NSNLI: First Workshop on Neuro-Symbolic methods for Natural Language Inference

Is Neuro-symbolic SOTA still a myth for Natural Language Inference

Somak Aditya, Microsoft Research
Maria Chang, IBM Research
Swarat Chaudhuri, UT Austin
Monojit Choudhury, Microsoft Research
Sebastijan Dumančić, KU Leuven

The Goal of the Workshop

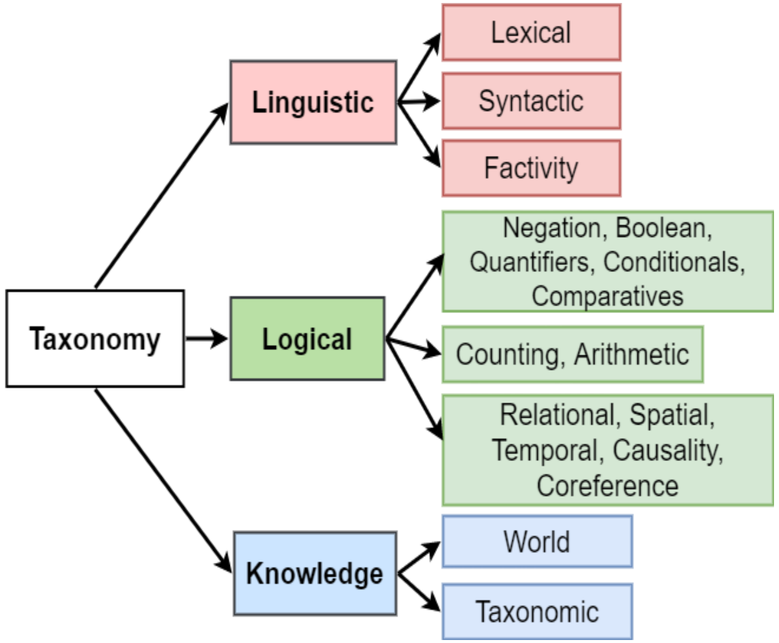
Observation: A discrepancy in performance of large language models on benchmarks vs out-of-distribution simpler examples.

Lack of reasoning. Lack of generalization.

For reasoning, logic may help. *But how? What type of reasoning?*

The NLI Task (a stand-in for NLU)

Text	Hypothesis	Judgement
An older and younger man smiling.	The older man is not smiling	Contradicti on
An older and younger man smiling.	Two men are laughing at the cats playing on the floor.	Neutral
An older and younger man smiling.	An older man is smiling.	Entailment



Task: Given a premise and hypothesis in natural language, identify whether the hypothesis contradicts, entailed by or neutral by the premise.

Critiques of NLI Systems

Jul 2018 | Mar 2019 | Feb 2020

	Breaking NLI	Fragments	Counterfactual NLI
Examples	P: The man is holding *saxophone*. H: The man is holding *electric guitar*.	P: Arthur visited Paris and New York H: Arthur did not visit Paris.	P: Students are inside of a lecture hall. OH: Students are indoors. (E) NH: Students are on the soccer field. (Contradiction)
Tests	Hypothesis and premise varies only one word.	Logical fragments: negation, Boolean, quantification.	Confounding factors: Entity, relations, actions. Semantic/Logical: Evidence, Negation.
Performance Drop	**20% on** LSTM based models	**40% on** Logic Fragments For Pre-trained BERT	**30% on** the New Set for Pre-trained BERT

Motivation: What Linguistic/logical phenomenon does the trained model capture? And observation: these model break easily.

The Reason Behind Such Discrepancy

ML/DL: Overfitting. Lack of generalization.

Logic, KR, Cognitive Science, and Linguistics:

- Lack of reasoning ability. *(Marcus 2018, 2020;)*
- Lack of commonsense knowledge. *(Davis & Marcus CACM 2015; Rao CACM 2021)*
- Lack of grounding and pragmatic abilities. *(Bender & Koller ACL 2020, Linzen ACL 2020)*

Which solution to Adapt?

At Least three Broad Camps

Deep Learning

Complete End-to-End NN Learning

Compile the symbolic knowledge (such as rule) to create data.

- Symbolic Math: Lample and Charton ICLR 2020

Mimic symbolic reasoning within a neural network

- Graph Neural Networks, Xu et.al. ICLR 2020

Probabilistic Logic/SRL

Cascade Neural and then Symbolic

Feed Neural outputs to symbolic reasoners.

No End-to-End Backprop

- Logic Tensor Networks. PSL-VQA AAAI '18, and NSCL 2019 (MIT)

End-to-End Backprop

- Extension of Probabilistic Logic, (**DeepProbLog NeurIPS '19**, NLProlog ACL '19)

Program Synthesis + ATP

Algorithmic Inference with Neural Sub-routines

Symbolic systems makes the final decision. Neural is helping in steps.

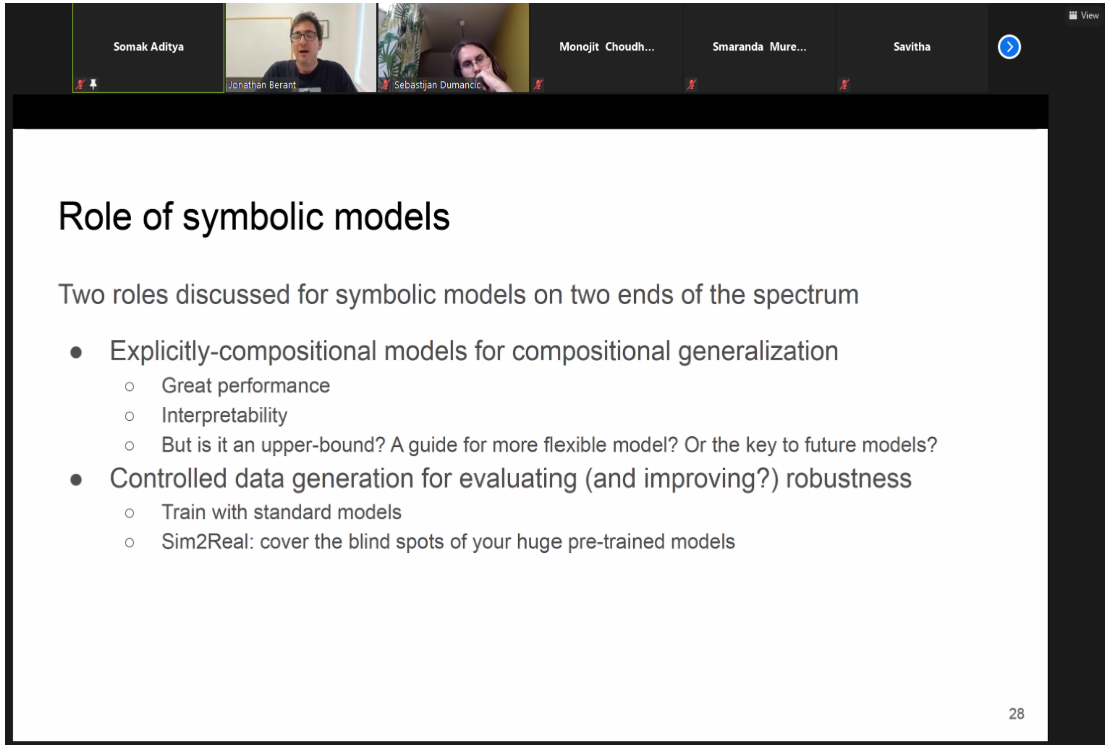
- Example 1: ATP Provers (HOList)
- Example 2: Self-driving cars
- Example 3: Program synthesis+ML

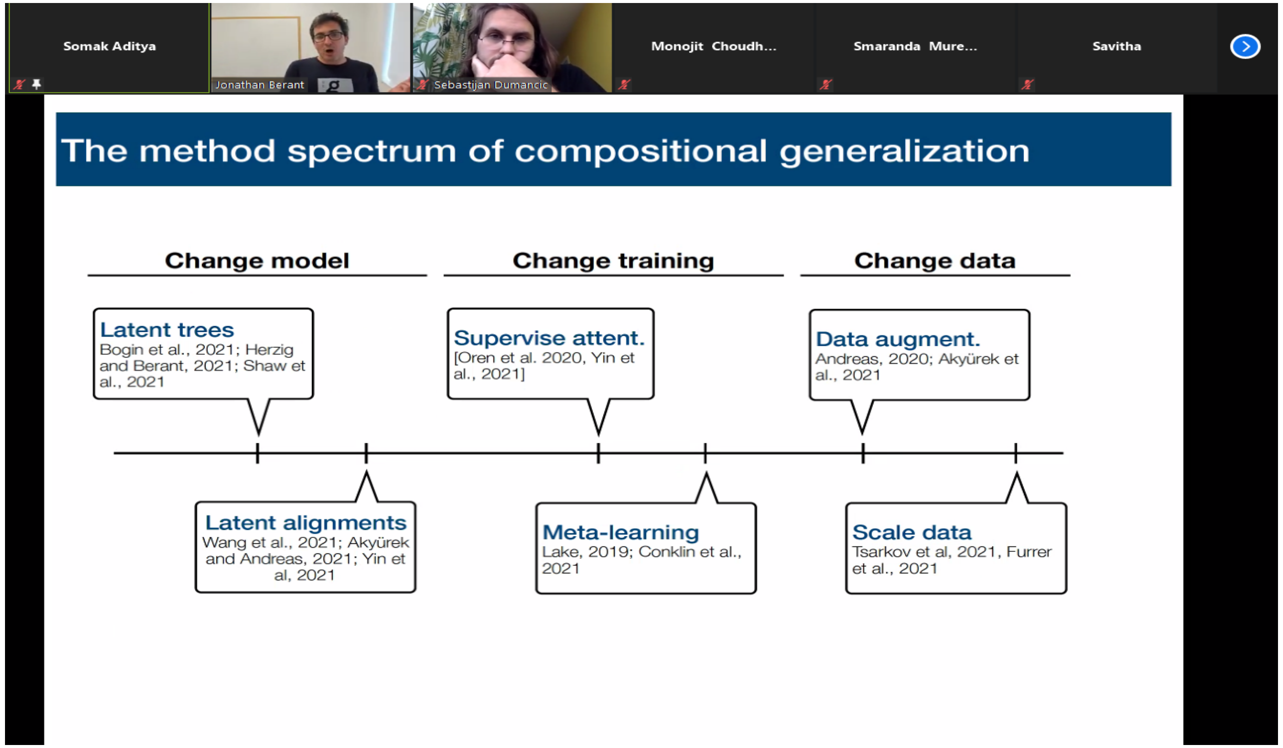
Iterative Programs and ML

- Example 4: Program synthesis+ML

*Robert Engelmore Lecture, Henry Kutz AAAI 2020

Session 1 – Jonathan Berant

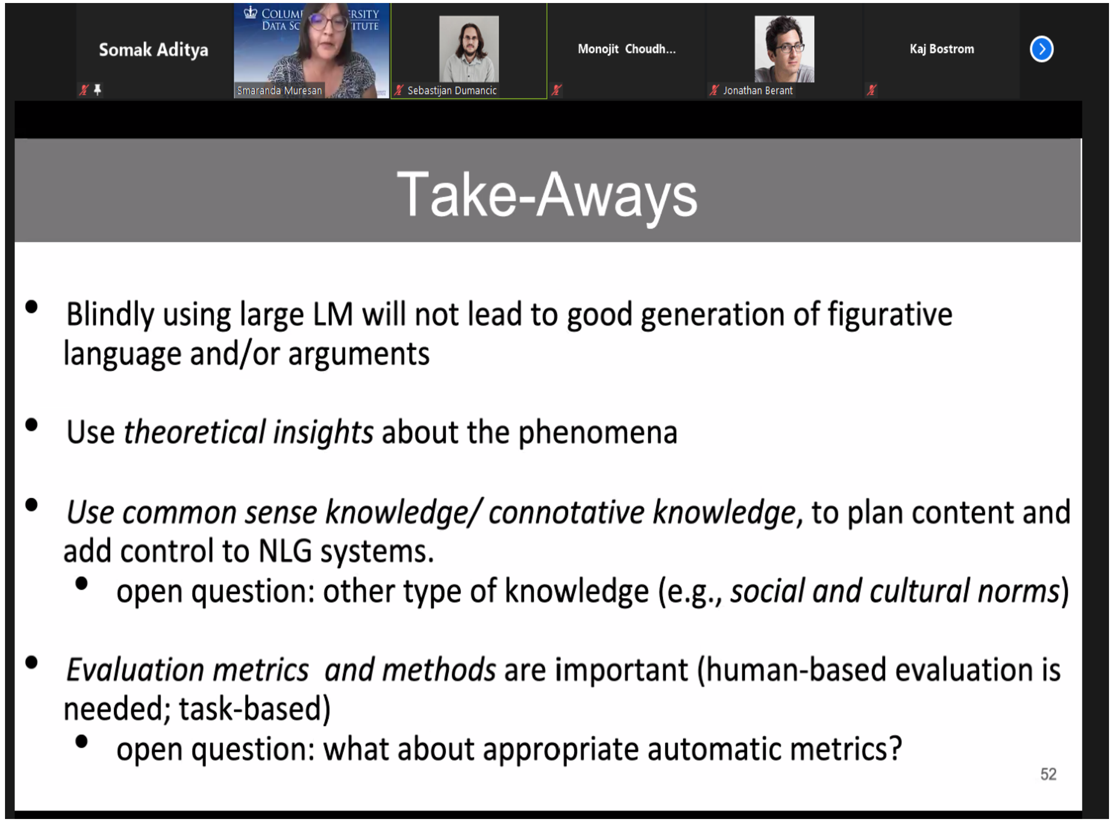


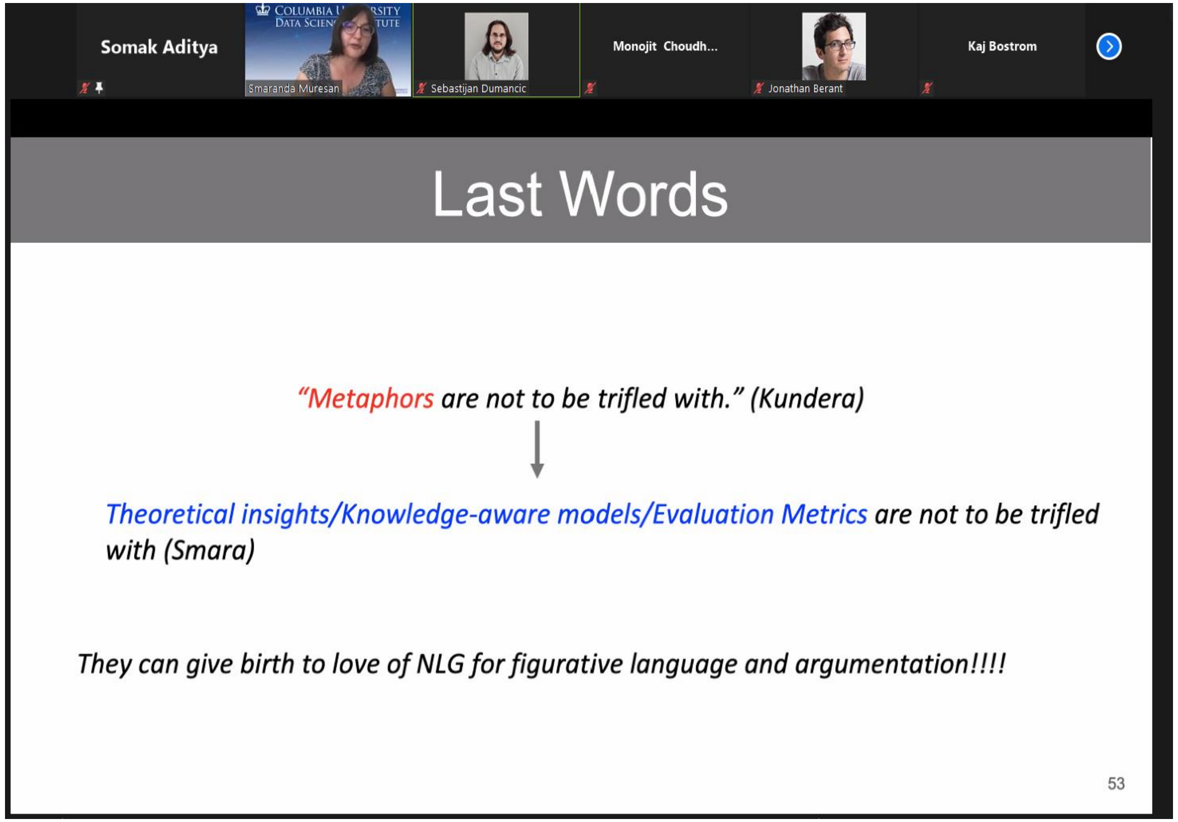


Neuro-symbolic models for understanding complex questions*

*Slides available in the workshop website

Session 1 – Smaranda Muresan

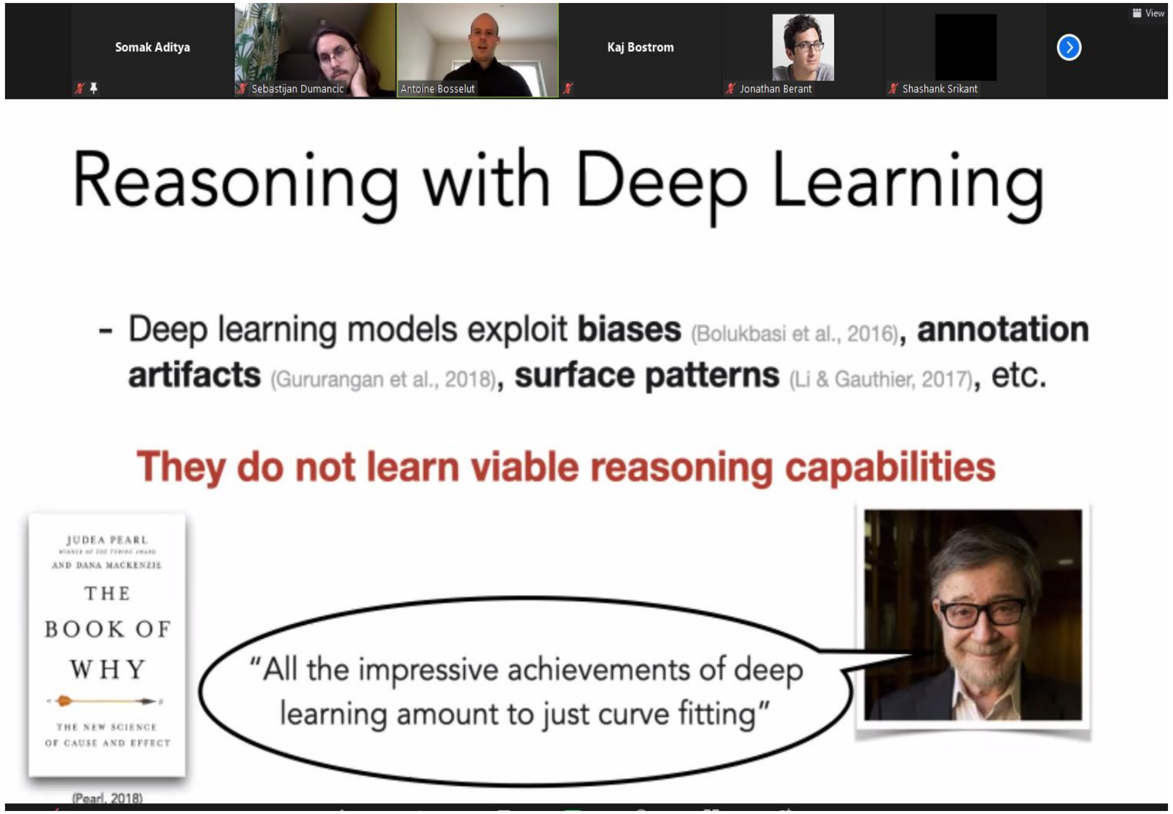


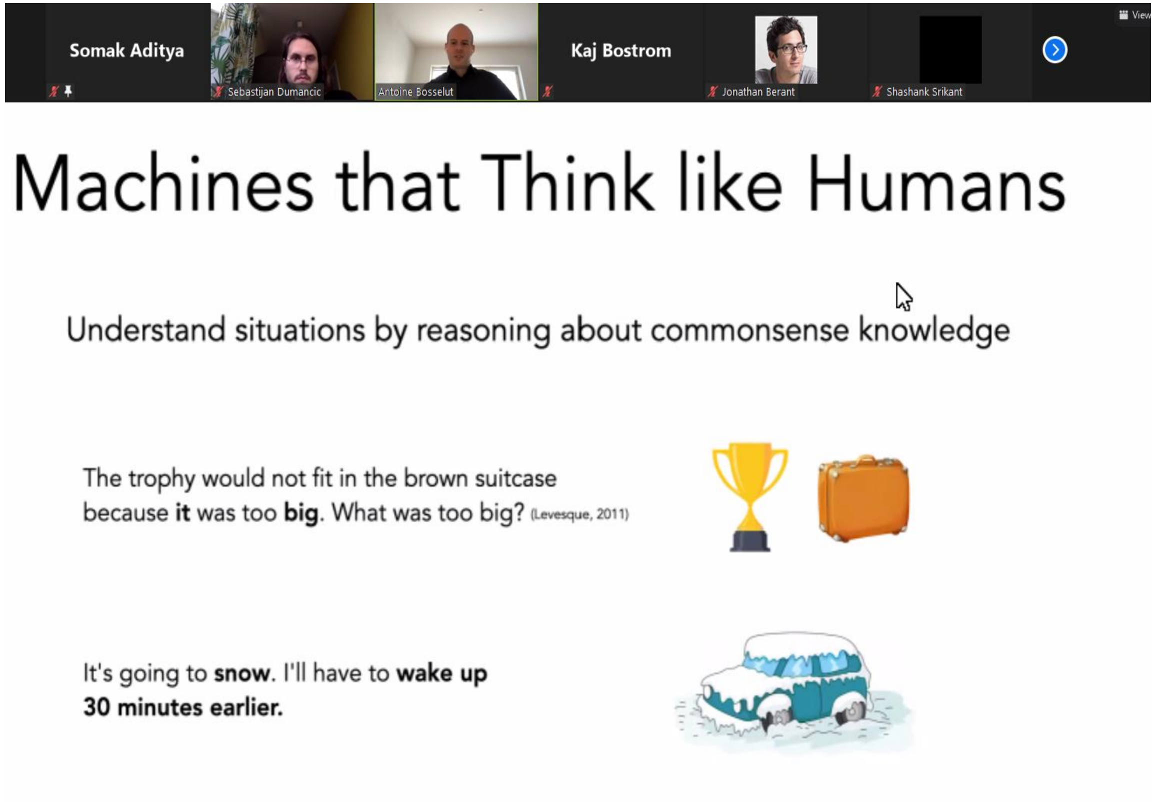


Knowledge-enhanced Text Generation: The Curious Case of Figurative Language and Argumentation

*Awaiting slides from speaker

Session 1 – Antoine Bosselut





Symbolic Scaffolds for Neural Commonsense Representation and Reasoning

*Awaiting slides from speaker

Thanks to Session 1 Speakers

Keynote Speakers:

- Jonathan Berant – *compositional model, controlled data generation*
- Smaranda Muresan – *knowledge-aware models, evaluation metrics, data generation*
- Antoine Bosselut – *representation and reasoning (with structure and un-structured)*

Invited and Contributed Papers:

- Meriem Beloucif – *BERT and the probing tasks*
- Shashank Srikant – *multiple demand system and language center*
- Kaj Bostrom – *alluding to a generative proof tree*

The open-floor QA!

A Hybrid QA-cum-panel.

- Host: Dr. Monojit Choudhury, Principal Researcher, Microsoft Research
- Questionees: Keynote and Invited Speakers
- Questioners: Audience
- Theme:

  1)The perceived disconnect between the philosophy and techniques for commonsense reasoning, and

  2) the need for more robust evaluation paradigms

Session 2

Schedule for the second Half-day session on August 22 (6:00 am – 9:30 am UTC).

6:00 – 6:30	**Welcome & Session 1 Summary**
6:30 – 8:30	**Invited and Contributed Talks** (30 min) ProLinguist: Program Synthesis for Linguistics and NLP paper (30 min) Reasoning using DeepProbLog paper (20 mins) Perception, Inference, and Memory: The Trinity of Machine Learning paper (20 min) A Generative-Symbolic Model for Logical Reasoning in NLU paper (20 min) Multi-hop Reasoning Analysis Based on The Bayesian Probability paper
8:30 – 8:50	**Break**
8:50 – 9:15	**Open-Floor Q&A**
9:15 – 9:30	**Workshop Closing Statement**

Thanks to Session 2 Speakers

Invited Papers:

- Partho Sarathi – *program synthesis to learnt phonetic rules*
- Robin Manhaeve – *DeepProblog and its applications in CLUTTR*

Contributed Papers:

- Adam Lindstrom - *memory to be separate, alluding to continual learning*
- Jidong Tan – *Exciting application of generative neuro-symbolic network*
- Yitian Li – *Multi-hop Reasoning, Using intermediate hops to evaluate BERT*

Topics covered by Speakers

Transformers

- A Generative-Symbolic Model for Logical Reasoning in NLU
- Exploring Multi-hop Reasoning Process in NLU from the View of Bayesian Probability
- How can BERT Understand High-level Semantics?
- Keynote 2

Generative

- A Generative-Symbolic Model for Logical Reasoning in NLU
- Keynote 3

Probabilistic Logic

- Reasoning using DeepProbLog

Program Synthesis

- ProLinguist: Program Synthesis for Linguistics and NLP
- Flexible Operations for Natural Language Deduction
- Keynote 1

Cognitive science/NeuroScience

- Can Cognitive Neuroscience inform Neuro-Symbolic Inference Models?
- Perception, Memory, and Inference: The Trinity of Machine Learning