ACL 2024 Sharing

Zhu Liu

2024.09.26

ACL 2024

- 会议时间：8月11日到16日（17,18 还有2天workshop）
- 会议地点：泰国曼谷的Centara Grand and Bangkok Convention Centre at CentralWorld


centralwOrld

曼谷整体印象

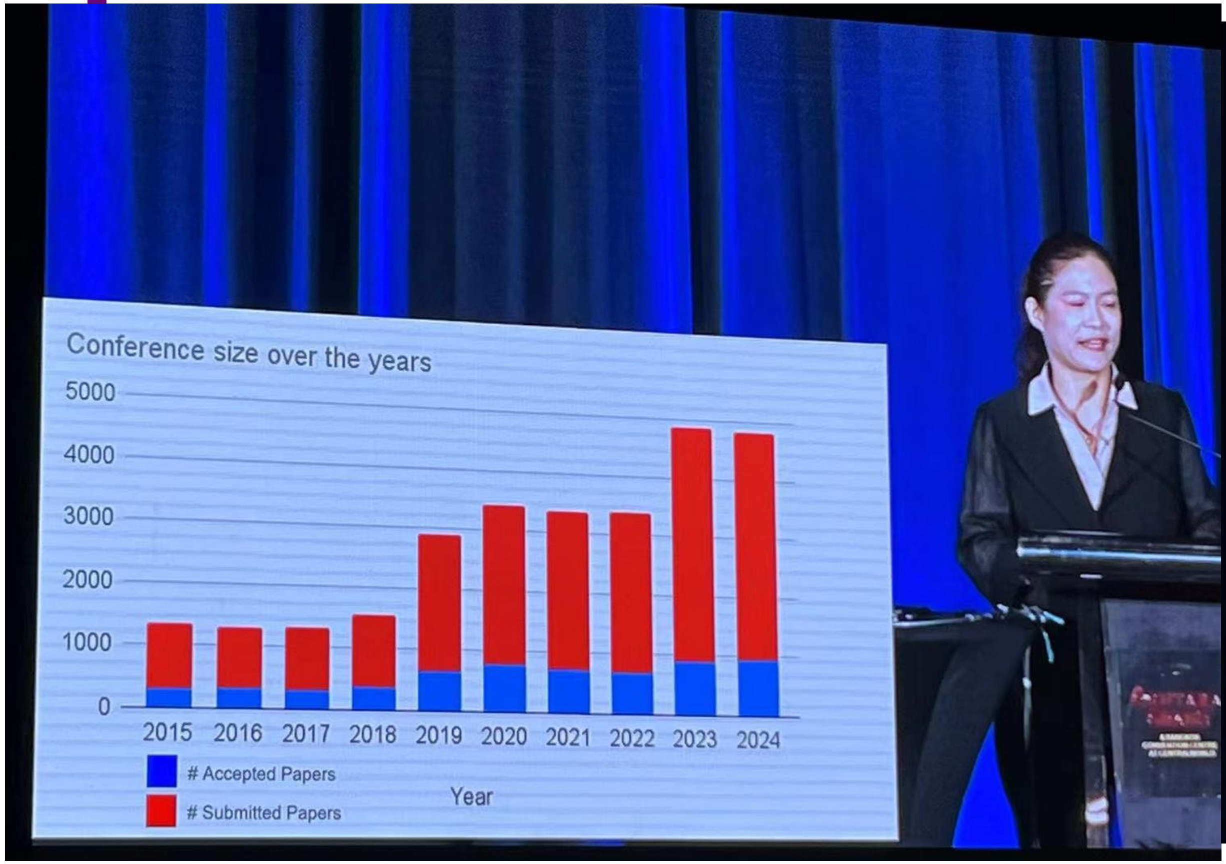
- 传统和现代的交融
  - 传统：很多佛教元素、保留王室、神秘的建筑
  - 现代：高楼大厦、便捷的购物体验...
- 天气湿热
- 饮食偏辛辣
- 一些可能吸引办会的点：国际化程度高、交通便捷、签证方便、英语普及度较高、很安全

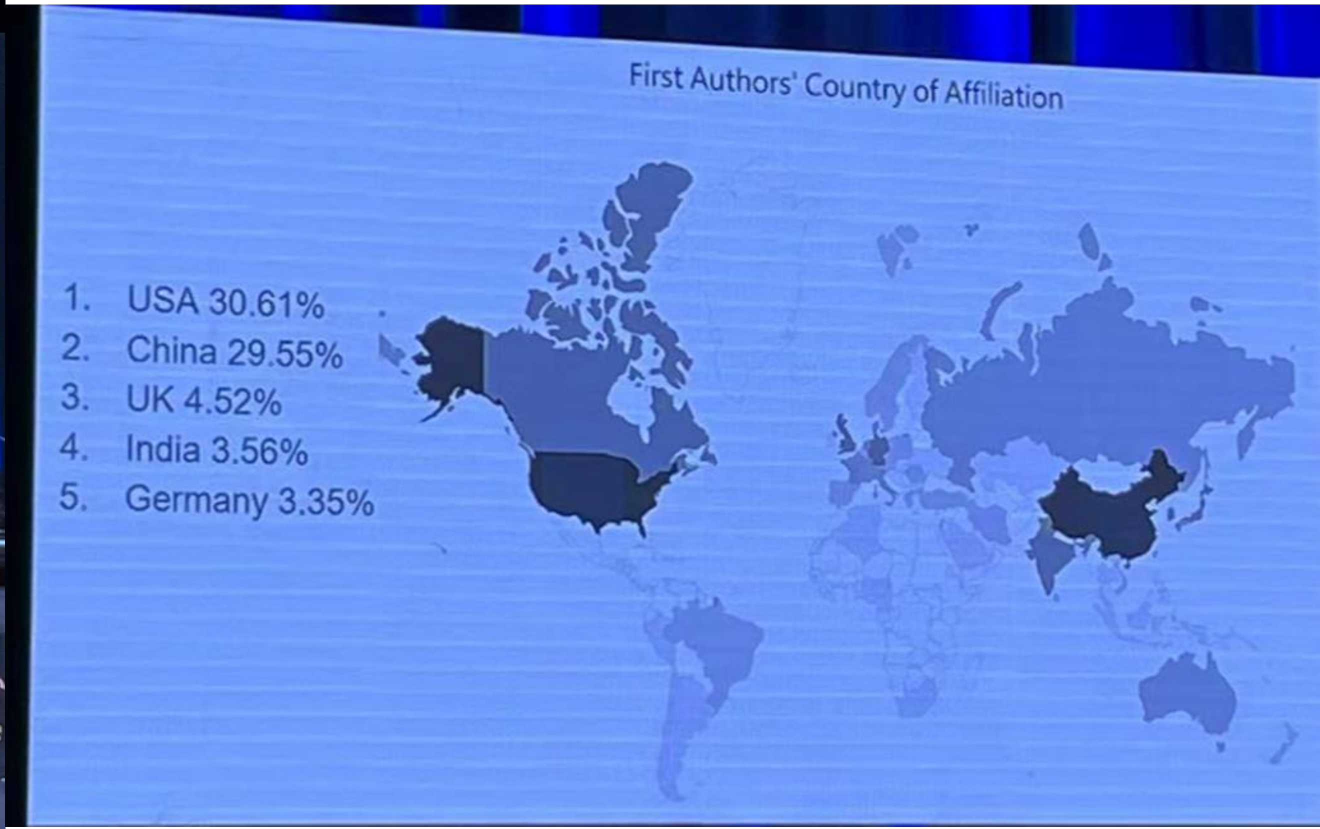
议程

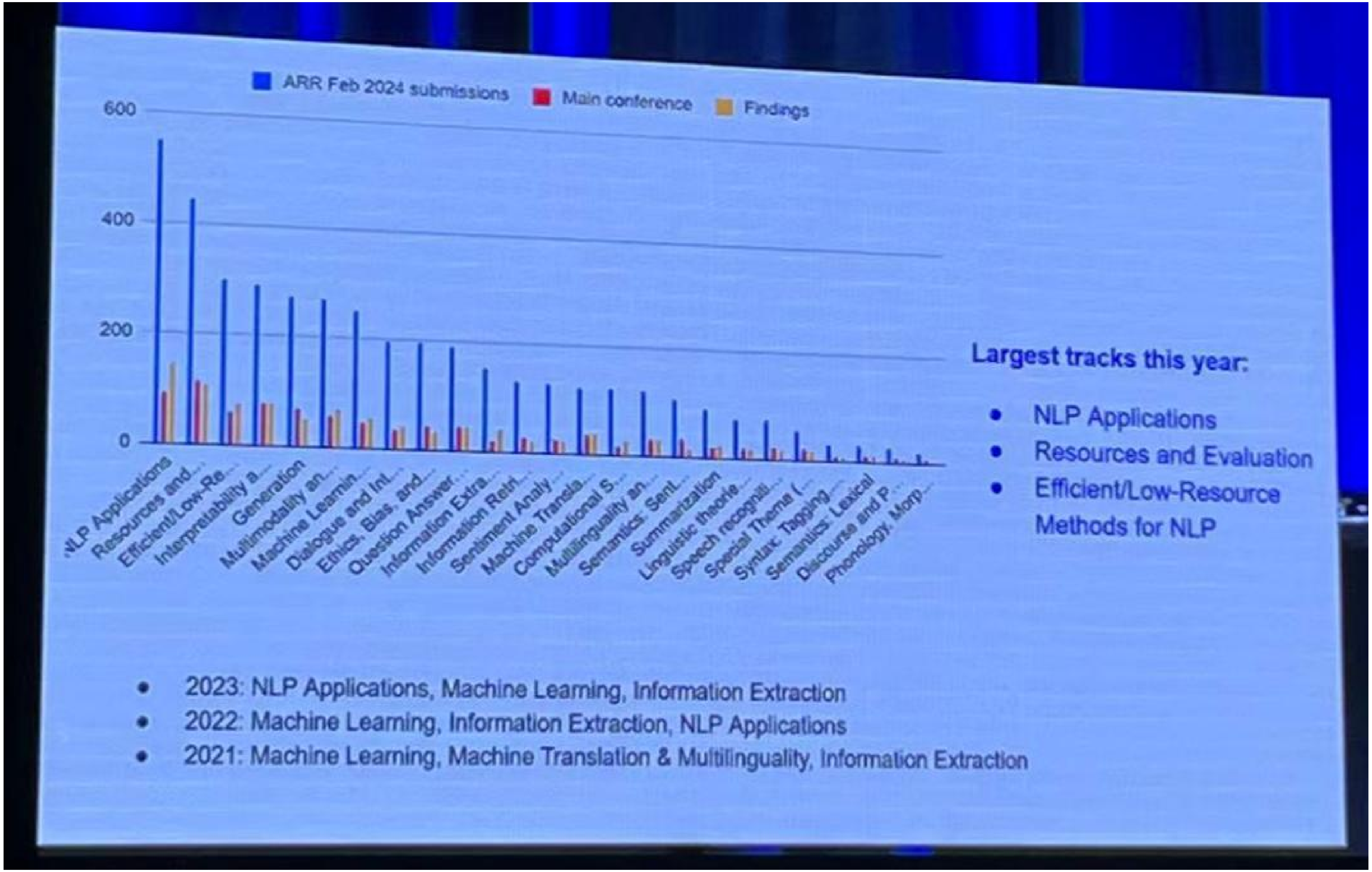
日期	时间	活动内容
8月10日	14:00-19:00	注册
8月11日	07:30-20:30	注册、教程、欢迎招待会
8月12日	07:30-17:30	注册、主题演讲、口头报告/海报/演示
8月13日	08:30-16:30	注册、主题演讲、圆桌讨论、社交晚宴
8月14日	08:30-16:30	注册、主题演讲、奖项与闭幕
8月15-16日	08:00-16:30	注册、研讨会

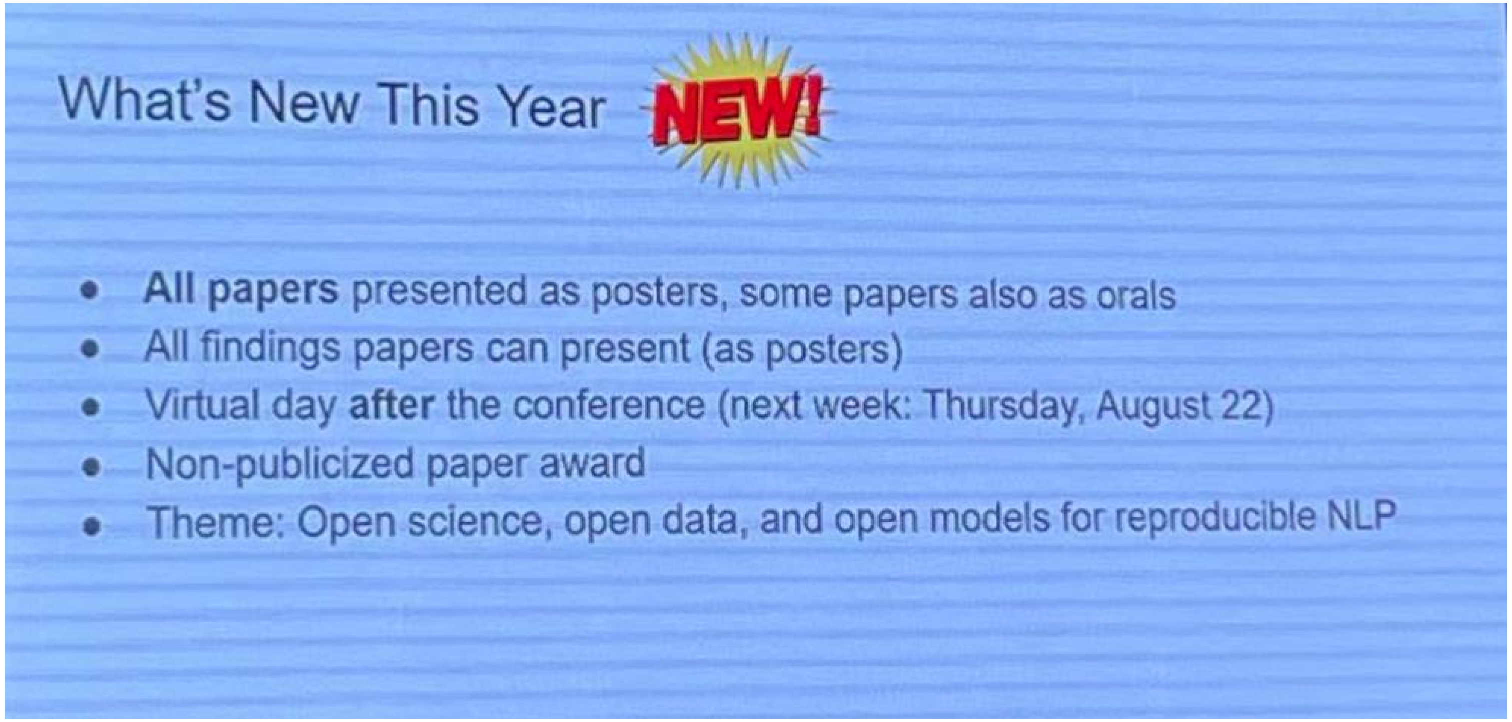
ACL 2024 介绍

- 主会
  - 约5000篇投稿，72个SAC，716个AC，4208个审稿人
  - 940篇main（21.3%），975篇findings（22.1%），6 JCL，31 TACL
  - 3个主旨演讲，1个圆桌讨论
- 其他
  - 18个workshop，6个教程，38个demos和60篇SRW的论文


Conference size over the years
5000
4000
3000
2000
1000
0
2015 2016 2017 2018 2019 2020 2021 2022 2023 2024
Accepted Papers
Submitted Papers
Year


First Authors' Country of Affiliation
1. USA 30.61%
2. China 29.55%
3. UK 4.52%
4. India 3.56%
5. Germany 3.35%


ARR Feb 2024 submissions
Main conference
Findings
600
400
200
0
NLP Applications
Resources and...
Efficient/Low-Re...
Interpretability a...
Generation
Multimodality an...
Machine Learnin...
Dialogue and Int...
Ethics, Bias, and...
Question Answer...
Information Extra...
Information Retri...
Sentiment Analy...
Machine Transla...
Computational S...
Multilinguality an...
Semantics: Sent...
Summarization
Linguistic theorie...
Speech recogniti...
Special Theme (...
Syntax: Tagging...
Semantics: Lexical
Discourse and P...
Phonology, Morp...
Largest tracks this year:
NLP Applications
Resources and Evaluation
Efficient/Low-Resource Methods for NLP
2023: NLP Applications, Machine Learning, Information Extraction
2022: Machine Learning, Information Extraction, NLP Applications
2021: Machine Learning, Machine Translation & Multilinguality, Information Extraction


What's New This Year
NEW!
All papers presented as posters, some papers also as orals
All findings papers can present (as posters)
Virtual day after the conference (next week: Thursday, August 22)
Non-publicized paper award
Theme: Open science, open data, and open models for reproducible NLP

主旨演讲

- Does In-Context-Learning Offer the Best Tradeoff in Accuracy, Robustness, and Efficiency for Model Adaptation?
- Can LLMs Reason and Plan?
- Are LLMs Narrowing Our Horizon? Let's Embrace Variation in NLP!

主旨演讲

- Does In-Context-Learning Offer the Best Tradeoff in Accuracy, Robustness, and Efficiency for Model Adaptation?
- 主讲人：Sunita Sarawagi
  - Professor, IIT Bombay
  - sequence models for text and time-series, domain adaptation, effective human intervention in learning, graphical models and structured learning

Background

- The model adaptation problem: from one domain (training set) to another (test set)
- Challenges:
  - Accuracy (overfit vs. no change)
  - Robustness
  - Efficiency
- Related topics: domain adaptation/transfer learning/few-shot learning
- Methods: Fine-tuning and its variants; Mixture of experts; Task-vectors; Matching-based methods

Evaluation

(small T)	Fine-tuning	MoE	Task vectors	Matching
Accuracy	✓✓	✓✓✓	✓	✓
Robustness	✓	✓✓	✓✓✓	✓✓
Efficiency				
Adapt	✓	✓	Online ✓✓✓	Online ✓✓✓
Test	✓✓✓	✓✓✓	✓✓✓	✓
Others		✓ Difficult to train		

LLM-era

- In-context learning (ICL): Adaptation is just a forward pass
- Why does it work?
  - H1: Transformers implement gradient descent algorithm over IC examples
  - H2: IC examples recognize tasks in pre-training, e.g. via **task vectors**
  - H3: Self-attention implements **matching-based** adaptation via induction heads

Re-evaluation

(small T)	Fine-tuning	ICL
Accuracy	✓✓	✓✓
Robustness	✓	✓
Efficiency Adapt	✓	Online ✓✓✓
Test	✓✓✓	✓✓
Others		✓✓✓ Ease of use in multi-tenancy models

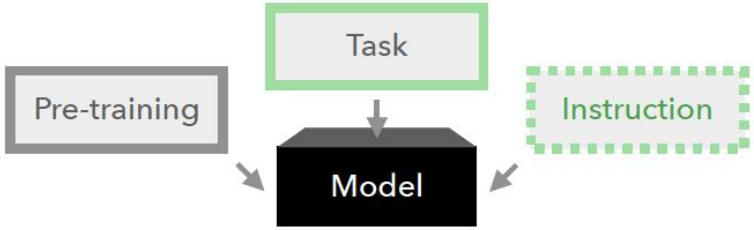
Improvement:
- Structuring of prompts
- Pre-training strategies
- Fine-tuning
- Example retrieval
- Model architectures

Can LLMs Really Reason and Plan?

- Speaker: Subbarao Kambhampati
- Arizona State University
- 关注的问题：LLMs在推理过程中会出现幻觉等问题
- 作者对LLM常用的微调和提示语进行了很多测试
- 结论：LLMs的训练和使用方式并没有显示出它们能够进行原则性推理的迹象。

Are LLMs Narrowing Our Horizon? Let’s Embrace Variation in NLP!

- Speaker: Barbara Plank LMU Munich & IT University of Copenhagen
- LLMs & Trust: A Short Look at AI history
- Regrain trust from four aspects:



D1 Knowledge about Model Input

D2 Knowledge about Model Behaviour

Trust arises from **knowledge of origin** as well as from **knowledge of functional capacity.**

Trustworthiness - Working Definition by David G. Hays, 1979

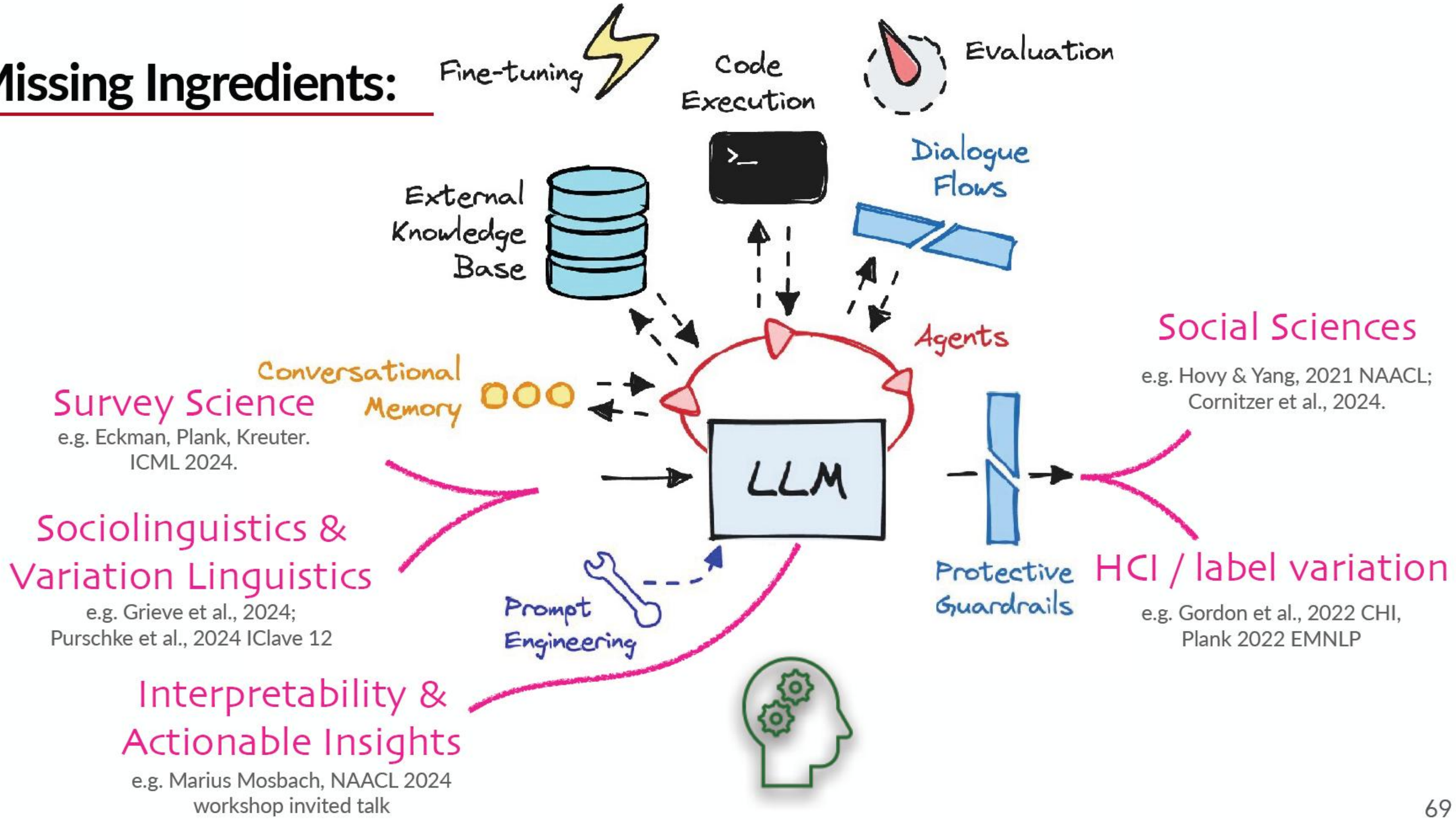
D3 Knowledge of Evaluation Protocols

D4 Knowledge of Data Origin

We need to embrace variation holistically

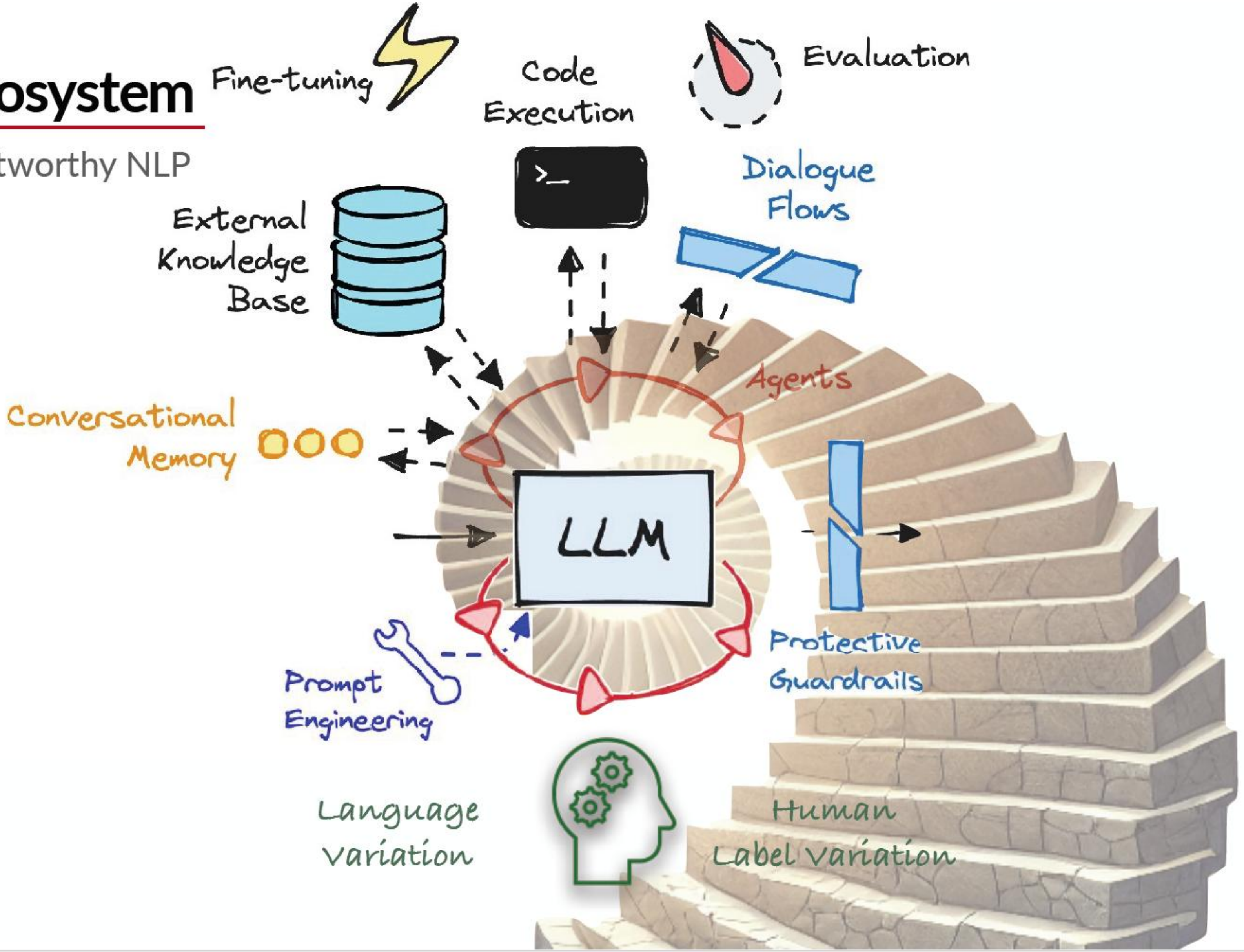
- Inputs: linguistic variation, low-resource languages & dialects
- Outputs: human label variation as signal (not error) [Uncertainty]
- Research: Language as the Bridge

Missing Ingredients:



Trust LLM Ecosystem

Human-facing, Trustworthy NLP



Tutorials

- **Tutorial 1– Computational Linguistics for Brain Encoding and Decoding: Principles, Practices and Beyond**

  Room : Lotus 1-4 (Level 22)

  Jingyuan Sun, and Shaonan Wang, and Zijiao Chen, and Jixing Li, and Marie-Francine Moens

- **Tutorial 2 - Automatic and Human-AI Interactive Text Generation (with a focus on Text Simplification and Revision)**

  Room : Lotus 5-7 (Level 22)

  Yao Dou, and Philippe Laban, and Claire Gardent, and Wei Xu

- **Tutorial 3 - Vulnerabilities of Large Language Models to Adversarial Attacks**

  Room : World Ballroom B (Level 23)

  Yu Fu, Erfan Shayegan, and Md. Mamun Al Abdullah, and Pedram Zaree, and Nael Abu-Ghazaleh, and Yue Dong

Tutorials

14:00 - 17:30 Tutorial 4 - 6

- **Tutorial 4 – Computational Expressivity of Neural Language Models**

  Room: Lotus 1-4 (Level 22)

  Alexandra Butoi and Ryan Cotterell and Anej Svete

- **Tutorial 5– Watermarking for Large Language Model**

  Room: Lotus 5-7 (Level 22)

  Xuandong Zhao, and Yu-Xiang Wang, and Lei Li

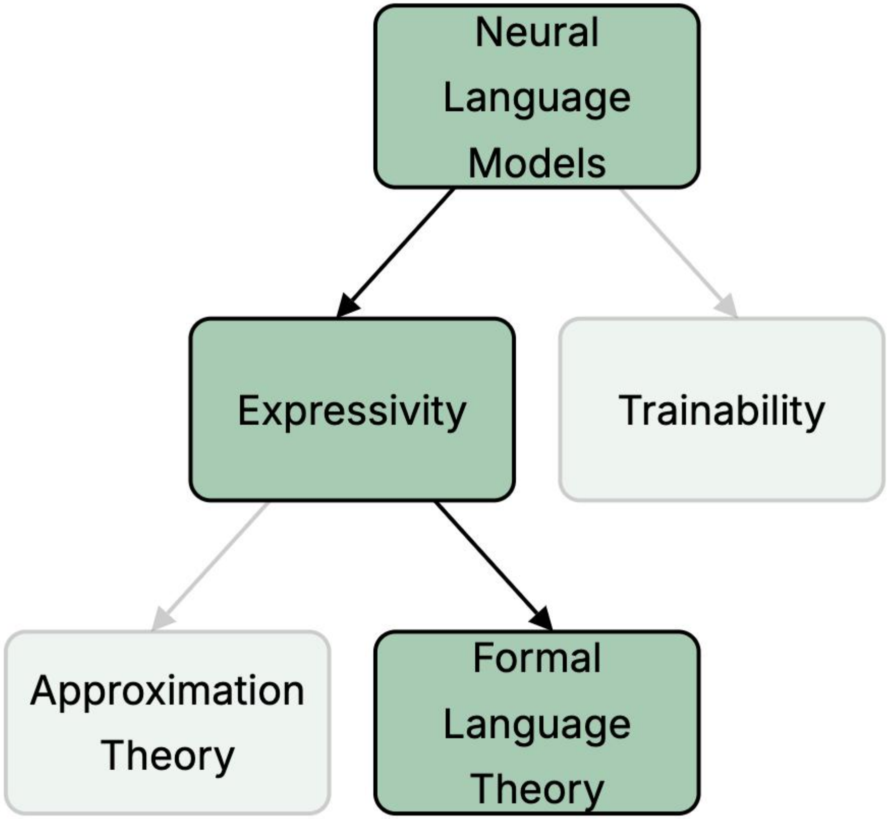
- **Tutorial 6; Presentation Matters: How to Communicate Science in the NLP Venues and in the Wild?**

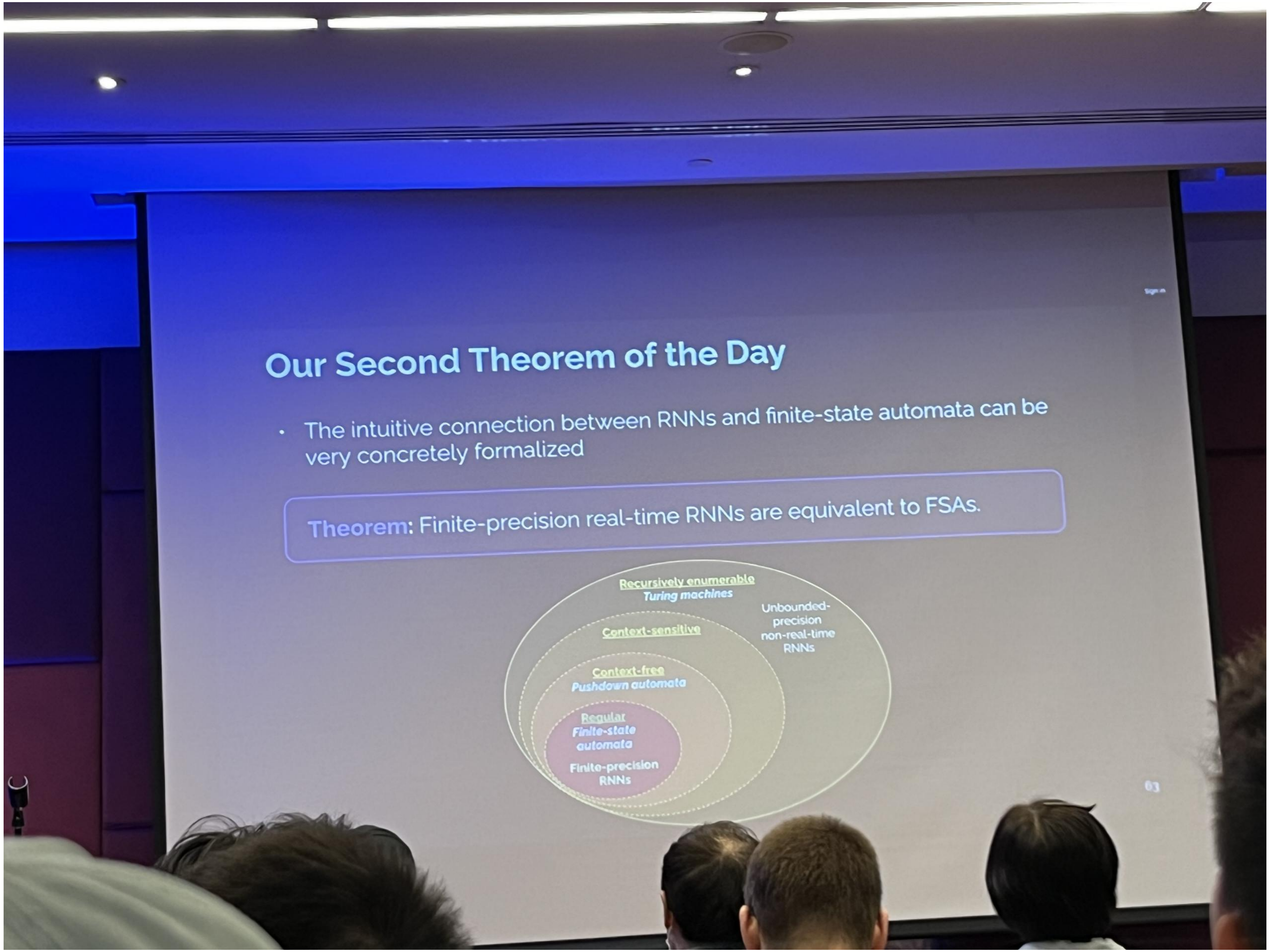
  Room: World Ballroom B (Level 23)

  Sarvnaz Karimi, and Cecile Paris, and Gholamreza Haffari

T4: Representational Capacity of Neural Language Models

- A group of ETH Zürich led by Ryan Cotterell
- formal language theory to understand the representational capcacity of LLMs
- RNN vs. Finite-State Automata
- LLMs vs. finite-state automata and Turing machines.




Our Second Theorem of the Day
The intuitive connection between RNNs and finite-state automata can be very concretely formalized
Theorem: Finite-precision real-time RNNs are equivalent to FSAs.
Recursively enumerable
Turing machines
Unbounded-precision non-real-time RNNs
Context-sensitive
Context-free
Pushdown automata
Regular
Finite-state automata
Finite-precision RNNs

Semantics主题 (oral)

- - NounAtlas: Filling the Gap in Nominal Semantic Role Labeling

  Roberto Navigli, Marco Lo Pinto, Pasquale Silvestri, Dennis Rotondi, Simone Ciciliano, Alessandro Scirè
- - UG-schematic Annotation for Event Nominals: A Case Study in Mandarin Chinese

  Wenxi Li | Yutong Zhang | Guy Emerson | Weiwei Sun

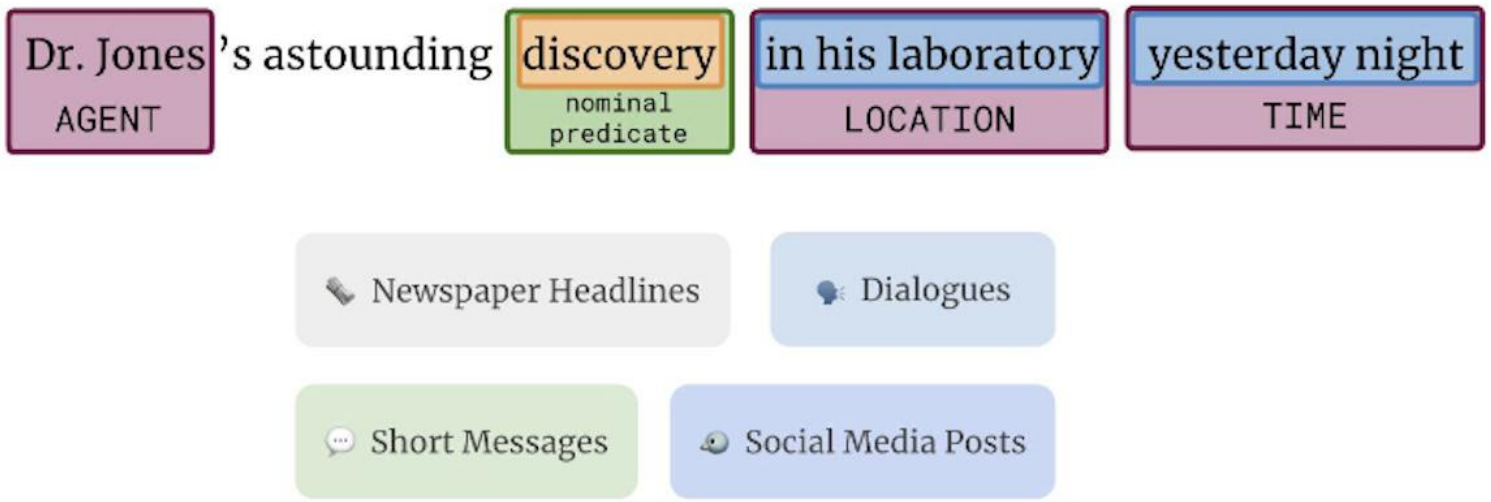
  Computational Linguistics, Volume 50, Issue 2 - June 2023
- Distributional Inclusion Hypothesis and Quantifications: Probing for Hypernymy in Functional Distributional Semantics

  Chun Hei Lo, Wai Lam, Hong Cheng, Guy Emerson

NounAtlas: Filling the Gap in Nominal Semantic Role Labeling

Nominal SRL is under studied

Existing SRL has been primarily focused on **verbal predicates.**
But **nominal predicates are frequent in real-world settings!**



Inventory

NounAtlas

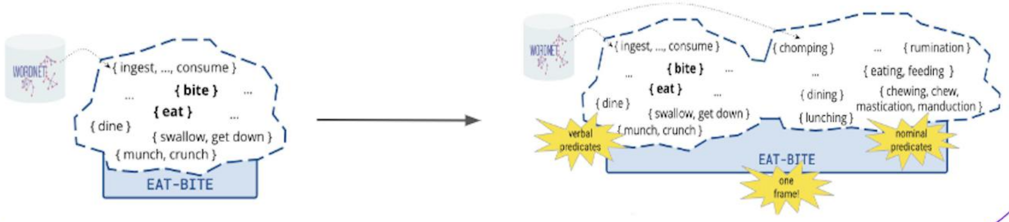
Filling the gap in nominal SRL with NounAtlas!

Creating a huge inventory for Nominal SRL

Expanding VerbAtlas frames

We start from the existing **VerbAtlas frames**, which groups together semantically-related **verbal predicates**.

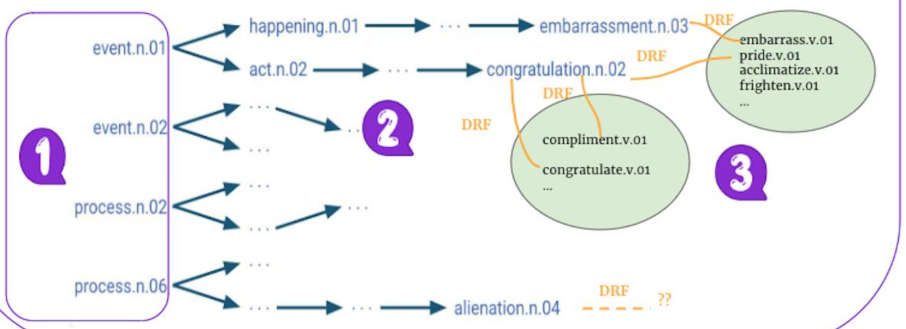
We expand the VerbAtlas frames by adding **WordNet nominal synsets** that convey similar semantics.



1 We manually selected the most prominent **nominal synsets from WordNet**

2 We collect all their **nominal descendants** (i.e. via direct/indirect WordNet **hyponym** relations)

3 We **link** nominal synsets to their verbal counterparts through WordNet's **Derivationally Related Forms (DRF)** edges

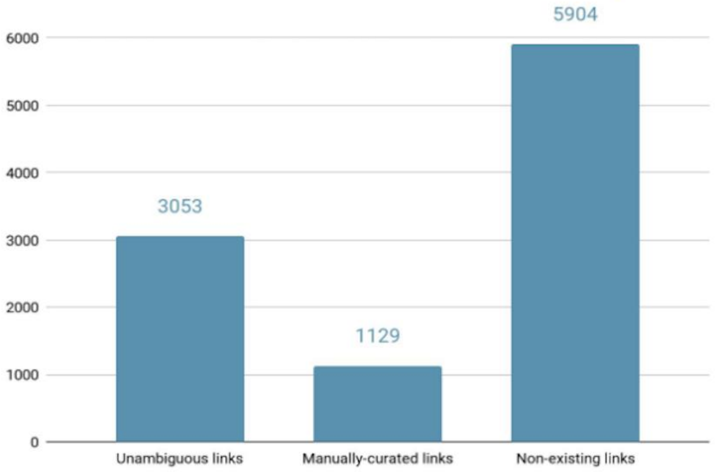


Evaluation

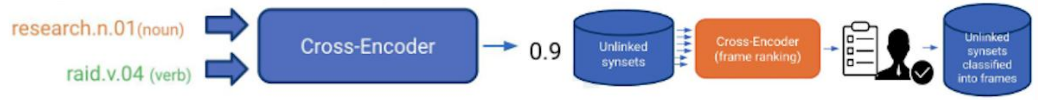
Wordnet-based synset-to-frame mapping: Evaluation

1. **Unambiguous links**, evaluated on 100 items: **82%** in **agreement** with **manual annotations**, 18% made up of equally-valid selections

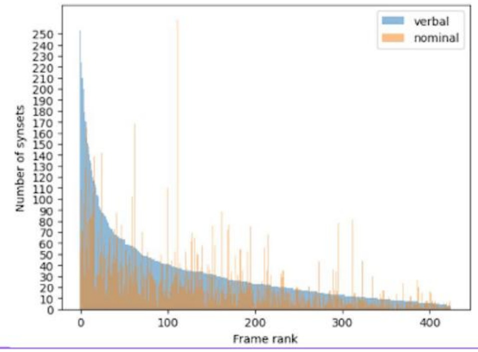
2. **Ambiguous links**, evaluated on 100 items: tie broken through **manual curation, Cohen's Kappa = 0.57 (moderate)**



3. Missing links: We use a Cross-Encoder trained on the other types of link to score predicates and, aggregated on ≤10 predicates, frames. We consider the top-5 frames and manually link.



- As a result of manual annotation, **99.6% of Unlinked synsets were added to a frame**
- **88.2% of correct frames** contained in our **model's top-5 ranking**, 49.3% in the top-1



Corpus + model

Creating a nominal SRL corpus

Predicate nominalization: Generation

Idea → start from **SemCor** (Miller et al. 1993):
Verbal predicates in SemCor annotated with WordNet synsets **are directly link**ed to VerbAtlas frames!

- We retain all SemCor sentences featuring verbal predicates
- We **generate** the corresponding **nominalized version** of each sentence using **Gemini-Pro**

Change the sentence by nominalizing the verb "married" indicated by **. Use exactly one of these deverbal nouns: "marriage ceremony", "marriage", "wedding". Indicate the chosen deverbal noun with **:
The couple was **married** Aug. 2, 1913.

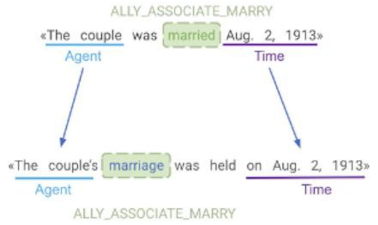
The couple's **marriage** was held on Aug. 2, 1913.

- We **validate** the generation by keeping the sentence if:

1. The deverbal noun is **identified**
2. The deverbal noun is a **noun**
3. Its **lemma is among the candidates** provided

Verbal-to-nominal role propagation

We annotate the verbal sentences with **InVeRo-XL** (Conia et al. 2021)



We validate ~500 random sentences to generate a **gold test set for Nominal SRL:**

unchanged sentences: **92.11%**
unchanged frames: **95.56%**
unchanged role spans: **77.11%**
unchanged role labels: **95.27%**

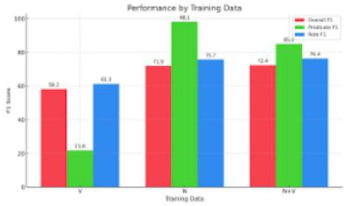
Robust approach

A unified model for SRL

Setup
Backbone model: RoBERTa-based model from Conia and Navigli (2020)
Training data:
- Nominal SRL: Our dataset
- Verbal SRL: OntoNotes 5.0 (V)
- The combination of the two datasets **(N+V)**

Test data:
- Our ~500 manually-curated nominal sentences (nouns)
- The OntoNotes test set (verbs) converted to VerbAtlas annotations

Gold-standard nominal test set



OntoNotes verbal test set

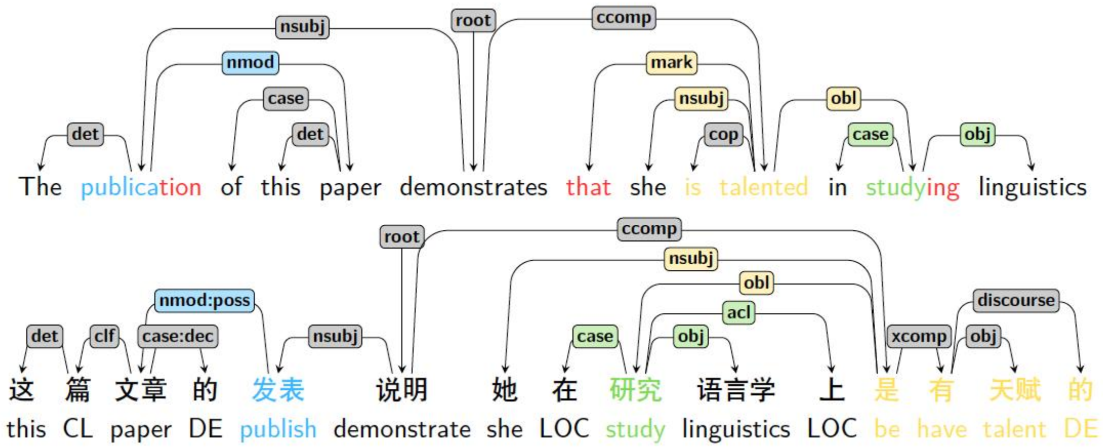
Our model, **jointly-trained on nominal and verbal data (N+V)**, achieves competitive performance on nominal SRL!

Also robust on the OntoNotes test set (verb frames).

Thanks to NounAtlas, we devised a unified approach for nominal and verbal SRL!

Background

Multilingual Heterogeneity of Event Nominals



- **Identification**: Languages differ in their use of functional morphemes for nominalizations and the types of morphemes they use.
- **Nominal SRL**: Languages differ in how they realize participants of event nominals through syntactic constructions

Research Question

Multilingual Heterogeneity of Event Nominals

	Mandarin	English	German
Subject with nominative case	–	?	+
Accusative case on object	–	?	+
Projection of outer Aspect	–	+	+
Modal or auxiliary verb	+	+	+
Complementizer	–	+	+
Verb suffix	–	+	+
Genitive/PP-subject	–	+	+
Genitive/PP-object	–	+	+
Gender	–	?	+
Quantity	–	+	+
Determiner	–	+	+
Noun suffix	–	+	+

Table: Verb- and noun-related features of event nominals

Research Question

How can we accommodate multilingually heterogeneous phenomena in a unified way, or more theoretically, uncover universals of the world's languages beneath their surface-level variations?

Distributional Inclusion Hypothesis and Quantifications: Probing for Hypernymy in Functional Distributional Semantics

DIH and Quantifications

DIH. r_2 is a hypernym of r_1 iff r_1's characteristic contexts $\subseteq$ r_2's.

Quantifications. A corpus with only universally quantified statements results in the reverse of DIH (rDIH).

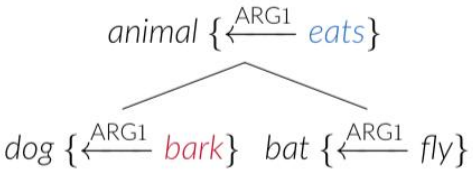


Figure 1. A taxonomic hierarchy of nouns. Next to each noun are the contexts applicable to it and its descendants.

Corpus 1 (DIH)	Corpus 2 (rDIH)
some dog barks	*every dog barks*
some animal barks	*every dog eats*
some bat flies	*every bat flies*
some animal flies	*every bat eats*
some animal eats	*every animal eats*

Table 1. Corpora generated from the hierarchy in Fig. 1.

FDS

Entity Vectors. $z \in \mathbb{R}^d$

Truth-Conditional Semantic Functions.

$$\begin{aligned} t^{(dog)}(z) &= P\left(dog(z) = \top \mid z\right) \\ &= \text{sigmoid}\left(v^{(dog)\top} z + b^{(dog)}\right) \end{aligned}$$

Representing Hypernymy.

$$\forall z \text{ s.t. } \|z\|_2 \leq 1\colon t^{(dog)}(z) < t^{(animal)}(z)$$

which is true iff $s(dog, animal) > 0$, where

$$s(r_h, r_H) = b^{(r_H)} - b^{(r_h)} - \left\| v^{(r_H)} - v^{(r_h)} \right\|_2$$

Model Training. By Lo et al. (2023), given a DMRS graph:

$$some \xrightarrow{\text{RSTR}} dog \xleftarrow{\text{ARG1}} bark$$

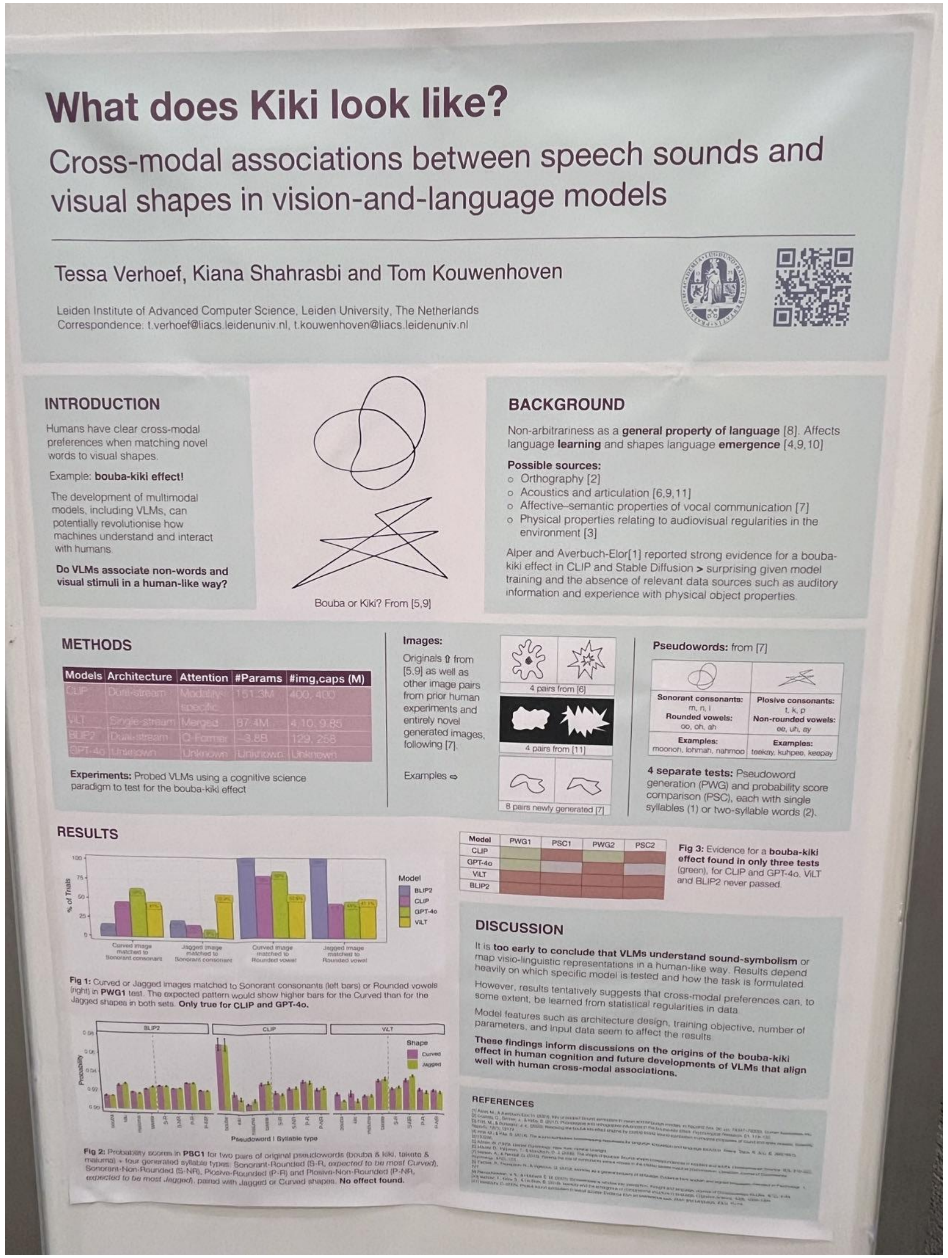
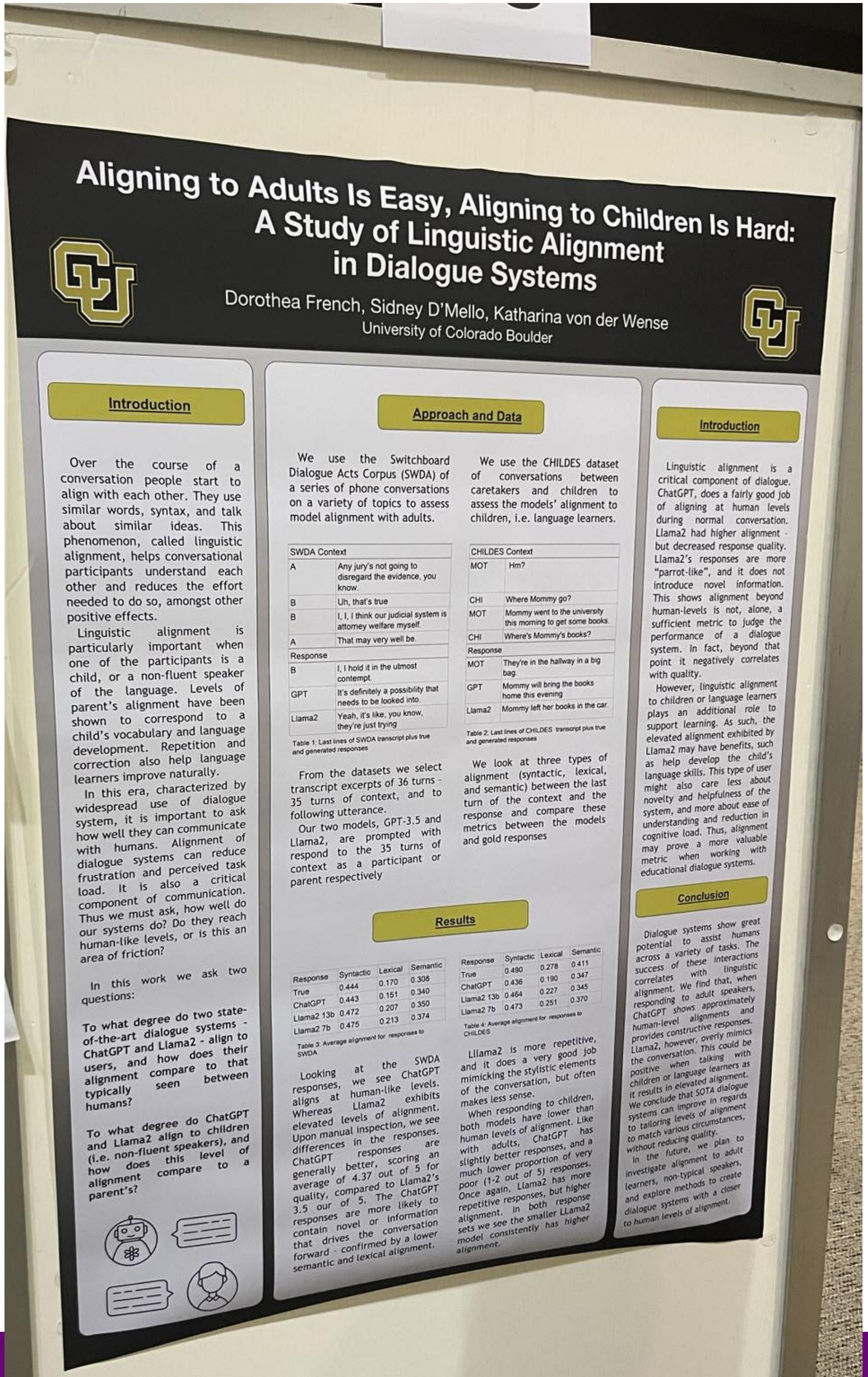
Variational Inference: $q_\phi(z \mid \xleftarrow{\text{ARG1}} bark)$

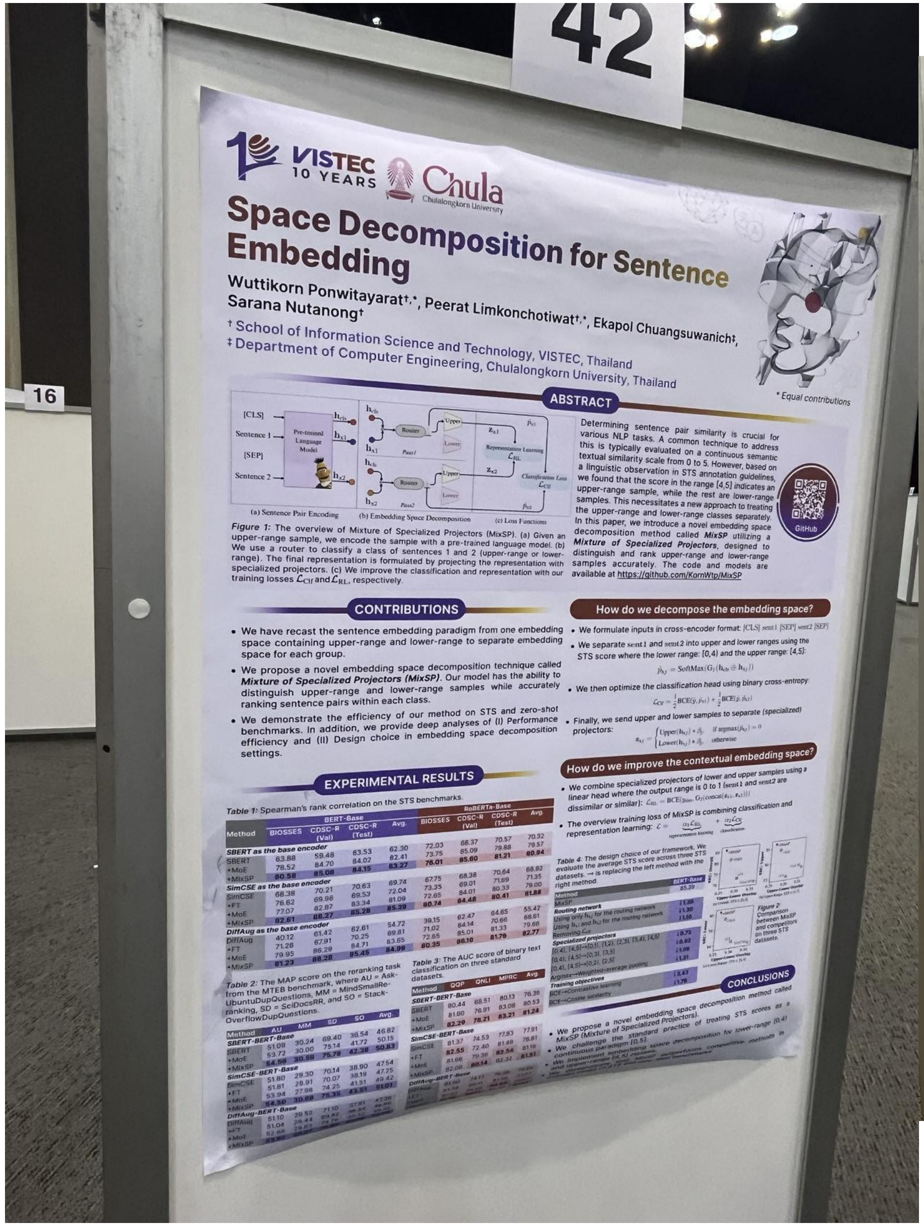
Reconstruction: $\max \ln \mathbb{E}_{q_\phi(z \mid \xleftarrow{\text{ARG1}} bark)} \left[t^{(dog)}(z)\right] + \ldots$

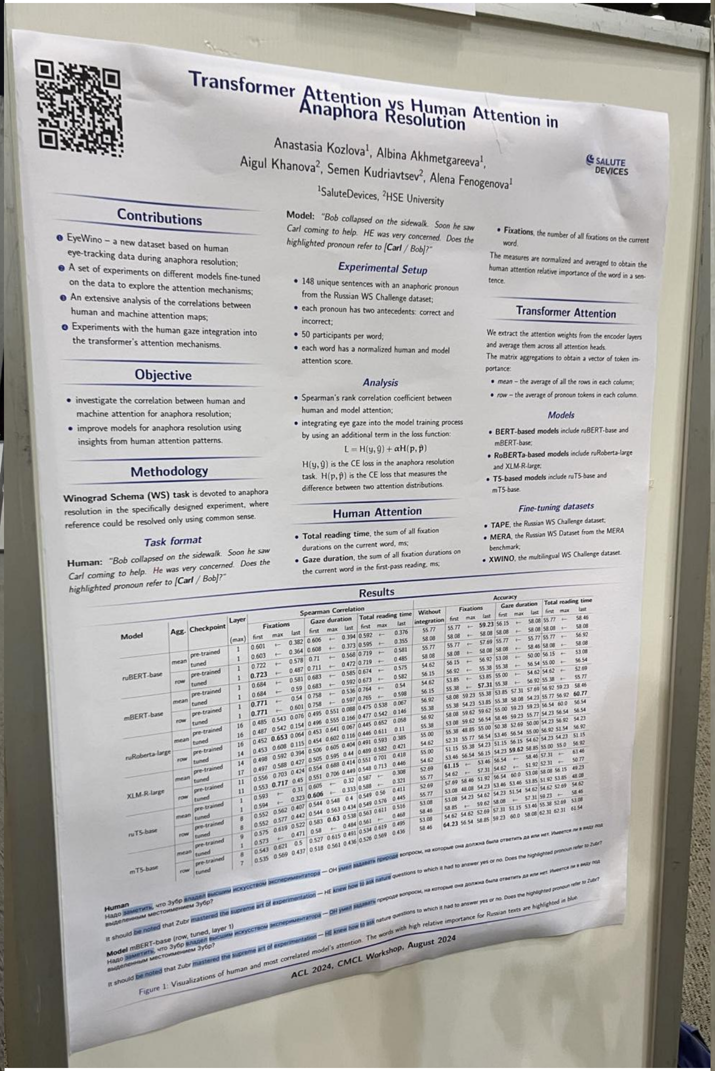
New objective (FDS$_\forall$). Optimizes over regions of the entity s
for handling universal quantifications

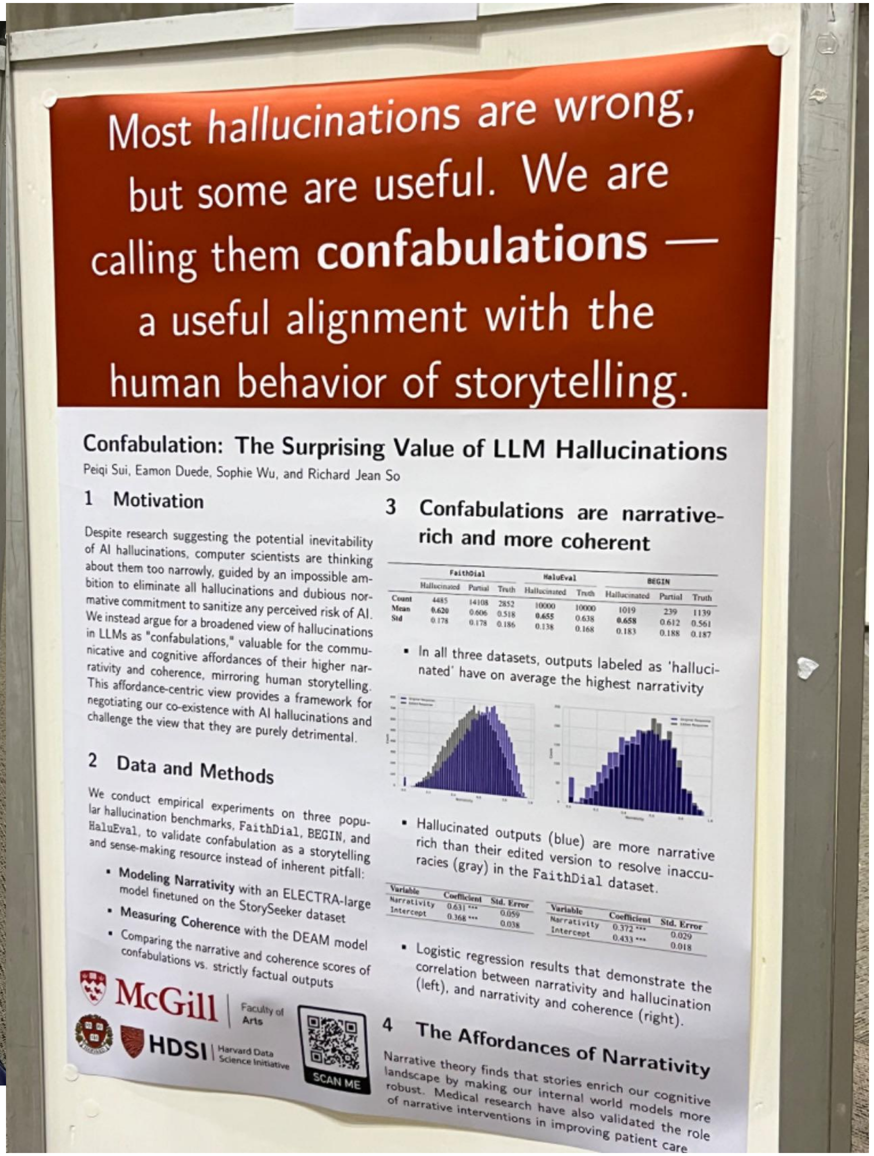
Posters


Fantastic Semantics and Where to Find Them:
Investigating Which Layers of Generative
LLMs Reflect Lexical Semantics
Introduction
Results
Aims
Method
References
ACL 2024

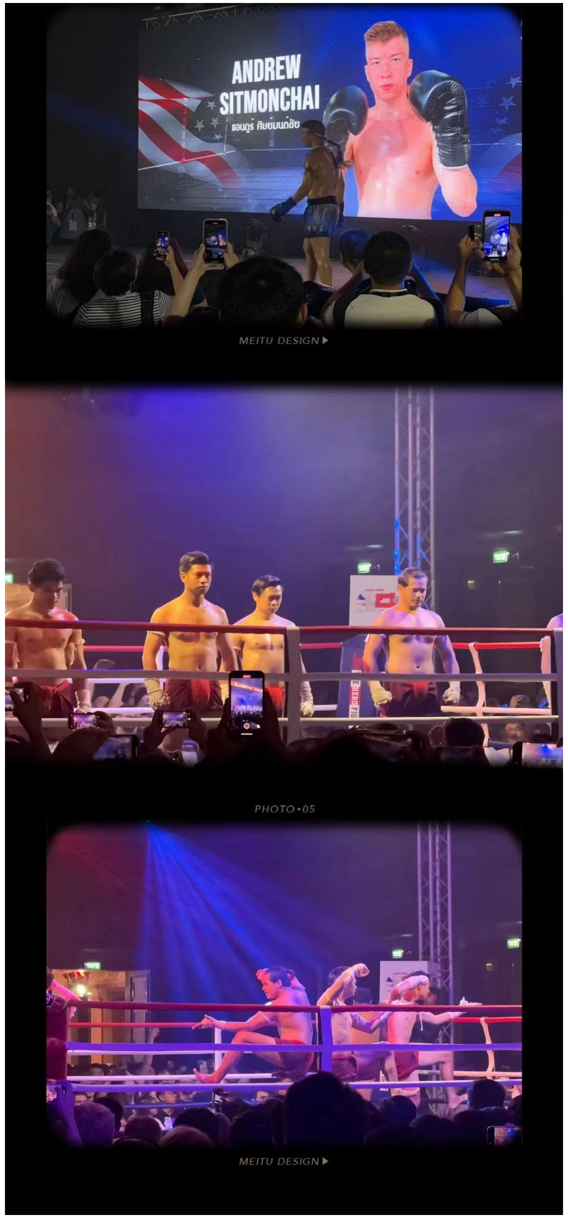

What does Kiki look like?
Cross-modal associations between speech sounds and visual shapes in vision-and-language models
Tessa Verhoef, Kiana Shahrasbi and Tom Kouwenhoven
Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands
Correspondence: t.verhoef@liacs.leidenuniv.nl, t.kouwenhoven@liacs.leidenuniv.nl
INTRODUCTION
Humans have clear cross-modal preferences when matching novel words to visual shapes.
Example: bouba-kiki effect!
The development of multimodal models, including VLMs, can potentially revolutionise how machines understand and interact with humans.
Do VLMs associate non-words and visual stimuli in a human-like way?
Bouba or Kiki? From [5,9]
BACKGROUND
Non-arbitrariness as a general property of language [8]. Affects language learning and shapes language emergence [4,9,10]
Possible sources:
Orthography [2]
Acoustics and articulation [6,9,11]
Affective-semantic properties of vocal communication [7]
Physical properties relating to audiovisual regularities in the environment [3]
METHODS
Images:
Pseudowords: from [7]
Experiments: Probed VLMs using a cognitive science paradigm to test for the bouba-kiki effect
RESULTS
DISCUSSION
REFERENCES

Aligning to Adults Is Easy, Aligning to Children Is Hard:
A Study of Linguistic Alignment in Dialogue Systems
Dorothea French, Sidney D'Mello, Katharina von der Wense
University of Colorado Boulder
Introduction
Approach and Data
Results
Conclusion


42
16
VISTEC 10 YEARS
Chula Chulalongkorn University
Space Decomposition for Sentence Embedding
Wuttikorn Ponwitayarat†,*, Peerat Limkonchotiwat†,*, Ekapol Chuangsuwanich‡, Sarana Nutanong†
† School of Information Science and Technology, VISTEC, Thailand
‡ Department of Computer Engineering, Chulalongkorn University, Thailand
* Equal contributions
ABSTRACT
CONTRIBUTIONS
How do we decompose the embedding space?
How do we improve the contextual embedding space?
EXPERIMENTAL RESULTS
CONCLUSIONS
GitHub


Transformer Attention vs Human Attention in Anaphora Resolution
Anastasia Kozlova[1], Albina Akhmetgareeva[1], Aigul Khanova[2], Semen Kudriavtsev[2], Alena Fenogenova[1]
[1]SaluteDevices, [2]HSE University
SALUTE DEVICES
Contributions
Objective
Methodology
Task format
Experimental Setup
Analysis
Human Attention
Transformer Attention
Models
Fine-tuning datasets
Results
ACL 2024, CMCL Workshop, August 2024


Most hallucinations are wrong, but some are useful. We are calling them confabulations — a useful alignment with the human behavior of storytelling.
Confabulation: The Surprising Value of LLM Hallucinations
Peiqi Sui, Eamon Duede, Sophie Wu, and Richard Jean So
1 Motivation
Despite research suggesting the potential inevitability of AI hallucinations, computer scientists are thinking about them too narrowly, guided by an impossible ambition to eliminate all hallucinations and dubious normative commitment to sanitize any perceived risk of AI. We instead argue for a broadened view of hallucinations in LLMs as "confabulations," valuable for the communicative and cognitive affordances of their higher narrativity and coherence, mirroring human storytelling. This affordance-centric view provides a framework for negotiating our co-existence with AI hallucinations and challenge the view that they are purely detrimental.
2 Data and Methods
We conduct empirical experiments on three popular hallucination benchmarks, FaithDial, BEGIN, and HaluEval, to validate confabulation as a storytelling and sense-making resource instead of an inherent pitfall:
Modeling Narrativity with an ELECTRA-large model finetuned on the StorySeeker dataset
Measuring Coherence with the DEAM model
Comparing the narrative and coherence scores of confabulations vs. strictly factual outputs
3 Confabulations are narrative-rich and more coherent
In all three datasets, outputs labeled as 'hallucinated' have on average the highest narrativity
Hallucinated outputs (blue) are more narrative rich than their edited version to resolve inaccuracies (gray) in the FaithDial dataset.
Logistic regression results that demonstrate the correlation between narrativity and hallucination (left), and narrativity and coherence (right).
4 The Affordances of Narrativity
Narrative theory finds that stories enrich our cognitive landscape by making our internal world models more robust. Medical research have also validated the role of narrative interventions in improving patient care
McGill Faculty of Arts
HDSI Harvard Data Science Initiative
SCAN ME

Social Events



颁奖环节

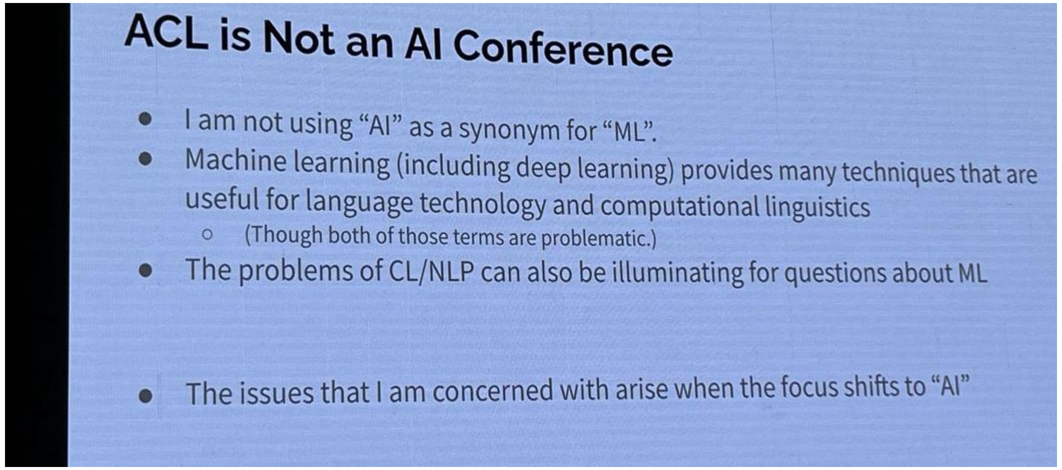
- The 2024 winner of the 1999 Test-of-Time Paper Award is:
- Lillian Lee. 1999. Measures of Distributional Similarity.
- In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL), pages 25–32, College Park, Maryland, USA.

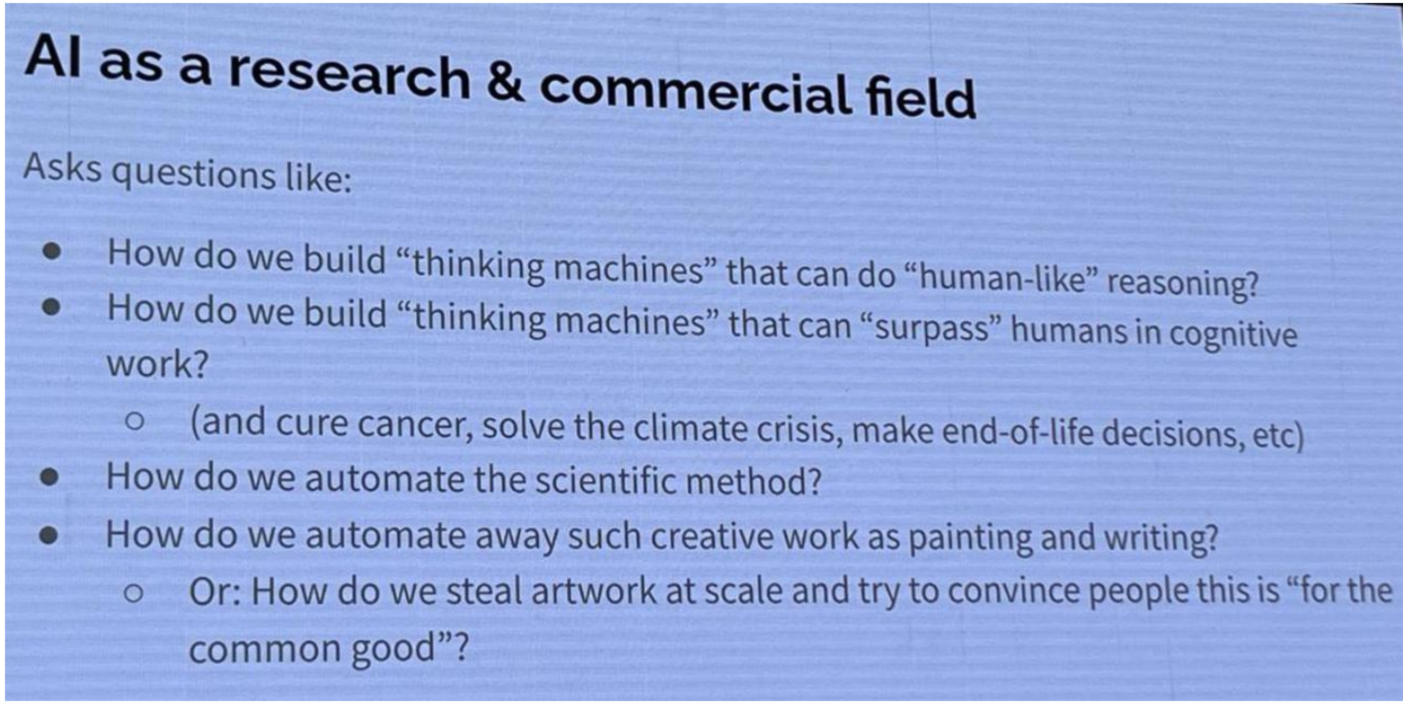
- The 2024 winner of the 2014 Test-of-Time Paper Award is:
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global Vectors for Word Representation.
- In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543, Doha, Qatar.

Best Paper Award

- Mission: Impossible Language Models
- Why are Sensitive Functions Hard for Transformers?
- Deciphering Oracle Bone Language with Diffusion Models
- Causal Estimation of Memorisation Profiles
- Aya Model: An Instruction Finetuned Open-Access Multilingual Language Model
- Semisupervised Neural Proto-Language Reconstruction
- Natural Language Satisfiability: Exploring the Problem Distribution and Evaluating Transformer-based Language Models

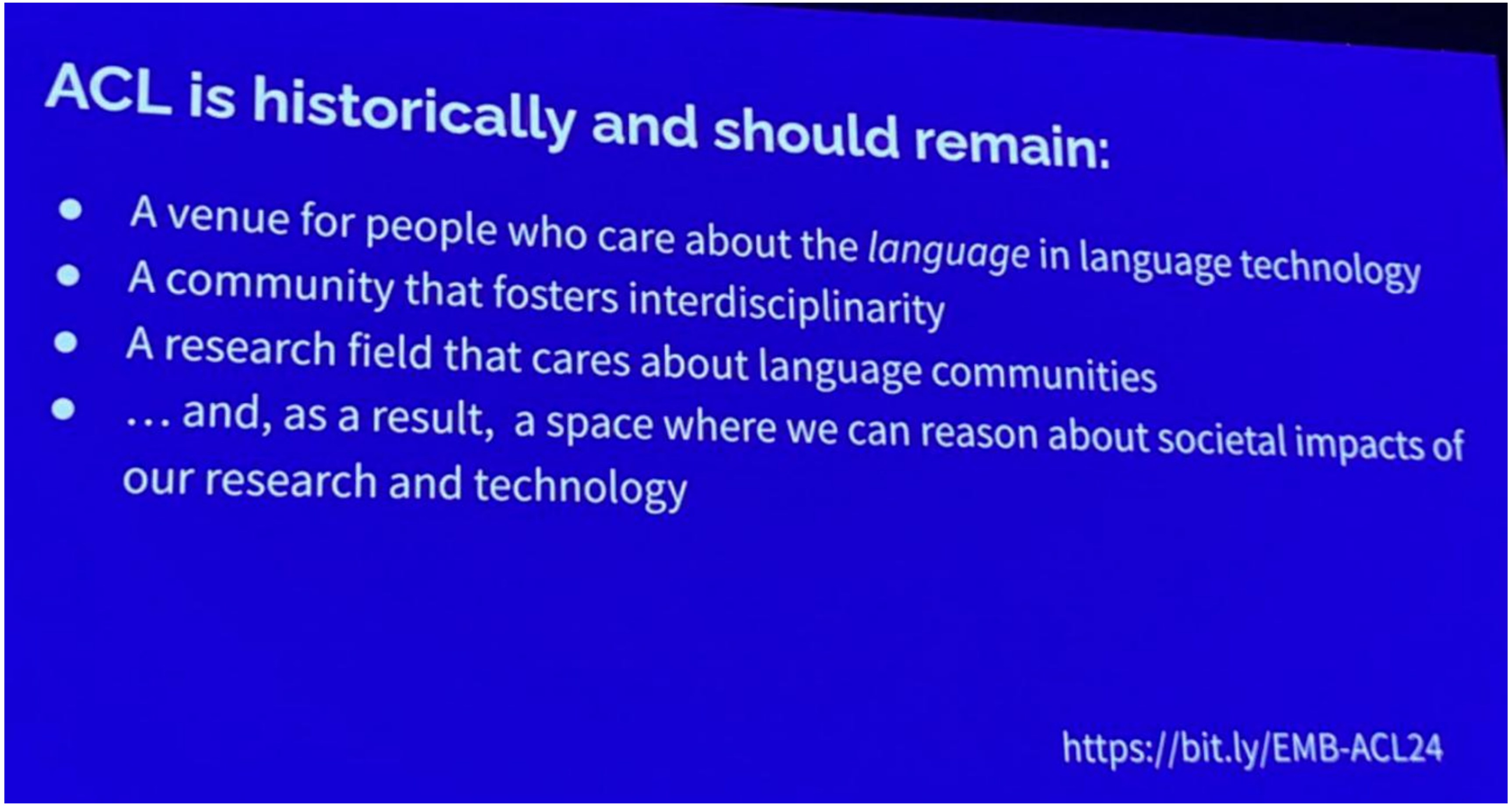
One more thing... from EMB

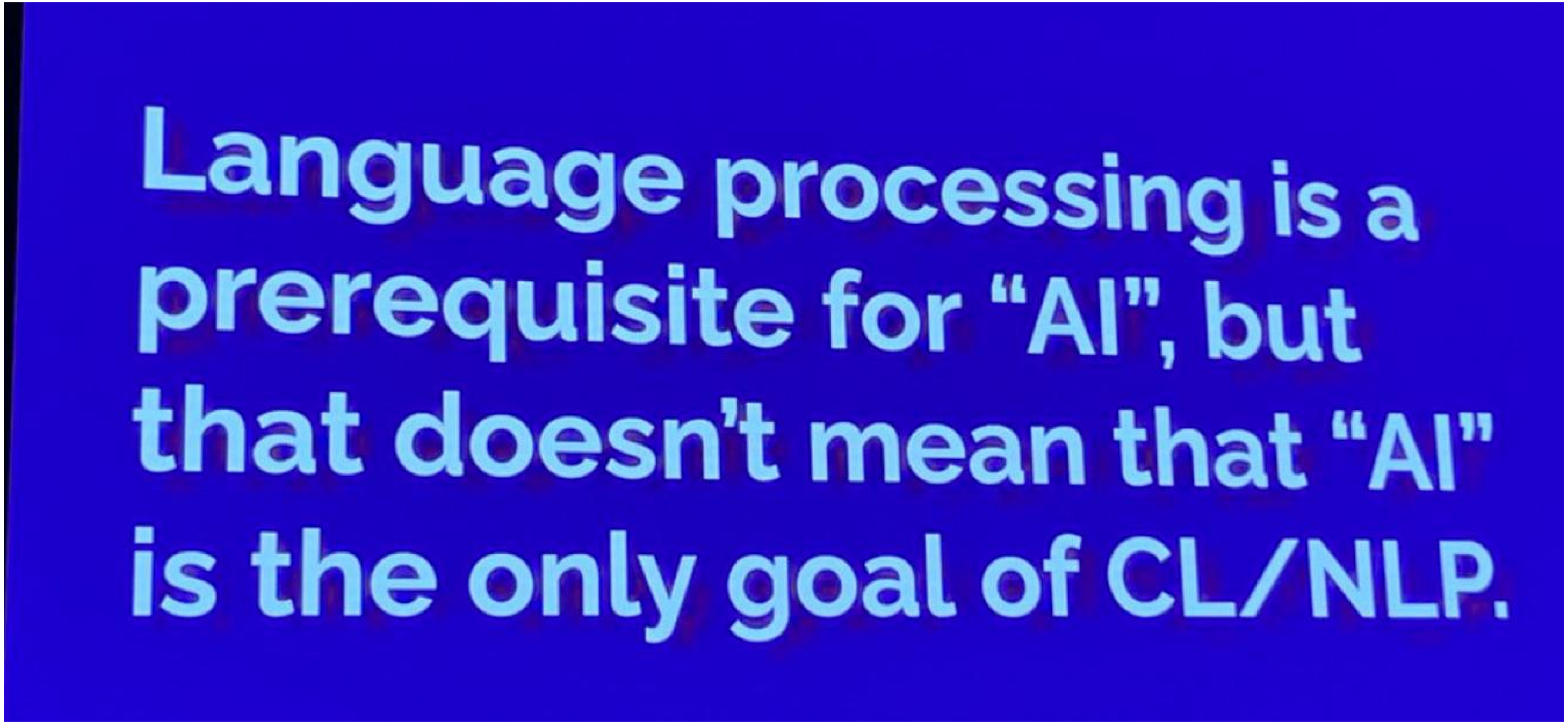




Compling/NLP asks questions such as

- How are languages similar/different?
- How is information represented in languages?
- How can we build technology that assists with: transcription, translation, summarization, information access … in different languages?
- How can we evaluate such technology?
- What kinds of intermediate representations are useful for such technology?
- How well do different ML techniques work for different tasks?
- How do language technologies interact with existing systems of power and oppression?


ACL is historically and should remain:
A venue for people who care about the *language* in language technology
A community that fosters interdisciplinarity
A research field that cares about language communities
… and, as a result, a space where we can reason about societal impacts of our research and technology
https://bit.ly/EMB-ACL24


Language processing is a
prerequisite for "AI", but
that doesn't mean that "AI"
is the only goal of CL/NLP.

Workshops

- Rep4NLP (Probing -> causal inference)
  - - Latent Space Exploration for Safe and Trustworthy AI by Hassan Sajjad
  - The gap between what language models say and what they know
  - Generalisation in LLMs – and beyond
  - Efficiency as an Inductive Bias: Towards Tokenizer-free and Dynamically Sparse Language Models
- Heng Ji: AI Plays Medicinal Chemist
- Anna Rogers: Ai for reasearch workflow

Q & A

THANK YOU

Note