Project Proposals

Jindřich Helcl, Jindřich Libovický, Dominik Macháček, Peter Polák, Michal Novák, Ondřej Bojar, Pavel Šindelář, Ondřej Dušek, Zdeněk Kasner

Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

Project reports

- Maximum 4 pages + references + overview of individual contributions

- Brief method overview (e.g., what is RAG)

- Experimental design

- Results

- Related work (what other did before for the task)

- Conclusions (lessons learned, recommendation for future work)

- 4-minute presentation at the final class

Next Steps

- Form teams of **5 members**

  (4 members absolute minimum, 6 absolute maximum)

- Until Friday, fill in a form with
  - Team name
  - Team members
  - Contact email
  - Three topic preferences

1 – Regional-Specific Question Answering

- Open ended QA with local questions
- Czech, Slovak, Ukrainian + English translations
- ~1000 textual and ~500 visual items per language
- Test RAG approaches with different data sources
  - English vs. local Wikipedia
- Compare with zero-shot (non-RAG) performance

Jindřich Helcl, Jindřich Libovický
{helcl, libovicky}@ufal.mff.cuni.cz

In which village does the comedy Slunce, seno, jahody by Zdeněk Troška take place?

In which year was Ludvík Svoboda elected as the president of the Czechoslovak Socialist Republic?

What is the name of the mascot of the Czech Radio Radio Junior station?

Q: What was the building in the picture used for between 1969 and 1990?

A: The Federal Assembly meetings

2 – Mu-SHROOM: Hallucination Detection

- Span labeling of factual errors
- Based on SemEval shared task 🍄
- Multilingual *(ar, ca, cs, de, en, es, eu, fa, fi, fr, hi, it, sv, zh)*
- Winning systems score ~65% overlap
  - *(54% on Czech, "mark all" baseline 26%)*
- Project goal: Develop competitive system

Jindřich Libovický, Jindřich Helcl
`{libovicky, helcl}@ufal.mff.cuni.cz`

Z jakého důvodu podal Petr Nečas v roce 2013 demisi?

*Petr Nečas, bývalý český premiér, podal demisi **25.** června 2013, poté co byl zveřejněn skandál kolem jeho blízké spolupracovnice, Jan**a Kohoutové**, **která byla jeho bývalou partnerkou**. Skandál se týkal korupce a úplatkářství v **jeho** kanceláři a údajné úplatkářské praktiky, které měly být prováděny za účelem získání politických výhod.*

*Nečas se rozhodl podat demisi, aby zabránil dalšímu úniku informací a **aby se vyhnul možnému úniku z funkce**. Demise byla přijata prezidentem republiky **Václavem Klausem** a následně byla sestavena nová vláda, kterou vedl Jiří Rusnok.*

– Llama 3.1 8B, 2024

Kdo namluvil Rákosníčka?

*Rákosní**čka** je postavou z **knihy "Mlok a Rákosníček" od českého spisovatele Jana Wericha. Tento humoristický román vznikl v roce 1957. Hlavními postavami jsou Mlok a Rákosníček, kteří se setkají v lese a vytvářejí mezi sebou různé dobrodružství. Rákosníček je charakterizován jako nevěřící, skeptický a ironický pták, který se snaží rozbíjet iluzi a přinášet realistický pohled na svět.***
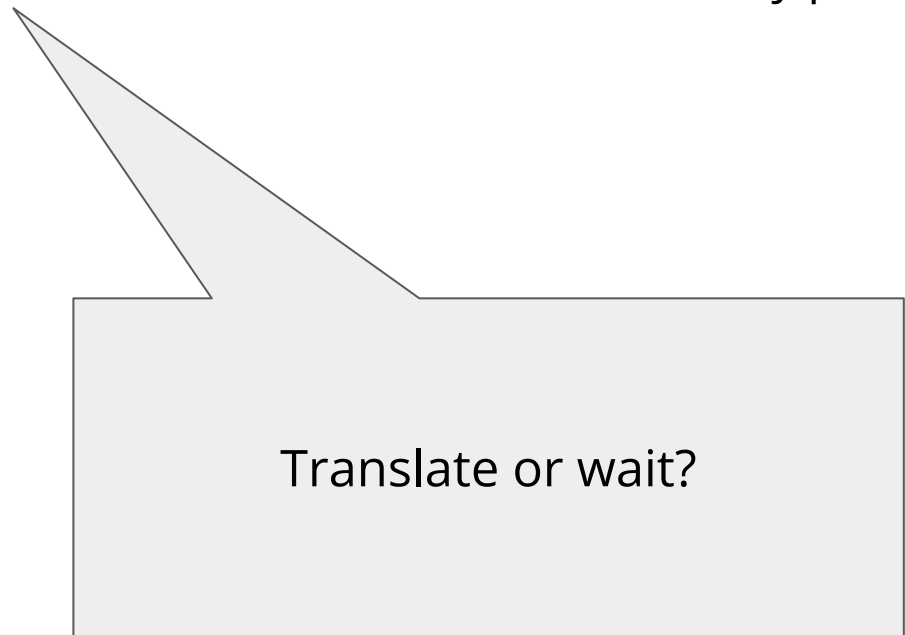
– Mistral 7B, 2024

Mu-SHROOM SemEval-2025 Task 3

Doctor wore a white coat   ...   Doktor měl na sobě bílý plášť

Translate or wait?

Dominik Macháček, Peter Polák
{machacek,polak}@ufal.mff.cuni.cz

Doctor wore a white coat   ...   Doktor**ka** měl**a** na sobě bílý plášť

when she talked to the patients.   ...   když mluvil**a** s pacienty.

Translate or wait?

Let's ask LLM.

Dominik Macháček, Peter Polák
{machacek,polak}@ufal.mff.cuni.cz

3 – Simultaneous Translation

- Proposal 1: **Text-to-text LLM** (for everyone)
  - Speech-to-text by Dominik
  - Given (partial source, partial prev. translation) => translate
  - Prompting / in-context learning / CoT / finetuning / ...
- Proposal 2: **Speech-to-text** (for those who prefer)
  - Model: Whisper (or something else?)
  - PyTorch / training a module / adversarial learning / ...
- Objectives:
  - Non-computational latency (= fast computation not needed) and MT quality (metrics like COMET, ...)
- Languages
  - Czech->English, English->{German, Chinese, Japanese}
- **Bonuses** (extra work)
  - IWSLT shared task in mid April.
  - Integration to live demo, serving real users, join our business project
  - And more. (student faculty grant, publications, theses, ...)

Dominik Macháček, Peter Polák
{machacek,polak}@ufal.mff.cuni.cz

- text analysis
- clustering mentions referring to the same entity

Michal Novák
mnovak@ufal.mff.cuni.cz

Postižených **cestovek** bylo víc, **dvě** z nich udělaly bankrot, vyprávěla **nám** Ludmila Janočková , majitelka **agentury** .

Na **krach** to vypadalo i u **nás** , odešlo **mi** proto několik šikovných spolupracovníků.

já Musela jsem se zříci výhodné prodejní kanceláře v Kotvě a rychle hledat *někoho* způsob, jak **firmu** zachránit.

Především *ona* informovala **své věřitele** (jen **hotelům** *ona* dlužila půl druhého milionu) o nastalé situaci a požádala **je** o poshovění.

Pokud *oni* nechtěli nebo nemohli odklad **platby** povolit, *ona* použila k **úhradám** zálohy vybrané na připravené zájezdy.

Něco málo *ona* zaplatila i z osobních půjček poskytnutých příbuznými a přáteli.

Michal Novák
mnovak@ufal.mff.cuni.cz

Michal Novák
mnovak@ufal.mff.cuni.cz

Prof. Dr. Semih Koray'ın, dergimizde *bizim* ki eski yazılarından *onun* derlenerek kaleme alınan makalesini *onun* dosyamızın *bizim* sunuş yazısı *onun* yaptık *biz* .

Koray, **bilim** etkinliğinin *onun* **tarih** içinde *onun* ki serüvenini *onun* geleceğe de uzanarak analiz ederken *o* dosyamıza *bizim* güzel bir giriş de yapıyor *o* .

Dosyamız *bizim* esas olarak iki **bölümden** oluşuyor *o* .

1592년 (**선조 25년**) **임진왜란이** 일어나자 **세자**(世子)로 책봉되었고, **함경도와** **전라도** 등지에서 **군수품과** 의병을 직접 모집하고 **군량미를** 모으는 데에도 힘썼다.

임진왜란 동안 **세자의** 몸으로 **부왕을** 돕고 **전쟁** 승리에 적지 않은 공을 세웠으나, **부왕의** 인정과 칭찬은 커녕 견제와 냉대를 당해야만 했다.

Michal Novák
mnovak@ufal.mff.cuni.cz

Michal Novák
mnovak@ufal.mff.cuni.cz

4 – Multilingual Coreference Resolution

- **task:** detecting mention spans (even zero ones) and clustering them to entities
- **data:** CorefUD 1.3 (≥22 datasets for ≥16 languages)
- opportunity to participate in the **running shared task** and compare with other contestants

Michal Novák
mnovak@ufal.mff.cuni.cz

http://ufal.cz/corefud/crac25

5 – ELOQUENT Sensemaking

One of "labs" in [ELOQUENT/CLEF shared tasks](https://).

Can your language model prep, sit, or rate an exam for you?

Ondřej Bojar, Pavel Šindelář
bojar@ufal.mff.cuni.cz,
pavlik.sindelar@centrum.cz

LLMs store immense amounts of information and easily answer questions about it. Can they *limit their knowledge* to information provided in given materials?

	Inputs	Output
"**Teacher**" submissions	input materials	quiz of multiple questions
"**Student**" submissions	input materials and quiz questions	answers to the questions
"**Evaluator**" submissions	input materials, a question and an answer	score

5 – ELOQUENT Sensemaking Example

Original speech transcript:
"Impfungen sind wahnsinnig wichtig für die Bekämpfung von Krankheiten. Aber eine immer größere Gruppe von Menschen scheint zu glauben, dass Impfungen unsere Gesundheit gefährden. Im Internet wird viel von negativen Impffolgen wie Allergien, Behinderungen und sogar Todesfällen berichtet."

Ondřej Bojar, Pavel Šindelář
bojar@ufal.mff.cuni.cz,
pavlik.sindelar@centrum.cz

MT into English:
"Vaccinations are incredibly important for fighting disease. But an increasingly large group of people seem to believe that vaccinations put our health at risk. There are many *reports on the Internet of negative vaccination consequences* such as allergies, disabilities and even deaths."

Sample question for this section:
"What *three negative effects* of vaccination are *mentioned on the Internet*?",
Corresponding answer:
"Allergic reactions, disability, death."

5 – ELOQUENT Sensemaking Timeline

25 April 2025: Registration closes. Evaluation starts.

10 May 2025: End of Evaluation Cycle [submission of runs]

30 May 2025: Submission of Participant Papers [CEUR-WS]

30 May–27 June 2025: Review process of participant papers

9 June 2025: Submission of Condensed Lab Overviews [LNCS]

16 June 2025: Notification of Acceptance for Condensed Lab Overviews [LNCS]

23 June 2025: Camera Ready Copy of Condensed Lab Overviews [LNCS] due

27 June 2025: Notification of Acceptance for Participant Papers [CEUR-WS]

7 July 2025: Camera Ready Copy of Participant Papers and Extended Lab Overviews [CEUR-WS] due

21-25 July 2025: CEUR-WS Working Notes Preview for Checking by Authors and Lab Organizers

…or proceed at *your pace* and get "just" **the credit for NPFL140**.

Ondřej Bojar, Pavel Šindelář
bojar@ufal.mff.cuni.cz,
pavlik.sindelar@centrum.cz

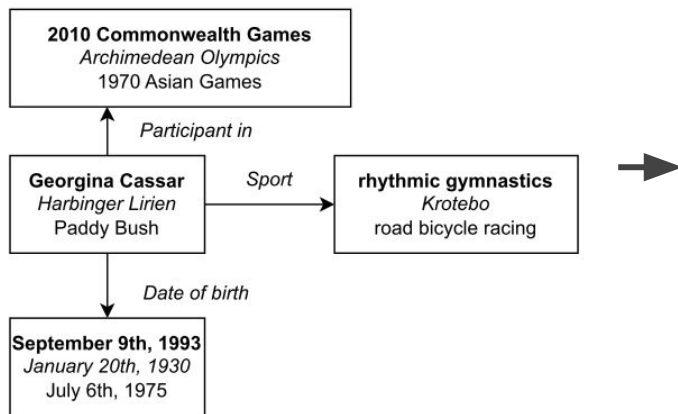# 6 – GEM: Multilingual graph-to-text generation

GEM 2024 Shared Task:

Generating short texts from knowledge graphs.

Triples: subject – predicate – object.
Subgraphs up to 7 triples.

Ondřej Dušek & Zdeněk Kasner
{odusek,kasner}@ufal.mff.cuni.cz

Georgina Cassar, born on 9 September 1993, competed in the 2010 Commonwealth Games in rhythmic gymnastics.

6 – GEM: Multilingual graph-to-text generation

Three types of data:

- ✅ **factual**: original data from the knowledge graphs
- 🔮 **fictional**: entities swapped using rules
- ⛔ **counterfactual**: new entities invented using LLMs

Ondřej Dušek & Zdeněk Kasner
{odusek,kasner}@ufal.mff.cuni.cz

✅ **factual**
Georgina Cassar, born on 9 September 1993, competed in the 2010 Commonwealth Games in rhythmic gymnastics.

🔮 **fictional**
Harbinger Lirien, born on 20 January 1930, competed in Archimedean Olympics in Krotebo.

⛔ **counterfactual**
Paddy Bush, born on 6 July 1975, competed in the 1970 Asian Games in road bicycle racing.

6 – GEM: Multilingual graph-to-text generation

Languages

- **Inputs** in English (concerns mostly relation names).
- Human-written **reference texts** for English and Swahili.
- Additional output languages in the shared task: Arabic, Chinese, German, Hindi, Korean, Russian, Spanish
  - But any language is possible in principle.

Ondřej Dušek & Zdeněk Kasner
{odusek,kasner}@ufal.mff.cuni.cz

Evaluation

- Basic automatic metrics for en+sw
- Human evaluation in the shared task → LLM-based evaluation for our purposes (?)
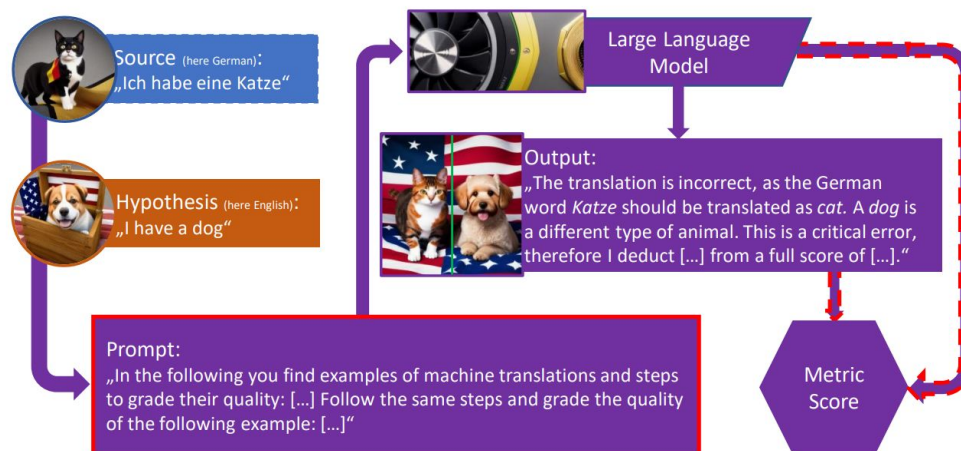
Eval4NLP 2023 shared task (github)

Prompting LLMs for evaluating texts (MT, summarization)

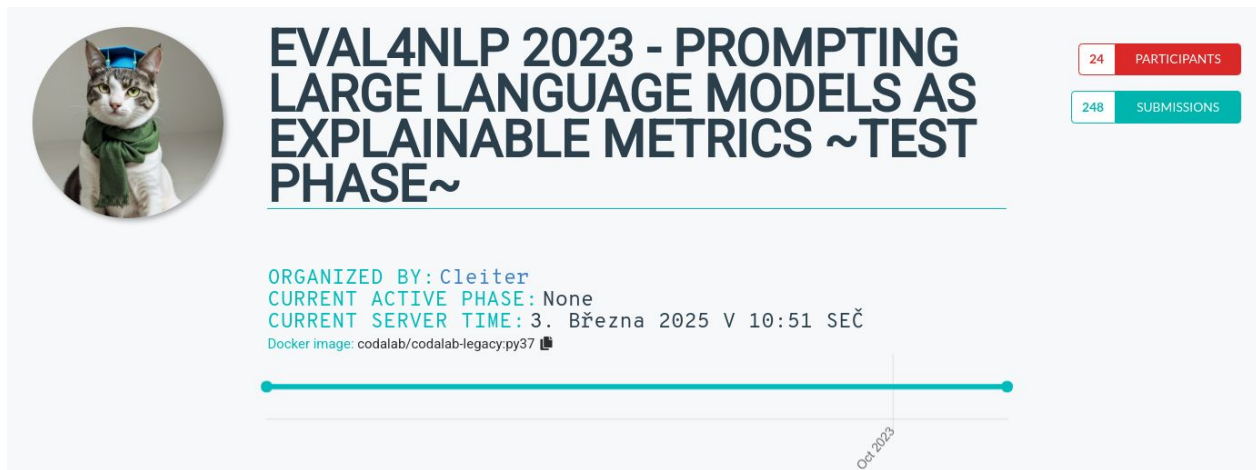Score + explanation

Zdeněk Kasner
{kasner}@ufal.mff.cuni.cz

Approaches: probabilities / direct output / agents…

Evaluation: scores from LLMs compared
with scores from human annotators.
→ Kendall's tau, Spearman and Pearson coefficient

Zdeněk Kasner
{kasner}@ufal.mff.cuni.cz

EVAL4NLP 2023 - PROMPTING LARGE LANGUAGE MODELS AS EXPLAINABLE METRICS ~TEST PHASE~

24 PARTICIPANTS
248 SUBMISSIONS

ORGANIZED BY: Cleiter
CURRENT ACTIVE PHASE: None
CURRENT SERVER TIME: 3. Března 2025 V 10:51 SEČ
Docker image: codalab/codalab-legacy:py37

Oct 2023

Focus on prompting techniques (finetuned models not allowed)

Zdeněk Kasner
{kasner}@ufal.mff.cuni.cz