **Pisa** | **Opera house** is found in **Sydney** | **Bus** is found in **Bus depot outdoor** | **Camry** is found in **Pub outdoor**

Object-object relationships mined



Van is a kind of/looks similar to **Ambulance** | **Eye** is a part of **Baby** | **Duck** is a kind of/looks similar to **Goose** | **Gypsy moth** is a kind of/looks similar to **Butterfly**

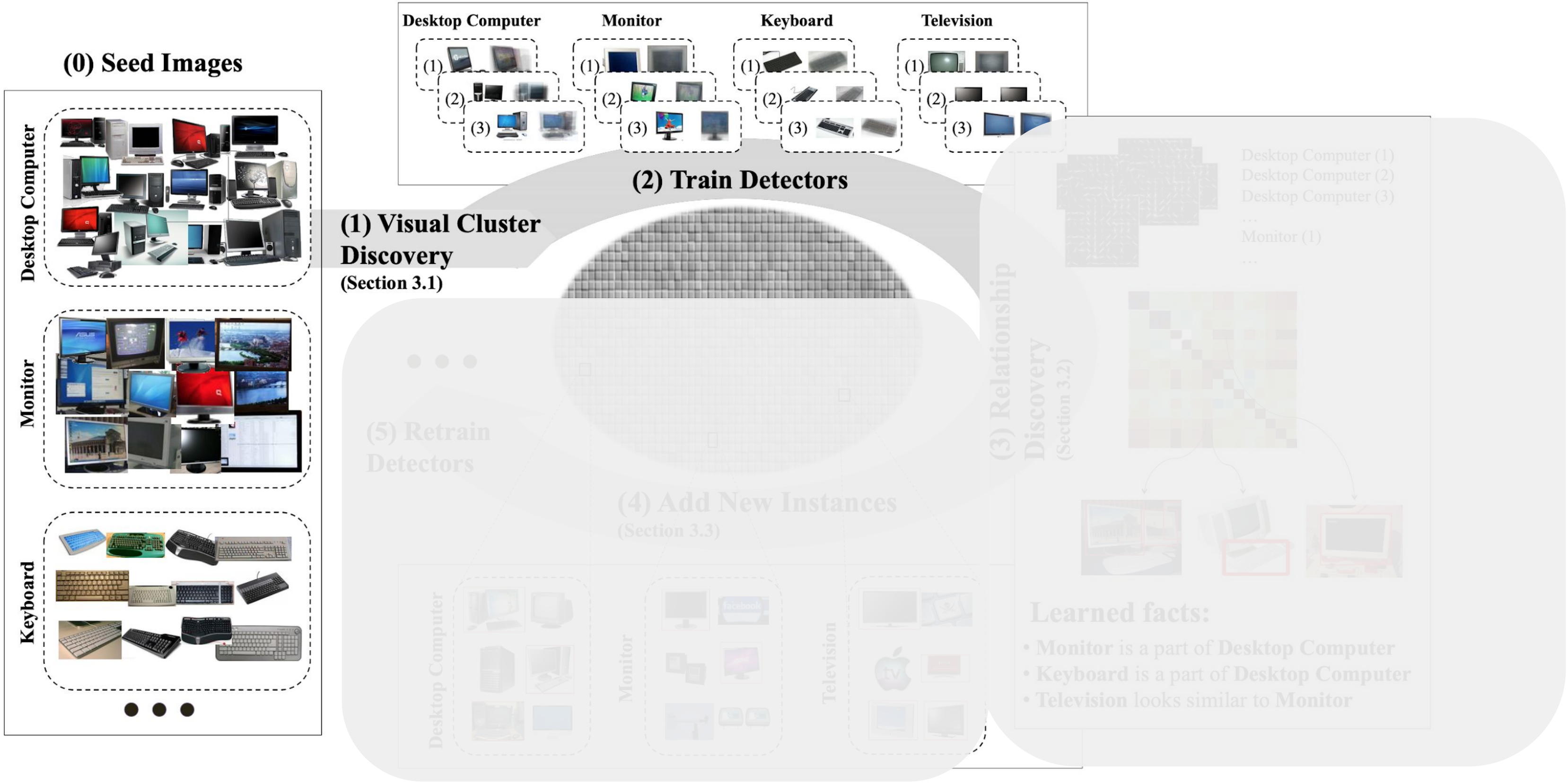
visual knowledge complements typical textual KG e.g. "monitor is expensive"



Monitor is a kind of/looks similar to **Desktop computer** | **Sparrow** is a kind of/looks similar to **bird** | **Basketball net** is a part of **Backboard**

NEIL: Extracting Visual Knowledge from Web Data, Chen et. al, ICCV 2013

NEIL KB: Approach



NEIL: Extracting Visual Knowledge from Web Data, Chen et. al, ICCV 2013

Visual Genome

visual attributes complement
typical textual KG attributes

Regions

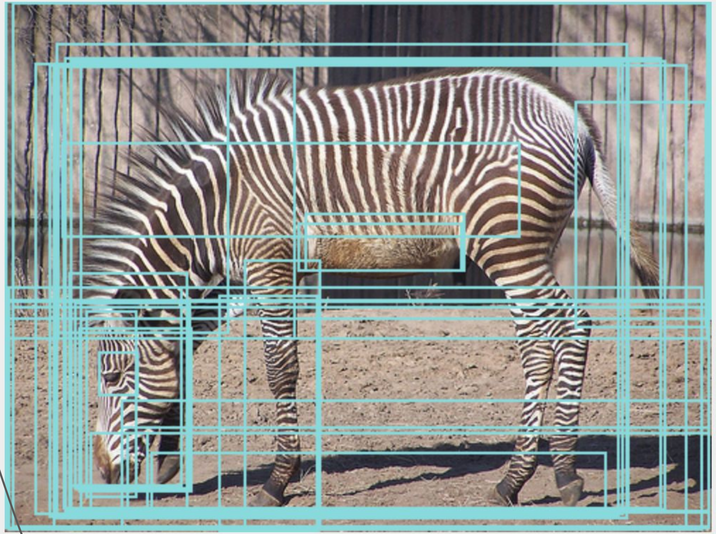
- This is a zebra
- Leg of a zebra
- Tail of a zebra
- Head of a zebra
- Ear of a zebra
- Mouth of a zebra
- Face of a zebra
- Young zebra sniffing the ground
- Striped head of zebra
- Pointed striped neck hair on zebra

Attributes

- zebra head is Striped
- zebra hair is Striped
- wooden fence is Old
- zebra is female
- black zebra is Black
- black zebra is white
- dirt field is dirt
- zebra is black
- zebra is white
- white zebra is white
- white zebra is black
- stripes is black

Relationships

- leg of a zebra
- zebra sniffing ground
- zebra hair ON zebra
- shadow ON ground
- belly ON zebra
- zebra IN corral
- zebra casting shadow
- black zebra walking through dirt
- zebra standing in dirt



Question Answers

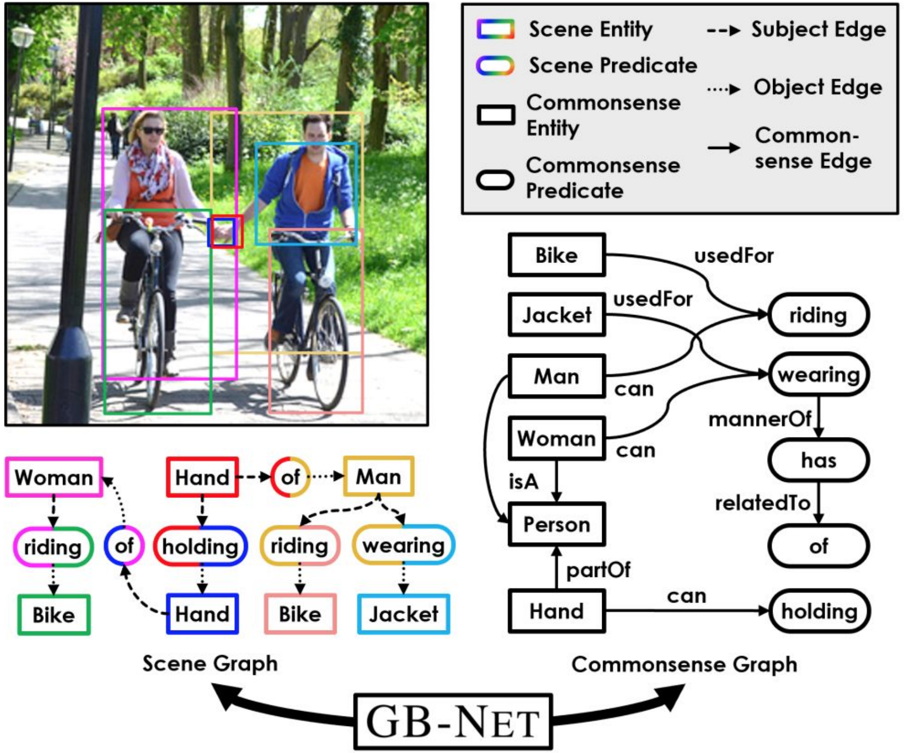
When does the scene occur?	Daytime.
What kind of animal is this?	A zebra.
Where are the shadows?	On the ground.
How many zebras are there?	One.
What is the ground made of?	Dirt.

similar to relationships in NEIL

Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations, Krishna et. al, 2016

GB-NET: from scene graphs to CSK graphs

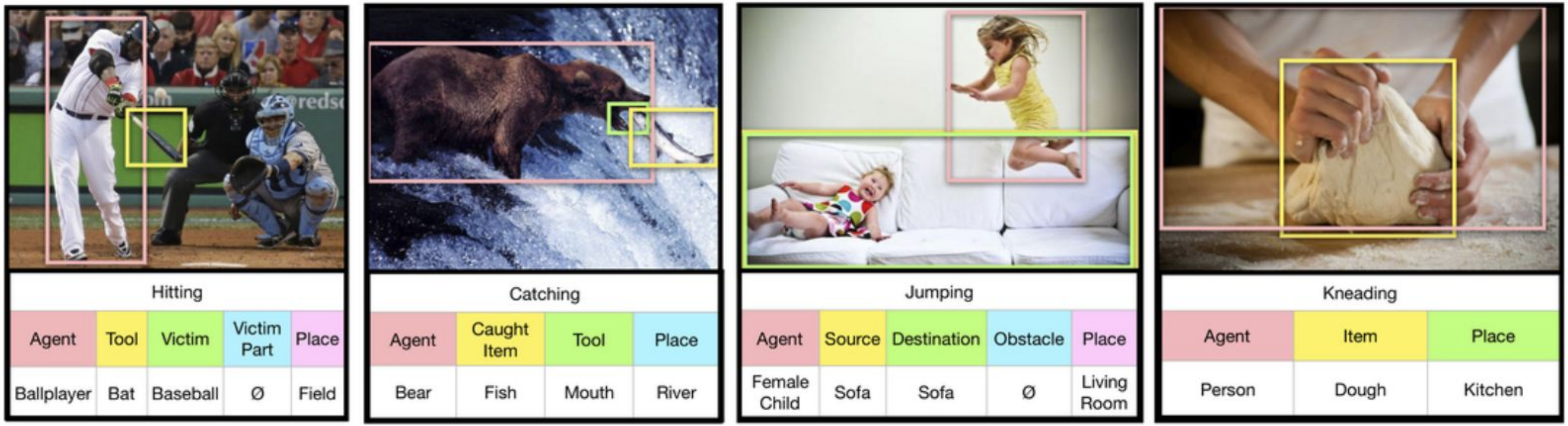
scene graphs are image dependent



commonsense graphs are image independent

Bridging Knowledge Graphs to Generate Scene Graphs, Zareian et. al, ECCV 2020

Situation with grounding data: SWiG

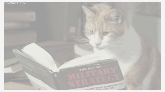


action specific tuples (frames)

Grounded Situation Recognition, Pratt et. al, 2020

Agenda

Part-I



Extract triples and organize

Visual commonsense knowledge

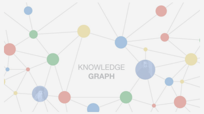
Rich complementary knowledge

Visual vs textual knowledge:
- Visual KG captures unmentioned knwl.
- Might also suffer from reporting bias

Future research directions:
- Extract (interaction) knowledge from videos
- More never-ending approaches like NEIL

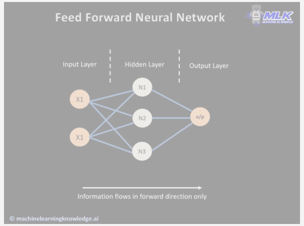
Part-II

symbolic representation



Neural helps symbolic

symbolic helps neural



neural representation of encoded knowledge in the hidden layers

Part-III

Evaluate KG

symbolic query, matching reqd

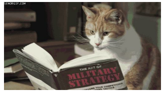
knowledge triple(s) retrieved

Task

decode query relevant knowledge

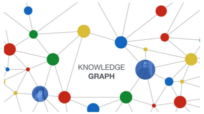
embed symbolic query, auto matching

From Knowledge base construction to Deep learning



Extract triples and organize

symbolic representation



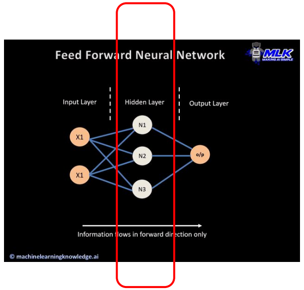
symbolic query, matching reqd

knowledge triple(s) retrieved

Task

decode query relevant knowledge

Train to fill missing word etc.



embed symbolic query, auto matching

neural representation of encoded knowledge in the hidden layers

Agenda

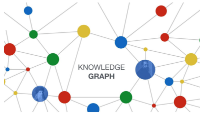
Part-I

Extract triples and organize

Train models

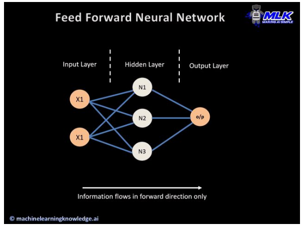
Part-II

symbolic representation



Neural helps symbolic

symbolic helps neural



neural representation of encoded knowledge in the hidden layers

Part-III

Evaluate KG

symbolic query, matching reqd

knowledge triple(s) retrieved

Task

decode query relevant knowledge

embed symbolic query, auto matching

5 min tour de Neural Language models

inference

Task: 😃😟

Conferences make you want to attend them

BERT

softmax

feed forward

finetune first
infer later

Transformer

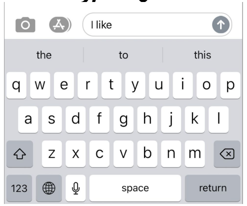
architecture

to

P(x / I like)

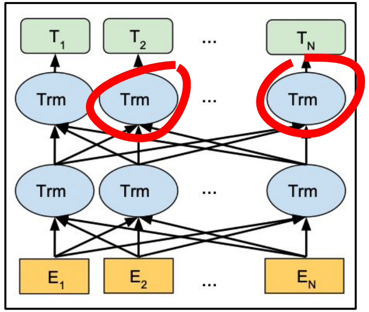
decode

Task: typing assist



GPT

training



like tutorials

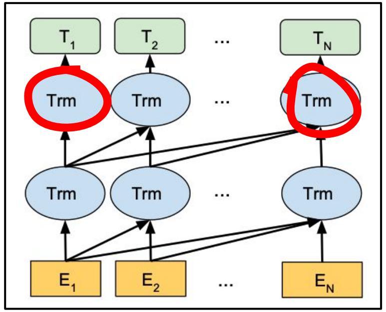
Bidirectional Encoder

I _ to attend _

I like to attend tutorials

Autoregressive Decoder

<s> I like to attend

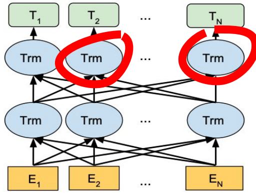


text corpus

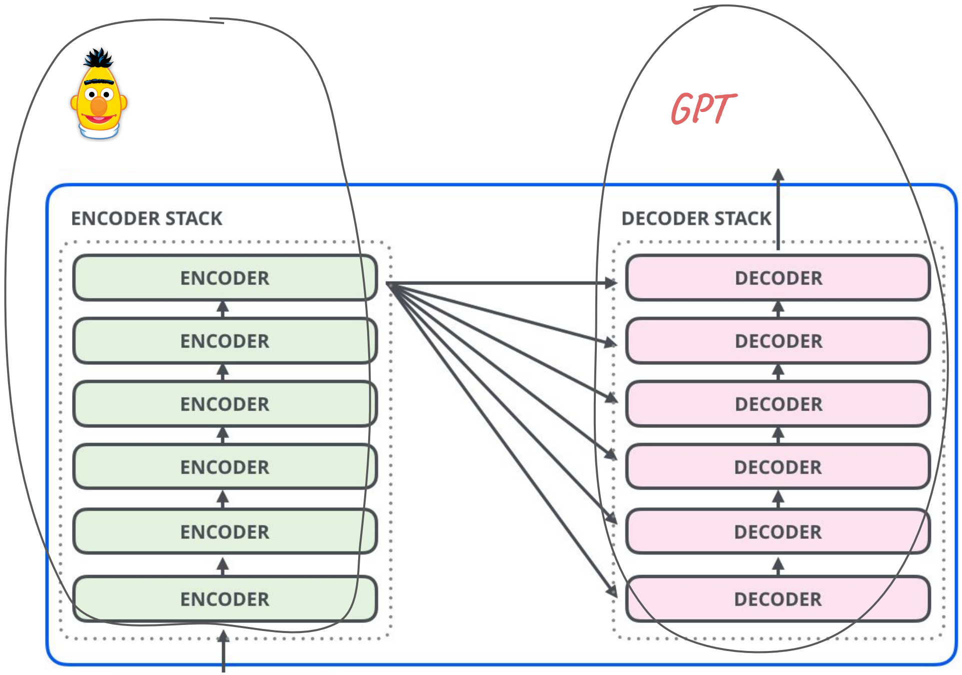
Wikipedia
Die freie Enzyklopädie

tour de Transformers

transform to a really good hidden representation



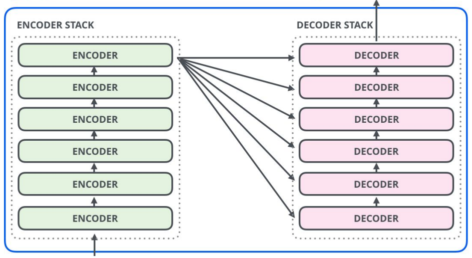
different layers might capture different low/high level aspects such as texture, color, shape, size or emotion, gender

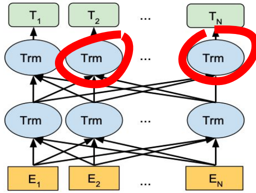


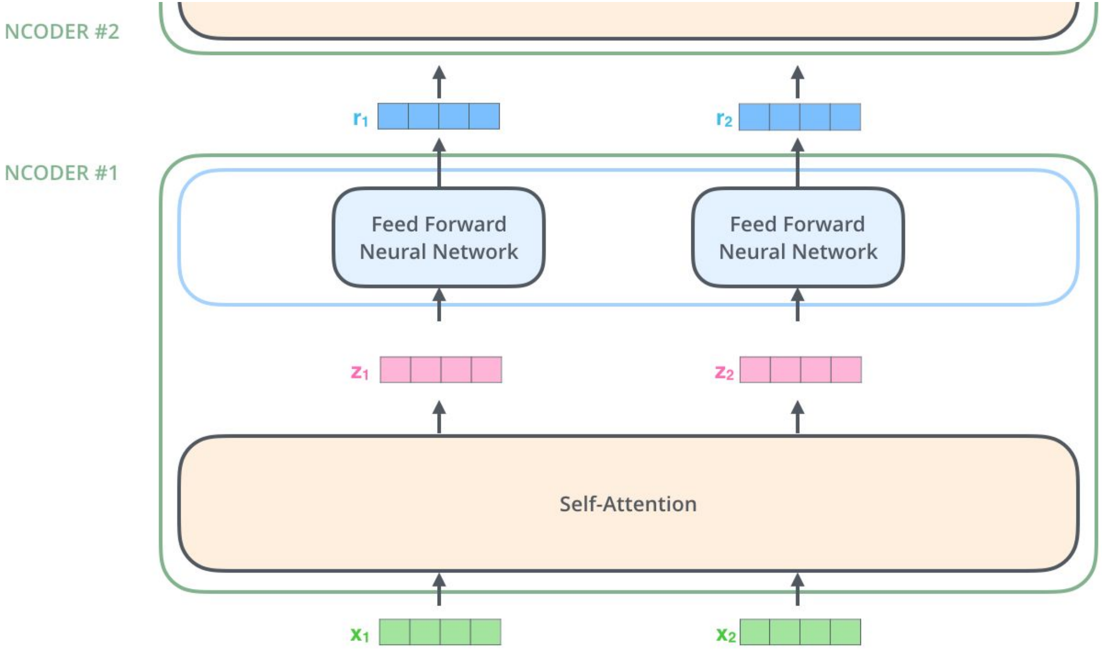
Credit: All the nice Transformer illustrations taken from http://jalammar.github.io/illustrated-transformer/

tour de Encoders in transformer

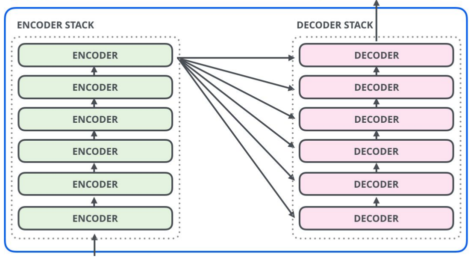
transform to a really good hidden representation





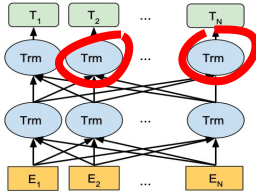


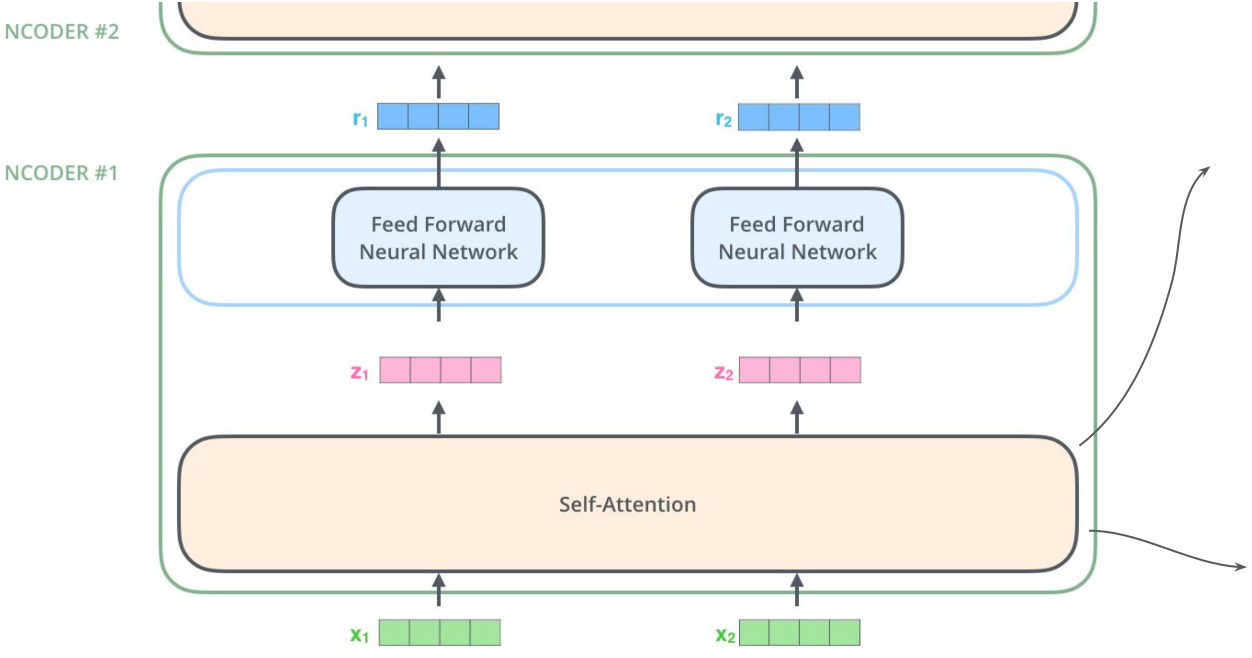
Credit: All the nice Transformer illustrations taken from http://jalammar.github.io/illustrated-transformer/

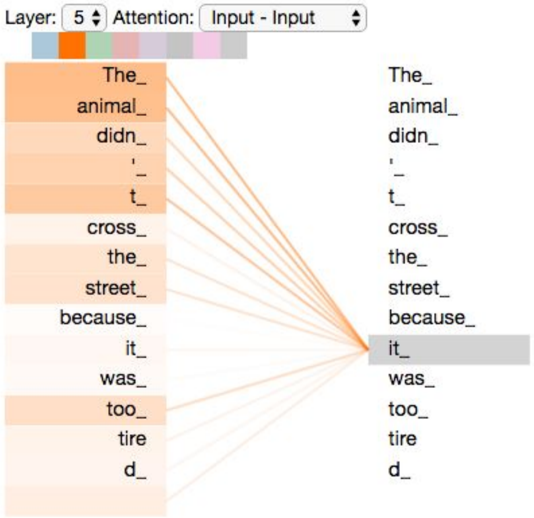


tour de Encoders in transformer

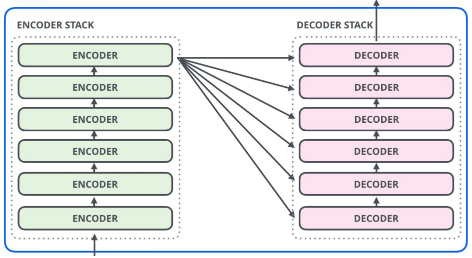
transform to a really good hidden representation





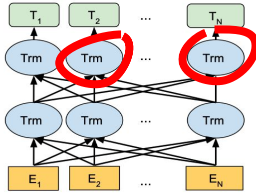


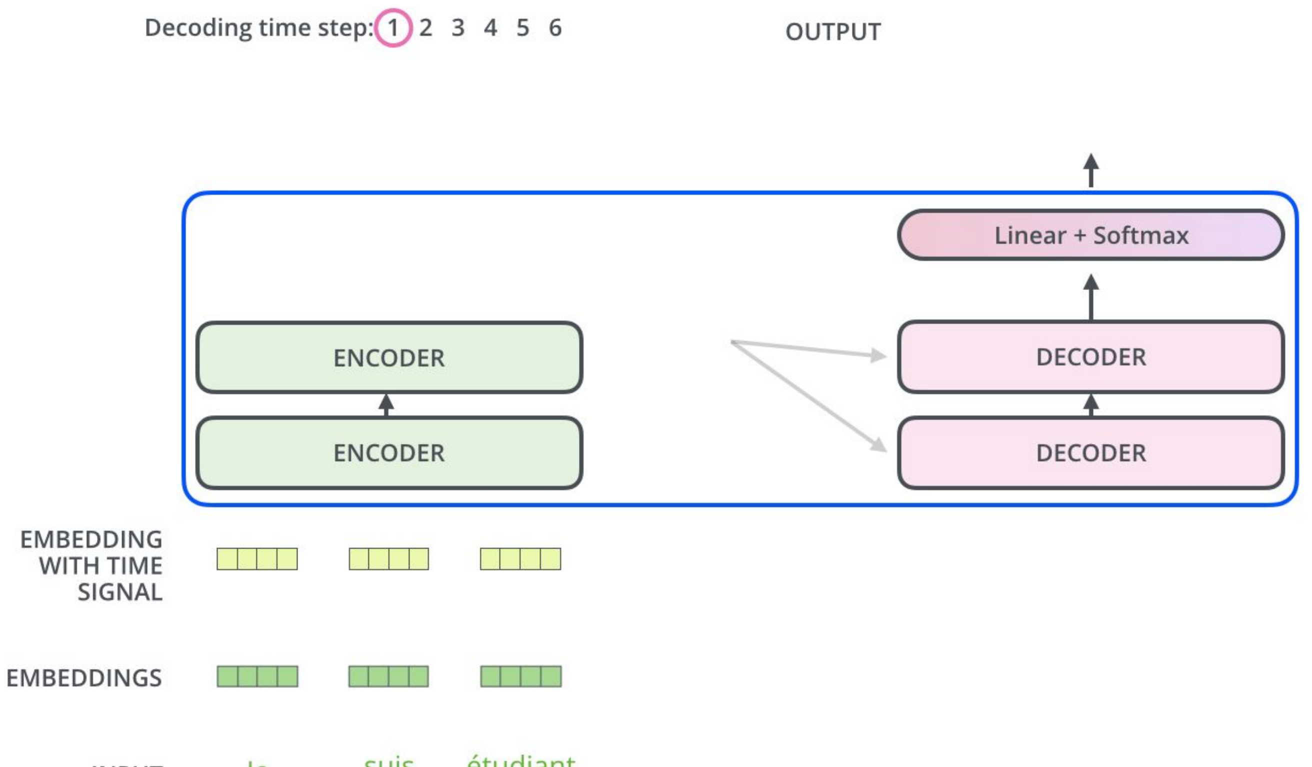
Credit: All the nice Transformer illustrations taken from
http://jalammar.github.io/illustrated-transformer/



tour de Encoders in transformer

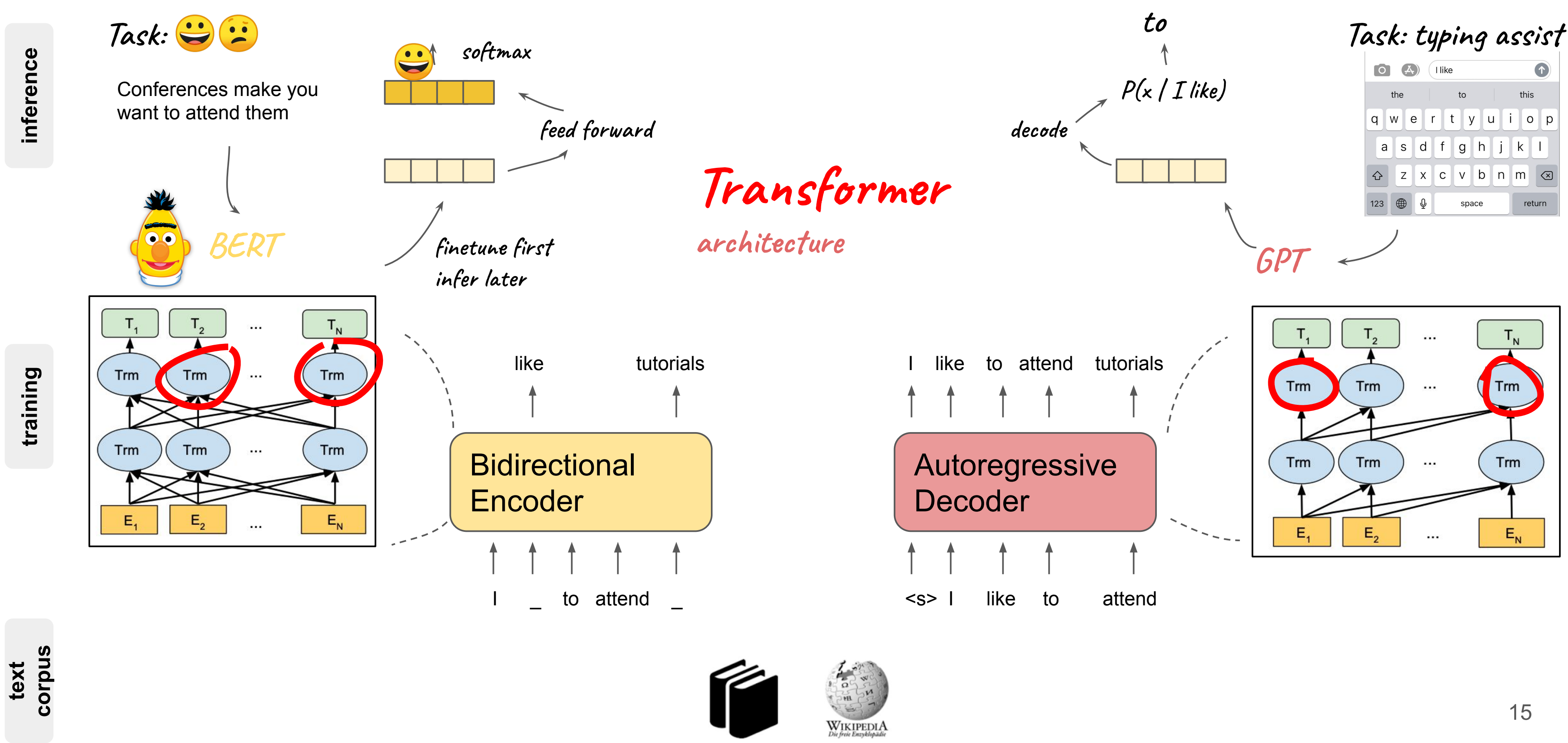
transform to a really good hidden representation





✓ (de) tour de models

current state of the art models: T5 (encoder + decoder architecture) and GPT3



Agenda

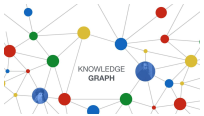
Part-I

Extract triples and organize

✓ *Train* models

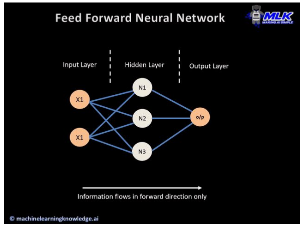
Part-II

symbolic representation



Neural helps symbolic

symbolic helps neural



neural representation of encoded knowledge in the hidden layers

Part-III

Evaluate KG

symbolic query, matching reqd

knowledge triple(s) retrieved

Task

decode query relevant knowledge

embed symbolic query, auto matching

1 of 4 : concept knowledge in neural LMs

untuned model[3] is not great

Tokens: [CLS] everyone knows that a bear has [MASK] . [SEP]

All Results:

1: teeth - 34.521595%

2: fangs - 15.836702%

3: wings - 5.113015%

4: horns - 4.042341%

5: claws - 3.797797%

6: eyes - 3.060219%

7: legs - 2.741149%

8: fur - 1.653655%

9: ears - 1.173016%

tuned model[4] is much better (like with any neural LM)

Context	Human Response	PF	RoBERTa-L Response	p_{LM}
Everyone knows that a bear has ___	fur	27	teeth	.36
	claws	15	claws	.18
	teeth	11	eyes	.05
	cubs	7	ears	.03
	paws	7	horns	.02

low correlation with human elicit properties but are coherent.

can also distinguish based on properties: “X has fur” vs “X has fur and is big”

[3] Bar Ilan demo., as of 2021 : link

[4] Weir et al., 2020

[5] Forbes et al., 2019

1 of 4 : concept knowledge in neural LMs

BERT

untuned model[3]

Tokens: [CLS] everyo ws th

All Results:

1: teeth - 34.521595%

2: fangs - 15.836702%

3: wings - 5.113015%

4: horns - 4.042341%

5: claws - 3.797797%

6: eyes - 3.060219%

7: legs - 2.741149%

8: fur - 1.65365

9: ears - 1.173016%

"neural language representations still only learn associations that are explicitly written down"[5],

even after being explicitly trained on a knowledge graph of objects and affordances.

Context	Human Response	PF	RoBERTa-L Response	p_{LM}
Everyone knows that a bear has ___	fur	27	teeth	.36
	claws	15	claws	.18
	teeth	11	eyes	.05
	cubs	7	ears	.03
	paws	7	horns	.02

"Perceptual or visual concepts such as *smooth*, can't be learned from text alone"[4],

ed on properties: "X has fur" vs "X has fur and is big"

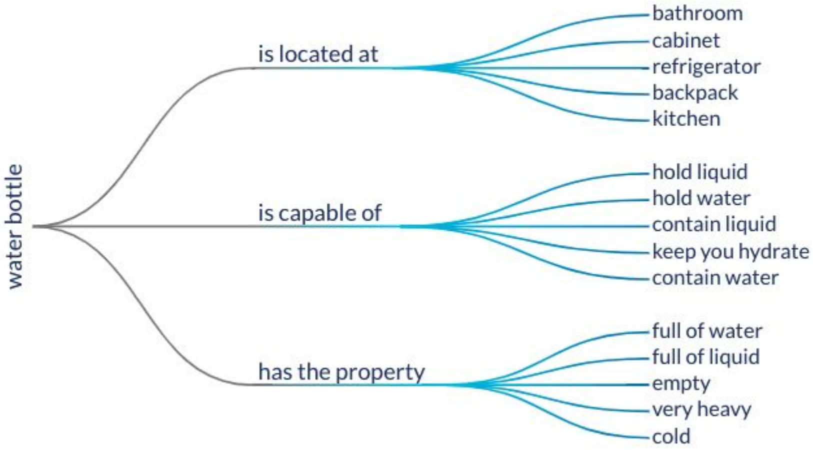
[3] Bar Ilan demo., as of 2021 : link

[4] Weir et al., 2020

[5] Forbes et al., 2019

2 of 4 : multi-relational & visual knowledge in neural LMs GPT

autoregressive model[2]
(GPT fine-tuned on ConceptNet)



promising knowledge

[1] (again, untuned is bad)
AllenNLP demo GPT2, as of 2021 : link

[2] COMeT demo., Bosselut et. al, as of 2021: link

2 of 4 : multi-relational & visual knowledge in neural LMs GPT

autoregressive model[2]

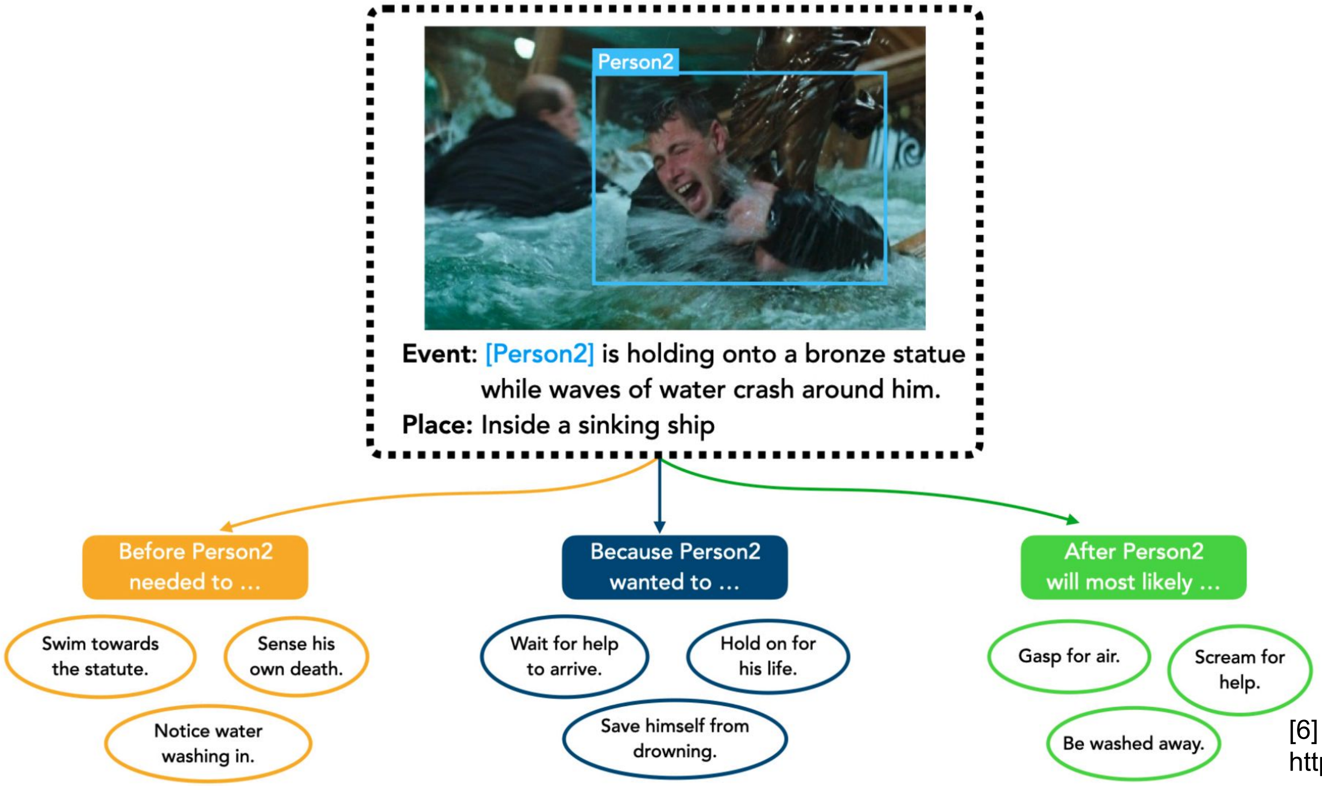
"Do not handle mutual exclusivity well and suffer from frequency bias (in general the outputs may be incoherent or inconsistent) "[4],

"Perceptual or visual concepts still hard to learn"[4],

[1] AllenNLP demo GPT2, as of 2021 : link

“Learn Perceptual or visual concepts”[4],

Task: Generate events before, after and intents at present given an image, <u>and</u> a description of the event in the image, <u>and</u> a plausible scene/location. Uses visual and language transformer.



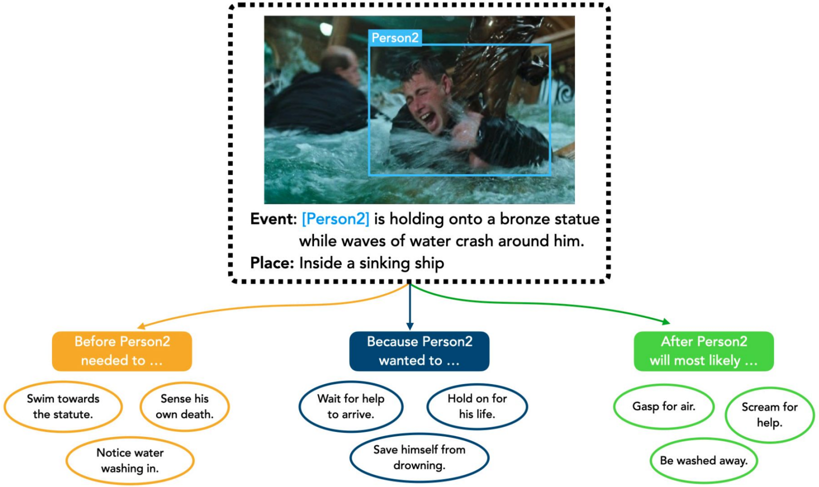
[6] Park et al., ECCV 2020
https://visualcomet.xyz

“Learn Perceptual or visual concepts”[4],

Task: Generate events before, after and intents at present given an image, and a description of the event in the image, and a plausible scene/location

Great start, future models could make fewer assumptions.

Output can still be inconsistent and incoherent.



[6] Park et al., ECCV 2020
https://visualcomet.xyz

3 of 4: neural LMs for CSKG completion

Commonsense Knowledge Mining from Pretrained M

Joshua Feldm

Similar ideas have been applied to correct a KG based on neural LM perplexity

Candidate Sentence S_i	$\log p(S_i)$
“musician can playing musical instrument”	-5.7
“musician can be play musical instrument”	[illegible]
“musician often play mus	[illegible]
“a musician can play a m	[illegible]

However, LMs can generate fictitious facts (distributionally similar but factually wrong)

Table 1: Example of [illegible]erating candidate sentences. Several enumerated sentences for the triple (musician, CapableOf, play musical instrument). The sentence with the highest log-likelihood according to a pretrained language model is selected.

Language Models as Knowledge Bases?

Fabio Petroni[1] **Tim Rocktäschel**[1,2] **Patrick Lewis**[1,2] **Anton Bakhtin**[1]
Yuxiang Wu[1,2] **Alexander H. Miller**[1] **Sebastian Riedel**[1,2]
[1]Facebook AI Research
[2]University College London

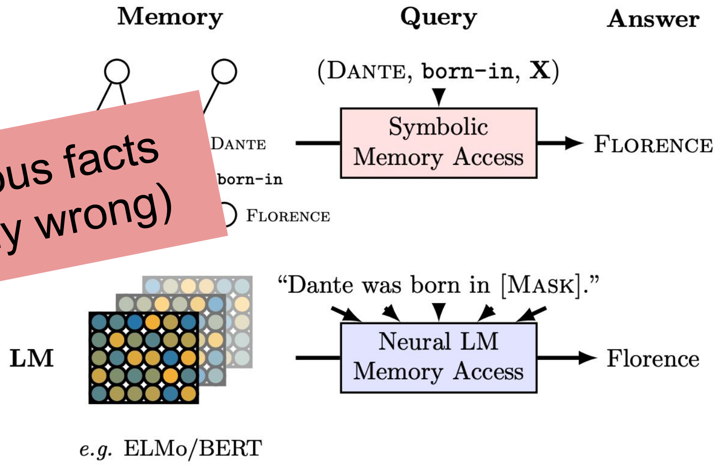
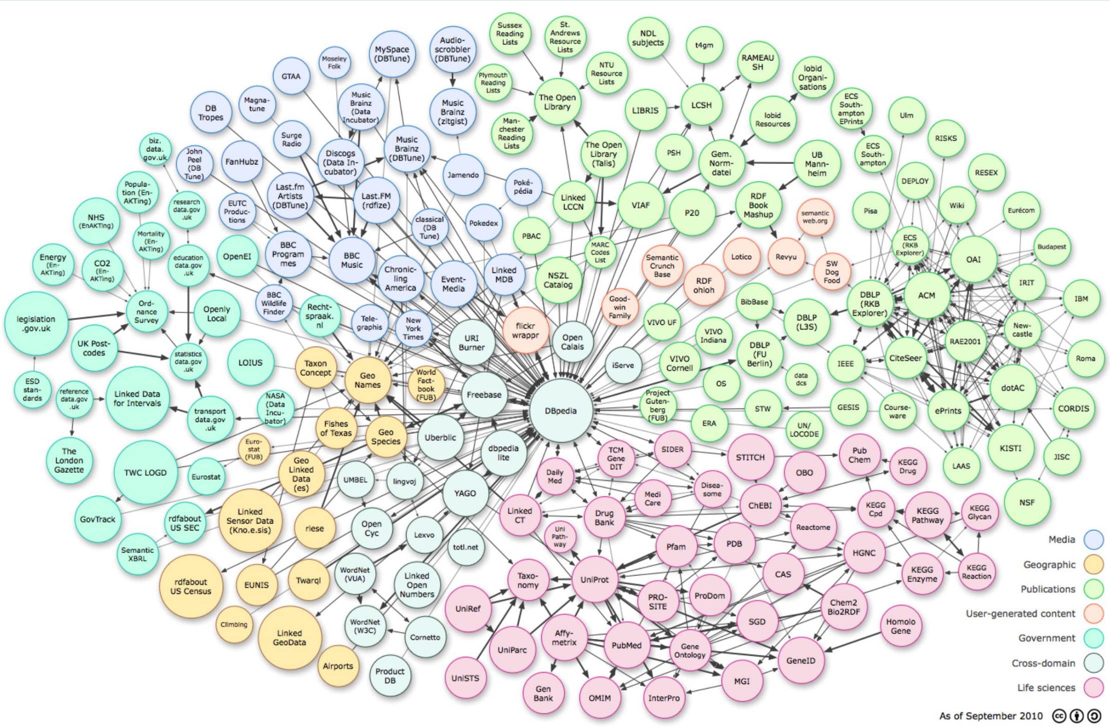


Figure 1: Querying knowledge bases (KB) and language models (LM) for factual knowledge.

4 of 4: fusing multiple CKGs

- Entity linkage: linking multiple taxonomies online is a massive, unsolved task.



4 of 4: neural LMs to ~~fuse~~ use multiple CKGs

- Entity linkage: linking multiple taxonomies online is a massive, unsolved task.
- Attention: need to first retrieve relevant subgraph.
- Multi-task learning: scalable, and embeds knowledge (e.g., UNICORN)

KNOWLEDGE GRAPH	SOCIALIQA
ATOMIC	**75.0**
CONCEPTNET	74.3
BOTH	74.8
single task	73.8

Entire KG (verbalized triples) is learned to complete as a task. So model trained on QA as well as KG prediction task.

No KG, model only trained on QA task

UNICORN on RAINBOW: A Universal Commonsense Reasoning Model on a New Multitask Benchmark, Lourie et al., AAAI 2020

Pros/cons of using neural over symbolic KGs

Knowledge acquisition	KG completion	KG correction	~~Fuse~~ use multiple KGs
✓	✓	✓	✓

Pros:

1. Real tasks/ queries representation space might be different, and it is difficult to align with the representation space/ or query the KG.
2. Typically, KGs do not come with context. This makes the KG lookup even more difficult. For example, things break when they fall but soft things do not.

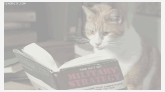
Pros/cons of using neural over symbolic KGs

Cons:

1. Symbolic KGs are more interpretable and easily debuggable, but neural models are hard to probe.
2. Promising direction of multi-task learning for using multiple KGs, but more work is needed.
3. LMs can generate fictitious facts-- this requires more work. e.g., grounding the knowledge to an established source such as Wikipedia.
4. More work is required (BOTH in symbolic and neural) to acquire perceptually grounded/ unmentioned knowledge, e.g, visual COMeT with fewer assumptions in the input -- and we need to make the output more consistent.

Agenda

Part-I

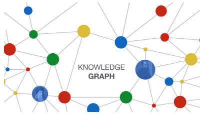


Extract triples and organize

✓ *Train* models

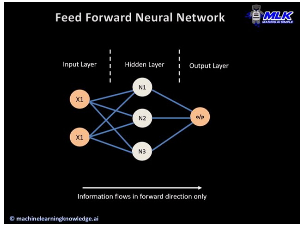
Part-II

symbolic representation



✓ Neural helps symbolic

symbolic helps neural



✓ neural representation of encoded knowledge in the hidden layers

Part-III

Evaluate KG

symbolic query, matching reqd

knowledge triple(s) retrieved

Task

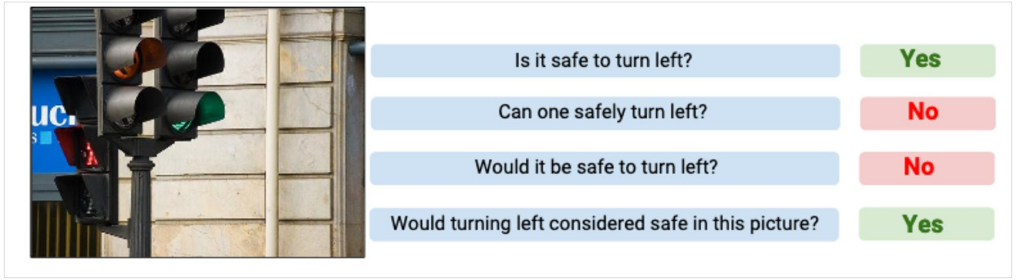
decode query relevant knowledge

embed symbolic query, auto matching

Can CSK help neural models

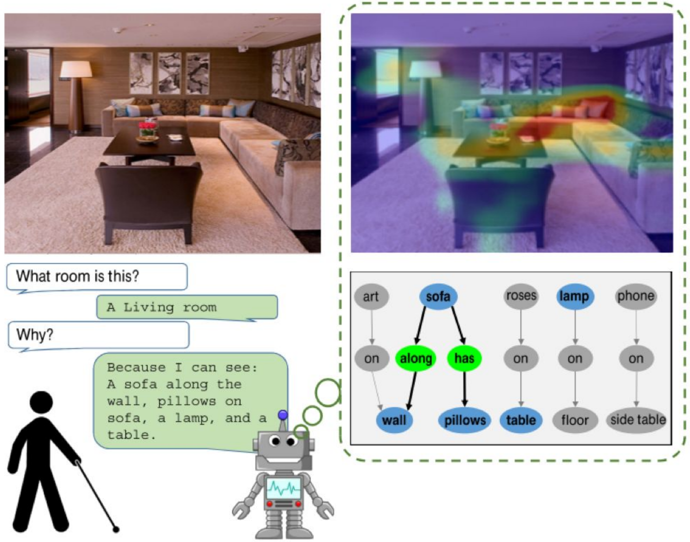
Robustness[d1]

Generating adversarial examples guided by commonsense knowledge[d2]



Explainability[d3]

Using attention map generated by a QA model (top right) to identify relevant components of a scene graph[d4]



[d1]: Cycle-Consistency for Robust Visual QA, Shah et. al 2019
[d2]: AdvEntuRe: Adversarial Training for Textual Entailment with Knowledge-Guided Examples, Kang et. al 2018
[d3] Generating Natural Language Explanations for Visual QA Using Scene Graphs and Visual Attention, Ghosh et al., 2018
[d4] Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations, Krishna et. al, 2016

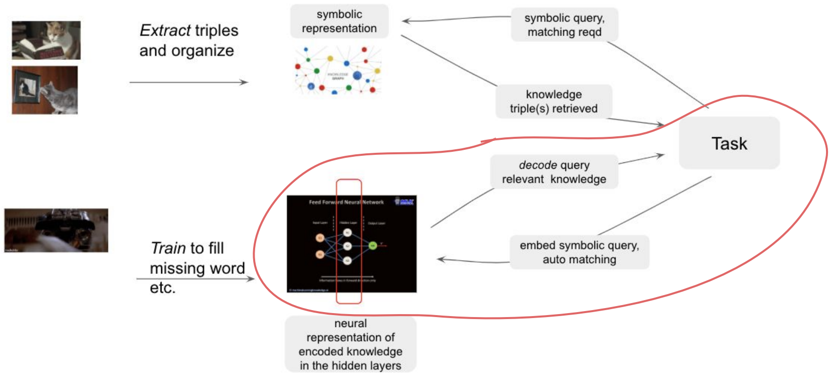
Can CSK help neural models

Limited training data[d5]

Inject commonsense knowledge[d6,d7,...d10]
to compensate for limited training data

Difficult to find training data for all types of scenarios, esp. rarely mentioned rules and facts

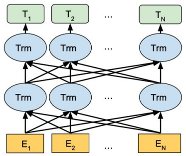
- Are shiny surfaces typically hard?
- What's bigger the moon or a wolf?
- If I put my socks in the drawer,
  will they still be there tomorrow?



[d5]: Tandon et. al 2018 : ProStruct
[d6]..[d10]: upcoming slides

Injecting commonsense knowledge into DL models

Task



injection method

scoring
attention

multitask learning

augmented input

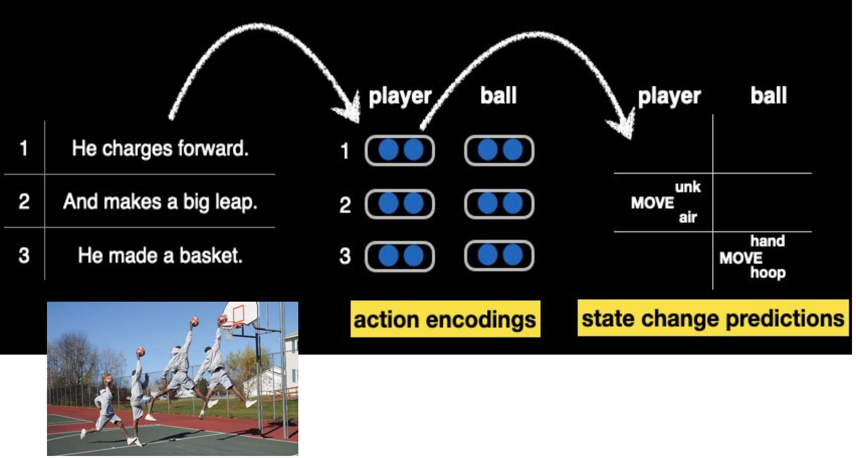
commonsense knowledge

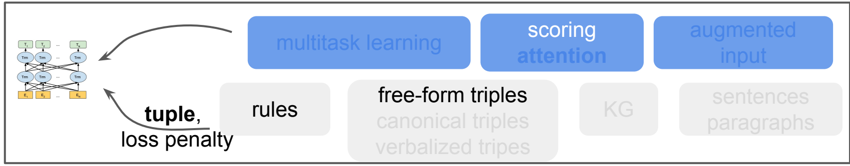
free-form triples
canonical triples
verbalized tripes

rules

preconstructed KG
on-the-fly generated KG

sentences
paragraphs





- During beam search decoding to find globally consistent results, probability mass moves away from implausible states.
- Model has seen insufficient data to learn these correlations, so use commonsense to steer away from unrealistic states.

$$P(\pi_t^j \mid e_j, topic)$$

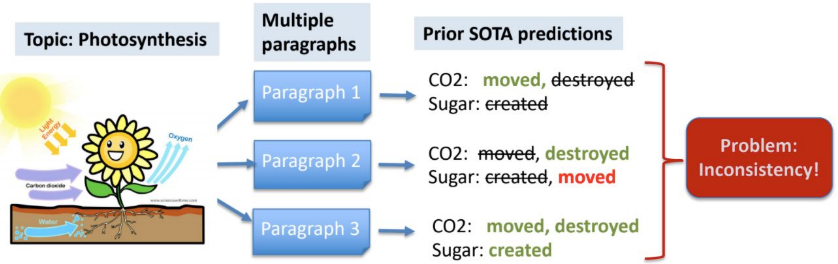
Things cannot move if they don’t exist yet

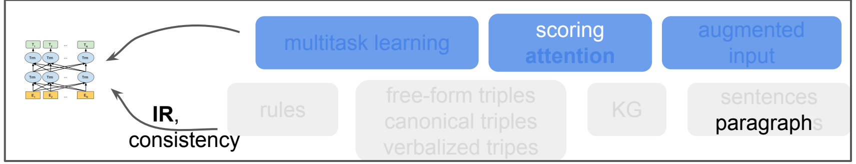
Ball will be destroyed at step 2 (less likely)

3 rules from SUMO ontology

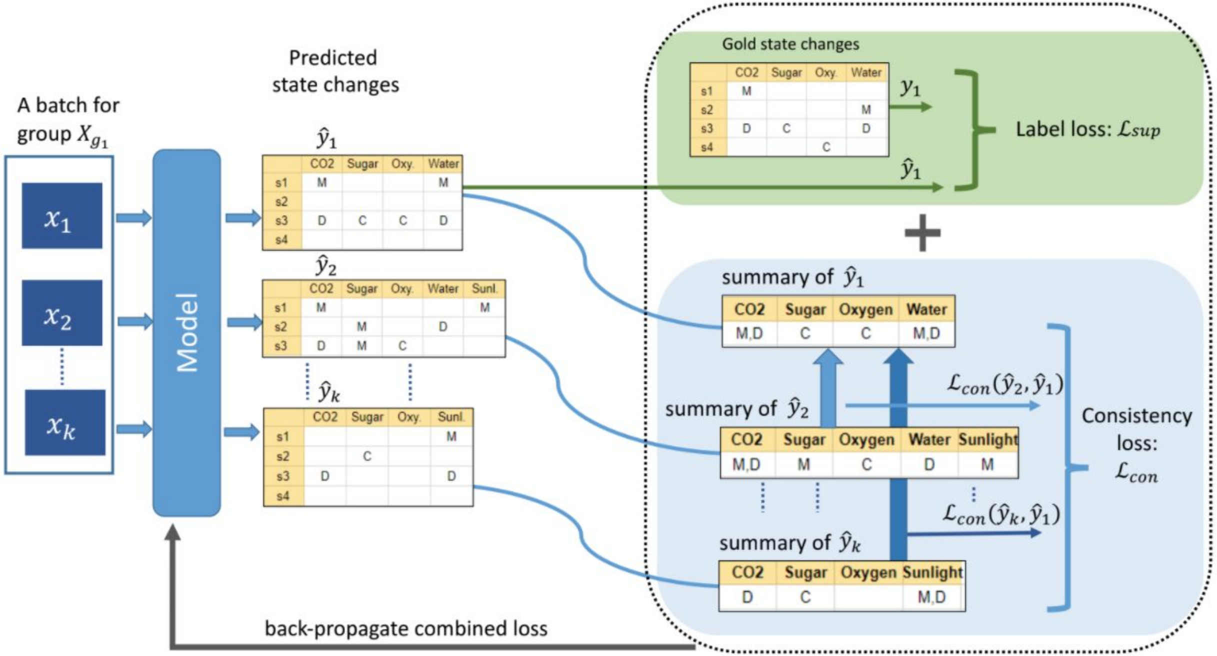
KB

Reasoning about Actions and State Changes by Injecting Commonsense Knowledge, Tandon et al EMNLP 2018

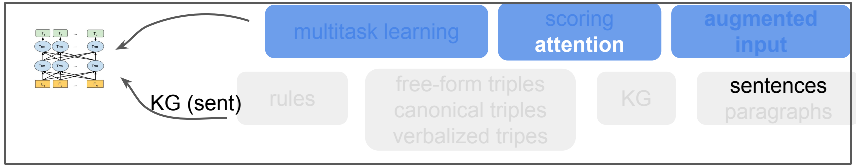




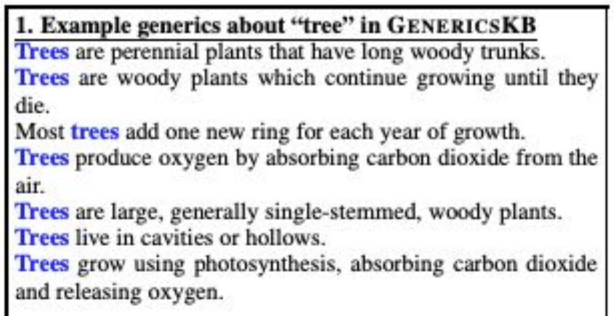
Adds consistency loss across paragraphs (derivable from a *CKG of paragraphs*) while training an end2end model.



Be Consistent! Improving Procedural Text Comprehension using Label Consistency. Du et al NAACL 2019



There is a recent thrust towards **unstructured entity specific sentence KGs**. It resolves the IR issues, and text can represent more complex commonsense knowledge.



1. Example generics about "tree" in GENERICSKB
Trees are perennial plants that have long woody trunks.
Trees are woody plants which continue growing until they die.
Most trees add one new ring for each year of growth.
Trees produce oxygen by absorbing carbon dioxide from the air.
Trees are large, generally single-stemmed, woody plants.
Trees live in cavities or hollows.
Trees grow using photosynthesis, absorbing carbon dioxide and releasing oxygen.

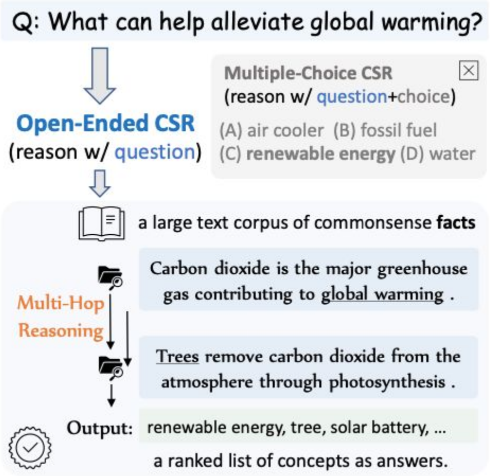
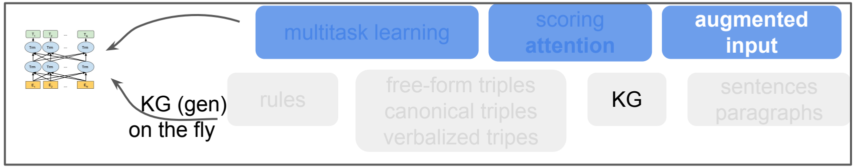


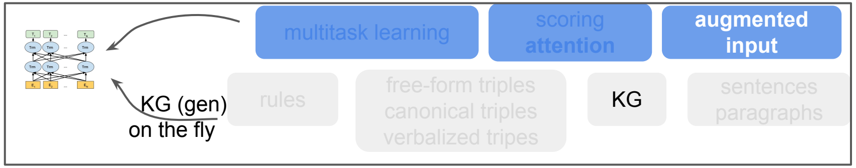
Figure 1: We study the task of open-ended commonsense reasoning (OpenCSR), where answer candidates are not provided (as in a multiple-choice setting). Given a question, a reasoner uses multi-hop reasoning over a knowledge corpus of facts, and outputs a ranked list of concepts from the corpus.

GenericsKB: A knowledge base of generic sentences. Bhakthavatsalam et al. arxiv 2020 https://arxiv.org/pdf/2005.00660.pdf
Differentiable Open-Ended Commonsense Reasoning , Lin et al. arxiv 2020 https://arxiv.org/pdf/2010.14439.pdf

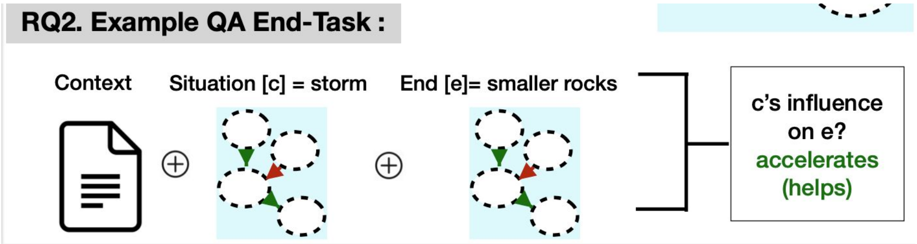


On the fly KG "generation" is another recent direction. When the KG is augmented to the input, QA performance boosts.

EIGEN model. Rajagopal et al. arxiv 2020



On the fly KG "generation" is another recent direction. When the KG is augmented to the input, QA performance boosts.



EIGEN model. Rajagopal et al. arxiv 2020

Agenda

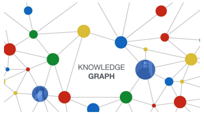
Part-I

Extract triples and organize

✓ *Train* models

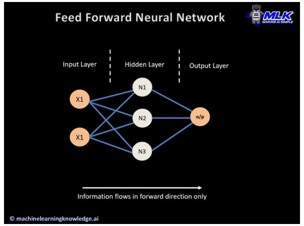
Part-II

symbolic representation



✓ Neural helps symbolic

✓ symbolic helps neural



✓ neural representation of encoded knowledge in the hidden layers

Part-III

Evaluate KG

symbolic query, matching reqd

knowledge triple(s) retrieved

Task

decode query relevant knowledge

embed symbolic query, auto matching

Commonsense for Interactive learning (LeapOfThought)

inference time (current models make mistakes that can be corrected)

Ask the AI a yes/no question

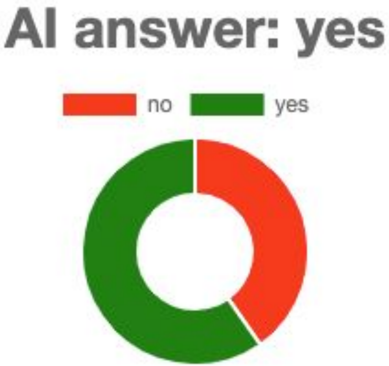
Does a whale have bellybutton?

Ask the AI a yes/no question

Does a whale have bellybutton?

Add rules to teach the AI if it answered incorrectly!

Whale is a mammal.



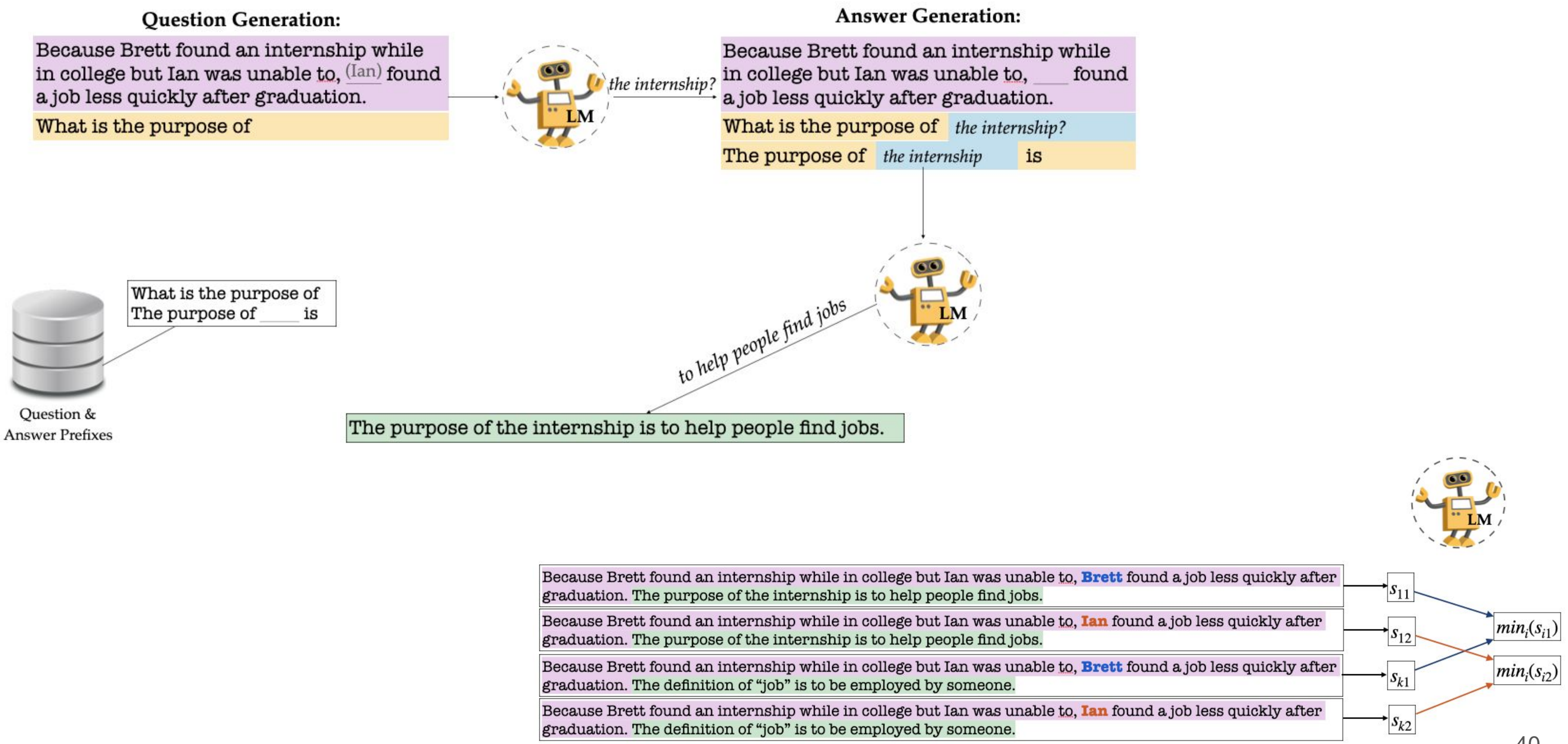
Commonsense for Interactive learning (LeapOfThought)

inference time (current models make mistakes that can be corrected)

- + Clearly shows that models will lack CSK and will benefit from having it.
- Model throws away the valuable user feedback after using locally.
- (risk) Model may learn false or fake information if the user tricks it.

Leap-Of-Thought: Teaching Pre-Trained Models to Systematically Reason Over Implicit Knowledge, Talmor et al , NeurIPS 2020

Generating required commonsense on the fly by querying LM



Unsupervised Commonsense QA with Self-Talk, Shwartz et al EMNLP 2020

One model that solves multiple commonsense tasks

TRANSFER	αNLI	COSMOSQA	HELLASWAG	PIQA	SOCIALIQA	WINOGRANDE
multitask	78.4	81.1	81.3	80.7	74.8	72.1
fine-tune	79.2	82.6	**83.1**	**82.2**	75.2	78.2
sequential	**79.5**	**83.2**	83.0	**82.2**	**75.5**	**78.7**
none	77.8	81.9	82.8	80.2	73.8	77.0

UNICORN on RAINBOW: A Universal Commonsense Reasoning Model on a New Multitask Benchmark, Lourie et al., AAAI 2020

Summary

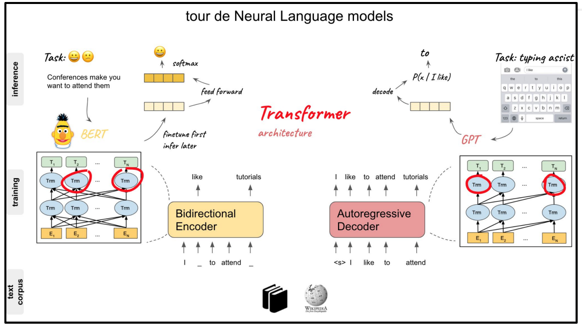
Neural helps symbolic

Contextual, plug-n-play, hard to interpret

Neural methods can help with:
Knowledge acquisition
KG completion
KG correction
~~Fuse~~ use KG

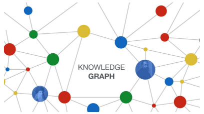
Future research directions:
- multitask learning with multiple KGs
- output needs to be faithful
- making model output coherent

High level overview of neural LMs



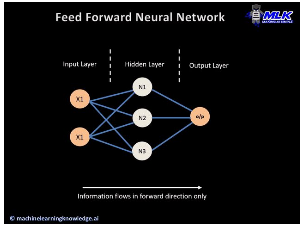
Part-II

symbolic representation



Neural helps symbolic

symbolic helps neural



Symbolic helps neural

Various ways to inject CSK

CSK can help with:
Robustness
Explainability
Limited training data

Future research directions:
- topic specific paragraph KGs
- interactive learning with CSK
- multitask learning unified models