Seminar CS715

Solving Complex Problems with Large Language Models

Presentation Schedule
### Presentation Schedule – Morning Session

<table>
<thead>
<tr>
<th>Slot</th>
<th>Topic</th>
<th>Student</th>
<th>Mentor</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:20</td>
<td>From Self-consistency to MedPrompt</td>
<td>Florian Bauer</td>
<td>Alexander Brinkmann</td>
</tr>
<tr>
<td>10:37</td>
<td>Prompt Search / Breeding</td>
<td>Shivam Suchak</td>
<td>Ralph Peeters</td>
</tr>
<tr>
<td>10:54</td>
<td>Contrastive Prompting</td>
<td>Ricarda Reiner</td>
<td>Keti Korini</td>
</tr>
<tr>
<td>11:11</td>
<td>LLM-based Agents</td>
<td>Max Meider</td>
<td>Christian Bizer</td>
</tr>
<tr>
<td>11:28</td>
<td>Agent Cooperation</td>
<td>Okan Göktepe</td>
<td>Christian Bizer</td>
</tr>
</tbody>
</table>

- Presentation time: 12 minutes. Discussion: 5 minutes
- Please email us your slides after your presentation
## Presentation Schedule – Afternoon Session

<table>
<thead>
<tr>
<th>Slot</th>
<th>Topic</th>
<th>Student</th>
<th>Mentor</th>
</tr>
</thead>
<tbody>
<tr>
<td>13:45</td>
<td>Limitations of LLMs</td>
<td>Aaron Koßler</td>
<td>Steffen Eger</td>
</tr>
<tr>
<td>14:02</td>
<td>Task Contamination</td>
<td>Saman Khursheed</td>
<td>Ralph Peeters</td>
</tr>
<tr>
<td>14:19</td>
<td>LLMs as Evaluation Metrics</td>
<td>Fabian Rajwa</td>
<td>Jonas Belouadi</td>
</tr>
<tr>
<td>14:36</td>
<td>LLMs with Tools as Evaluation Metrics</td>
<td>Priscilla Chyrva</td>
<td>Daniil Larionov</td>
</tr>
<tr>
<td>14:53</td>
<td>LLM Self-Evaluation during Fine-tuning</td>
<td>Sara Koni</td>
<td>Christoph Leiter</td>
</tr>
<tr>
<td>Break</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15:30</td>
<td>Evaluation of Code Writing Ability of LLMs</td>
<td>Eric Wieland</td>
<td>Ralph Peeters</td>
</tr>
<tr>
<td>15:47</td>
<td>WebAPI Query Planning Using LLM</td>
<td>Mayte Dächer</td>
<td>Keti Korini</td>
</tr>
<tr>
<td>16:04</td>
<td>Attribute Value Normalization Using LLMs</td>
<td>Avani Ghandi</td>
<td>Alexander Brinkmann</td>
</tr>
<tr>
<td>16:21</td>
<td>LLM for Literary Translation and Evaluation</td>
<td>Jiyeon Lee</td>
<td>Ran Zhang</td>
</tr>
<tr>
<td>16:38</td>
<td>Multimodal Reasoning</td>
<td>Thuy Nghiem</td>
<td>Steffen Eger</td>
</tr>
</tbody>
</table>
Next Steps
<table>
<thead>
<tr>
<th>Date</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thursday, 29.02.2024</td>
<td>Kick-off meeting and topic/mentor assignment</td>
</tr>
<tr>
<td>(15:30-17:00)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Read papers about your topic</td>
</tr>
<tr>
<td></td>
<td>Search for additional literature</td>
</tr>
<tr>
<td></td>
<td>Design experimental setup</td>
</tr>
<tr>
<td></td>
<td>Prepare outline and argumentation for your presentation</td>
</tr>
<tr>
<td>Until 20.03.2024</td>
<td>Meet with your mentor to discuss outline and/or experimental setup</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prepare draft of your presentation</td>
</tr>
<tr>
<td>Until 15.4.2024</td>
<td>Send draft presentation to your mentor</td>
</tr>
<tr>
<td></td>
<td>Finalize your presentation</td>
</tr>
<tr>
<td>Monday, 29.04.2024</td>
<td>Presentation and discussion of your topic (30 % of your final grade)</td>
</tr>
<tr>
<td>(10:00-12:00)</td>
<td></td>
</tr>
<tr>
<td>(13:45-17:00)</td>
<td></td>
</tr>
<tr>
<td>Friday, 21.06.2024</td>
<td>Submission of your seminar thesis (70 % of your final grade)</td>
</tr>
</tbody>
</table>
Formal Requirements

- Written report (paper)
  - 12-15 pages single column
    - including abstract and appendixes
    - not including bibliography
    - not including the page about generative AI usage
    - every additional page reduces your grade by 0.3
  - written in English
  - use latex template of Springer Computer Science Proceedings

- Final grade
  - 70% written report
  - 30% presentation
Statement About the Tools that You Used

Your report **must include an extra page** about

1. **which generative AI tools you used**
   - ChatGPT, OpenAI API, Dall3, Perplexity, …. 

2. **for which purposes**
   - structuring your paper
   - summarizing related work
   - writing text for specific chapters
   - improving English grammar and formulations
   - designing experimental setup
   - writing code
   - writing prompts
   - generating training data
   - error analysis
   - …. 

3. **How useful was each tool for this?**

## Example of a Generative AI Tools Declaration

<table>
<thead>
<tr>
<th>Tool</th>
<th>Purpose</th>
<th>Where?</th>
<th>Useful?</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChatGPT</td>
<td>Rephrasing</td>
<td>Throughout</td>
<td>+</td>
</tr>
<tr>
<td>ChatGPT</td>
<td>Summarization of related work</td>
<td>Sec 2</td>
<td>-</td>
</tr>
<tr>
<td>ChatGPT</td>
<td>Structure of thesis</td>
<td>Sec 1</td>
<td>~</td>
</tr>
<tr>
<td>Dall-E</td>
<td>Image generation</td>
<td>Fig 2,3</td>
<td>++</td>
</tr>
<tr>
<td>GPT4</td>
<td>Code generation</td>
<td>music.py</td>
<td>~</td>
</tr>
<tr>
<td></td>
<td></td>
<td>functions.py</td>
<td></td>
</tr>
<tr>
<td>GPT4</td>
<td>Training data augmentation</td>
<td>Extended_set.csv</td>
<td>++</td>
</tr>
</tbody>
</table>
Goals of Literature and Experimental Papers

- **Goals of Experimental Papers**
  1. describe the *(prompt engineering) techniques* from the selected papers
  2. summarize the *evaluation tasks and results* from the papers
  3. design *experimental setup* to evaluate technique on different task
  4. compare your results to the results from the paper

- **Goals of Literature Papers**
  1. describe the *problem / task*
  2. describe several *existing methods/systems* for handling the task,
  3. compare the methods/systems and their *evaluation* using a systematic *set of comparison criteria*
How to Structure Your Experimental Paper?

1. Introduction and Problem Statement
   • Which problem is addressed? What is the **overall approach** for addressing it?
   • Overview of the existing methods/papers and use cases for the evaluation (3 pages+)
   • Structure of your paper

2. Description of Your Experimental Design
   • How do you select **examples** for which **challenges**?
   • Which **prompt designs** and **language models** do you test?

3. Presentation of Experimental Results
   • Present the **results** of your experiments (tables containing values and deltas).
   • Present the results of your **error analysis** (types of errors, frequency of these types)

4. Conclusion
   • What did the experiments and the error analysis show?
   • How do your results compare to the experiments presented in the papers?

5. Bibliography
How to Structure Your Literature Paper?

1. Introduction and Problem Statement
   • Which problem/task is addressed? Why is the problem important?
   • Structure of your paper

2. Description of Existing Approaches
   • Overview of existing methods and features used by the methods
   • Detailed description of selected methods (likely two)
   • Comparison of the selected methods using a set of comparison criteria

3. Evaluation
   • Comparison and discussion of the evaluation tasks, metrics
   • Comparison of the evaluation results using a set of comparison criteria

4. Conclusion
   • What did the comparison of the methods and evaluation results show?
   • Can something be concluded for future work?

5. Bibliography
Learn from Examples

- Read survey articles and previous experimental papers and identify the structure from the previous slides
  - Why can this paragraph be found at that position?
  - What is the purpose of some section / subsection?

- Some relevant surveys

- Textbook on how to write a thesis
Citing Different Types of Publications

1. Journal article
   • Good to cite, current research results

2. Conference and workshop paper
   • Good to cite, current research results

3. Survey articles
   • Good to cite as overviews for specific topics, but prefer individual papers as reference for specific systems

4. Books (sometimes cited)
   • Textbooks
   • Collections of articles/papers => Cite specific paper in book

5. Websites
   • better not cited, exceptions are, e.g., documents like W3C Specifications
   • Do not cite Wikipedia, ever!
   • Use footnotes to refer to project pages, download pages, or technical documentation

6. Slide sets (especially from our lectures)
   • Never cite!
How to Find Relevant Publications?

1. Start with gathering relevant papers from the surveys

2. Exploit references: Given a relevant document x
   • Follow references in the past: papers y that x has cited
   • Follow references in the future: papers y that cited x
     ("cited by" functionality in Google scholar)

3. Use Google Scholar or Semantic Scholar
   • we use it a lot ourselves
Goals of Literature and Experimental Papers

- Goals of Literature Papers
  1. describe the problem / task
  2. describe several existing methods/systems for handling the task,
  3. compare the methods/systems and their evaluation using a systematic set of comparison criteria

- Goals of Experimental Papers
  1. describe the prompt engineering technique from the paper
  2. present evaluation task and results from the paper
  3. design experimental setup to evaluate technique on different task
  4. compare your results to the results from the paper
How to Structure Your Literature Paper?

1. Introduction and Problem Statement
   • Which problem/task is addressed? Why is the problem important?
   • Structure of your paper

2. Description of Existing Approaches
   • Overview of existing methods and features used by the methods
   • Detailed description of selected methods (likely two)
   • Comparison of the selected methods using a set of comparison criteria

3. Evaluation
   • Comparison and discussion of the evaluation tasks, metrics
   • Comparison of the evaluation results using a set of comparison criteria

4. Conclusion
   • What did the comparison of the methods and evaluation results show?
   • Can something be concluded for future work?

5. Bibliography
How to Structure Your Experimental Paper?

1. Introduction and Problem Statement
   - Which problem is addressed? What is the overall approach for addressing it?
   - Overview of the existing methods/papers and use cases for the evaluation
   - Structure of your paper

2. Description of Experimental Design
   - What is your How to you select examples for which challenges?
   - Which prompt designs and language models do you test?

3. Presentation of Experimental Results
   - Present the results of your experiments (tables containing values and deltas).
   - Present the results of your error analysis (types of errors, frequency of these types)

4. Conclusion
   - What did the experiments and the error analysis show?
   - How to your results compare to the experiments presented in the papers?

5. Bibliography
Learn from Examples

- Read survey articles and previous experimental papers and identify the structure from the previous slides
  - Why can this paragraph be found at that position?
  - What is the purpose of some section / subsection?

- Important
  - Read survey articles!
  - Read conference or journal papers

- Some relevant surveys

- Textbook on how to write a thesis
Citing Different Types of Publications

- **Journal article**
  - Good to cite, current research results
  - Survey articles (very good for an overview)

- **Conference and workshop paper**
  - Good to cite, current research results

- **Books (sometimes cited)**
  - Textbooks
  - Collections of articles/papers => Cite specific paper in book

- **Websites**
  - better not cited, exceptions are, e.g., documents like W3C Specifications
  - **Do not cite Wikipedia, ever!**
  - **Use footnotes** to refer to project pages, download pages, or technical documentation

- **Slide sets (especially from our lectures)**
  - **Never cite!**
How to Find Relevant Publications?

- Use Standard Search Engines
- **Use Google Scholar**
  - we use it a lot ourselves
- Search Engines of the University’s library
  - see slides from the library course
- **Exploit references:** Given a relevant document $x$
  - Follow references in the past: papers $y$ that $x$ has cited
  - Follow references in the future: papers $y$ that cited $x$
    ("cited by" functionality in Google scholar)