Conference Report: ACL 2025*

Roman Klinger

2025-08-01

Conference of the Association for Computational Linguistics (ACL)

End of July 2025, I participated in the ACL in Vienna. ACL is the biggest natural language processing conference, and (currently) one of only two conferenes which are considered A* (according to the CORE Ranking. The other A* conference is EMNLP – the conferences of the chapters are considerd A (EACL, NAACL) or B (IJCNLP). While I do not think that the differences between A and A* are too important; and sometimes also B conferences are preferably when they are topically more relevant for a paper, this aspect makes this conference to be a very popular.

Statistics

This has been the biggest conference ever, as far as I know. Regarding the number of participants, I heard various numbers, ranging from 5400 participants on site (plus 1500 online participants) to 6400 participants in Vienna. During the opening session, the following numbers were mentioned:

- 1700 main conference
- 1400 Findings papers
- 108 industry papers
- 800 workshop papers
- 104 student research papers
- 64 demo papers

The review process went, again through ARR, where papers are first reviewed independent of a concrete conference, and after a potential revise and resubmit, when the scores look like the paper could be accepted, it is “committed” to a conference. 4700 papers have been committed from the directly preceding review cycle, out of which 1600 were revisions. However, also 800 papers have been directly submitted from the review cycle preceding this one, presumably because the authors did prefer to submit to a conference in Europe.

One topic that triggered quite some “oh” and “ah” in the opening presentation was the aspect that chinese authors contributed 50% of the papers and the US only 19%

*Published as blog post at https://www.romanklinger.de/blog/2025-08-01-acl/

a substantial increase and decrease. In social media, some people attributed this presumed decrease in productivity to the new administration in the US. I find this questionable reasoning, there are many aspects affecting where papers are sent, and in 2025 also a NAACL took place in Albuquerque. At the same time, people in China might prefer to commit to a conference in Europe. South Korea continued to increase the number of contributions, followed by UK and Germany (all 3%).

Quite a diverse set of topics was present at the conference, but the pie chart is not representing well how many papers did focus on work with large language models. I think most papers did at least use LLMs for their modeling experiments in some way.

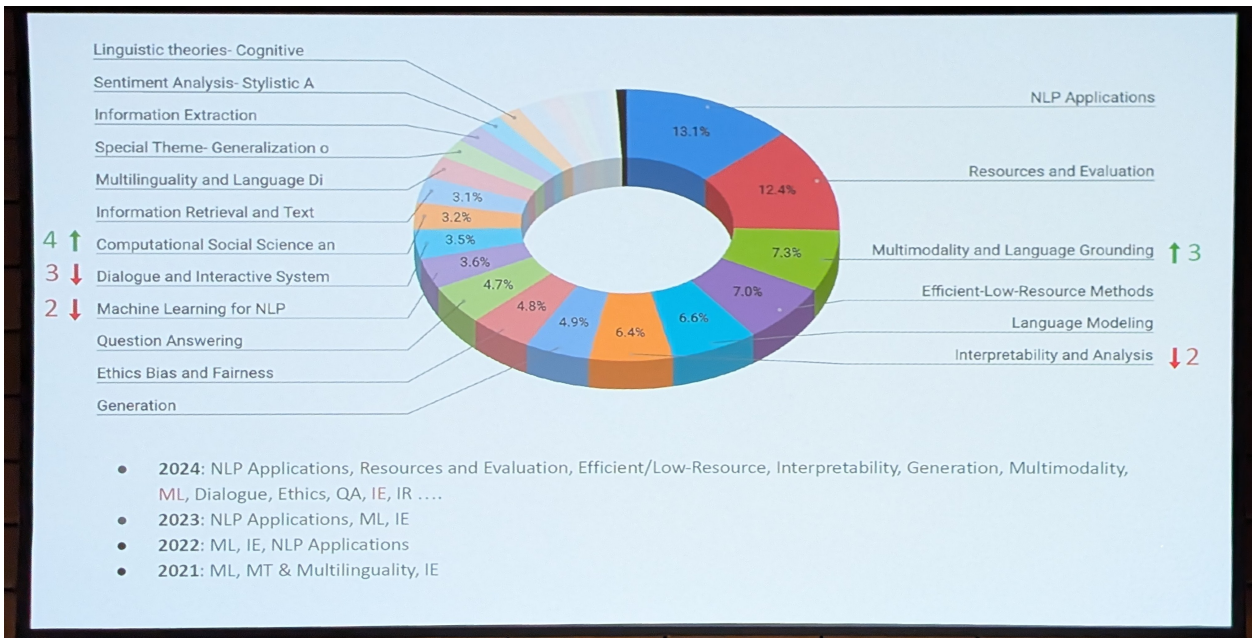

Figure 1: Distribution of topics

Organization

The organizers at this conference made a couple of decisions to do things differently than in previous editions. I'd like to share my opinion about them.

- *There was no poster session in parallel to oral presentations.* This lead to the poster sessions being very crowded, while the lecture halls were (presumably) empty. Maybe this needed to be like that for some reason, but I would have preferred to be able to chose between posters and talks in parallel sessions, and the poster sessions being smaller.
- *There was no apparent clustering of the poster.* I did not perceive a topical clustering of the posters, and for me, it was sometimes necessary to walk quite far distances from one interesting poster to another. Also, I like to randomly roam around posters that I might like. This did not work at all. One needed, before the poster session, make a list of posters to see and directly go there.
- *The talks were partially in tiny rooms.* First I thought "this is nice". The session on argument mining was in a room for about 40-60 people. That felt like an environment where people could actually discuss. Once the room was full, people

needed to wait outside and couldn't join, this impression was not that positive any more. This was particularly bad in one workshop for which I registered and in which I presented a poster. The poster presentation was 5 minutes away from the room for the workshop, and when I tried to come back, I couldn't participate in the workshop. This was very frustrating. My impression is that the venue just did not have enough rooms of a sufficient size; so that is something that couldn't easily be solved. But at least in the workshops, people who did register should have given preference.

- *Panel discussions after presentations*. This was an interesting experiments. After speakers gave presentations and answered questions, there was a short panel discussion amongst the speakers of one session, moderated by the session chair. In principle I think this is a great idea, but the papers were not close enough topic-wise in the sessions I participated in with this format such that this worked out. I'd like to see a second iteration of this idea though.

Altogether, this was, for me, the most difficult conference to navigate, so far. I am not sure if this was because of the shere size or because of other reasons. Clustering posters by topic similarity would definitely be a big wish from me for the next conferences.

Figure 2: Poster Session

Own Contributions to the Conference

We had a set of contributions to this conference, which I would like to briefly summarize in the following.

- Schäfer et al. (2025) reports on our experiments with socio-demographic prompting for offensive language detection with instruction-tuned models. We tested if

socio-demographic prompting (make prediction from the perspective of a person with a particular age or gender) has an effect stronger than pseudo-demographic prompts (the house number of a person). Further, we tested to which demographics a prediction is most similar if no demographics are provided in the prompt. We found, as expected, that some particular demographics seem to be better represented in large language models. This paper was the result of a joint effort from the whole group – writing one paper in one week during our first retreat. We are very happy that this paper made it into the main conference; but we also agree that the stress level of writing a paper in this short amount of time was too high. By the way, thanks to Steffen Eger for the idea to do such type of retreat. It worked really well to get to know each other and that was clearly a success in the retreat.

- Bagdon et al. (2025) studied various ways to get emotion and appraisal annotated data. In our project ITEM (with Carina Silberer from Stuttgart), we investigate how and why social media users express their emotions, particularly implicitly with text and images. In the paper we published at the main conference of ACL 2025, we wanted to understand if asking crowdworkers to create a post for a given emotion and including an image that would realistically use from an image data base works as a reasonable approximation for realistic data. The advantage would be that the data has less data privacy issues and copyright issues than real data. We compared these data to "donation" (we paid for them though) of real posts from users. We find that the experimentally elicited data is fine as training data, but to study the phenomena one needs real test data.
- Greschner, Wührl, and Klinger (2025) presented her work on the question if we can automatically detect aspects that influence the perceived quality of life of people with mental disorders from social media, which are not yet known. To do so, she annotated data, built classifiers, and did topic modeling and found a set of aspects that were not yet represented in standardized test instruments.
- Papay, Klinger, and Padó (2025) proposed a method to consider long-distance relations in text on the output level – with a conditional random field (joint work with Sebastian Pado from Stuttgart). This CRF could be put on top of a neural network and is therefore a relevant option for an output layer. Most importantly, Sean found a way to decode in linear runtime, which is not generally possible for loopy probabilistic structures.
- Jiahui Li and Klinger (2025) published the first paper from our INPROMPT project, in which we develop prompt optimization and engineering methods that involve a human user whenever automatic optimization is not sufficiently successful. Therefore, the proposed methods support human prompt developers. The paper in the student research workshop summarizes the project plans and discusses the upcoming research questions and tasks.
- Hofmann, Sindermann, and Klinger (2025) has been presented by me, but the work has been conducted mainly by Jan Hofmann (in collaboration with Cornelia Sindermann from Stuttgart and Ulm). In this work, we studied language model based agents which learn which posts in a social media profile are helpful for personality profiling. The data is only annotated on the profile level, so we use a reward function for reinforcement learning to learn to distinguish relevant and irrelevant posts. The method could readibly be transferred to any other long-text analysis and shows substantial runtime and cost savings, because the prompt that makes the personality prediction can work without a lot of provided context.

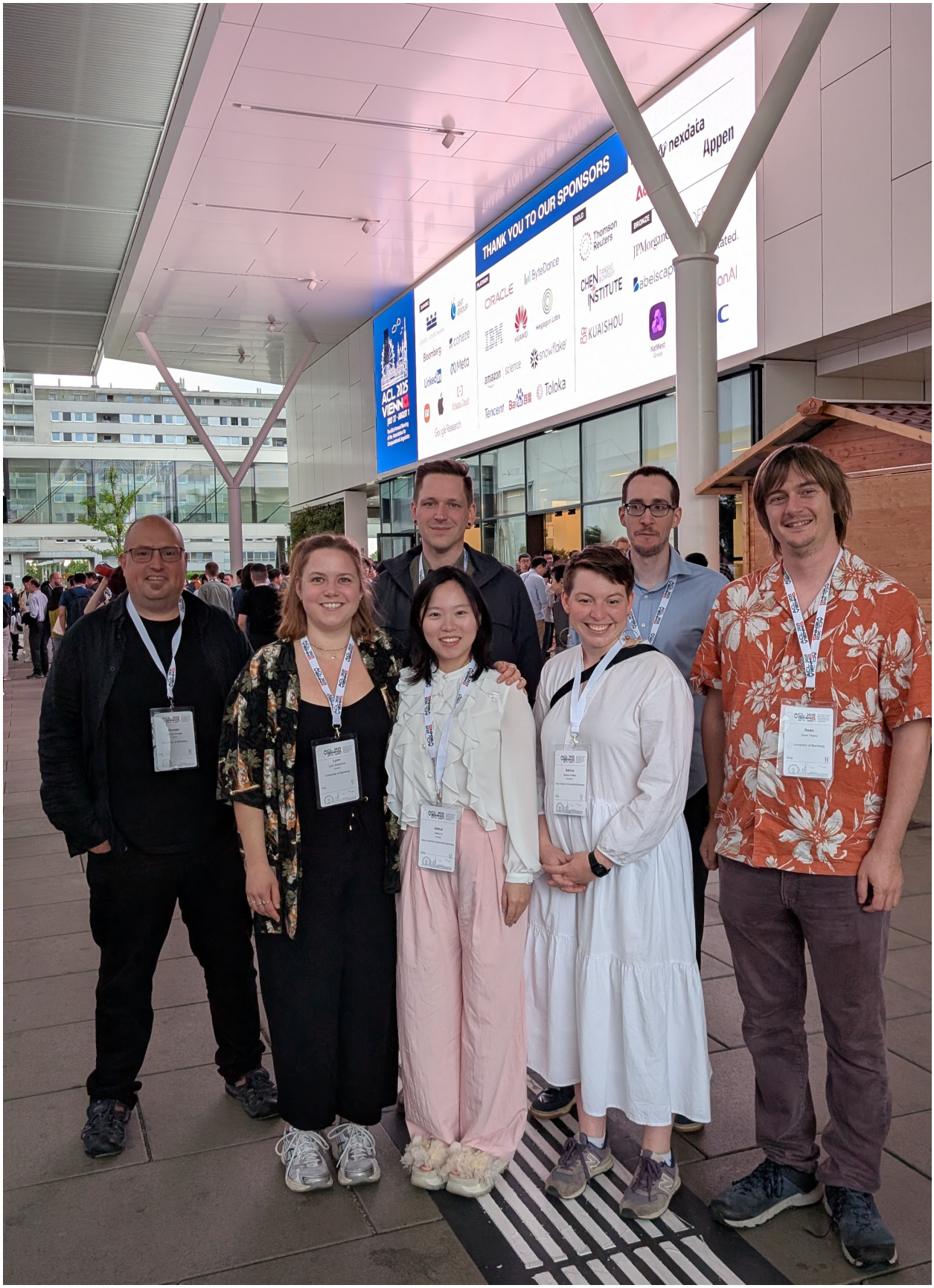

Figure 3: Photo of the BamNLP Group in front of the conference venue

My Favorite Contributions

In the following, I want to highlight some papers that have been presented (or at least published) at this ACL. As I said - I found this conference particularly difficult to navigate; and if I don't mention a paper that you would expect me to like it doesn't mean that I didn't like it. I probably just missed it (and I'd appreciate if you told me about that paper that I should read.).

Emotion analysis

- Palma et al. (2025) aim at understanding where emotion and sentiment information is represented in large language models. They then train small models on the local emotion/sentiment representation which works better than fine-tuning the whole model (and it is cheaper).
- Du and Hoste (2025) propose to calculate annotator disagreement not based on categorical values but instead map them to a valence and arousal space in which the continuous values are used for an error estimation. They show that such disagreement calculation is a more realistic estimate.
- Barz et al. (2025) do also focus on inter-annotator agreement, but more on understanding (the reasons for) disagreement. The authors annotate a corpus on environmental aspects and analyze it for topics and emotion distributions. Understanding disagreement was mostly analyzed based on qualitative interviews and less on statistical analyses. One main reason for disagreement were different perspectives, another reason to build personalized models and include contextuali information in corpora (like we did for instance in Troiano, Oberländer, and Klinger (2023), but the idea of qualititative interviews in this ACL2025 paper are a good idea that I really like).
- Lee, Lee, et al. (2025) detect neurons that are particularly relevant for particular emotions and show that removing them comes with a drop in emotion classification performance.
- Jiayi Li et al. (2025) reproduce prior work that shows that readers have limited ability to reproduce writer's emotions; and LLMs are better than humans. The particular novelty is that the authors distinguish ingroup and outgroup annotations. The related work section is unfortunately a bit limited in this paper - there has been work that failed to show such influence of demographic factors in natural language processing (while its known across other modalities). I still would like to understand which factors influence if in/outgroup context matters or not.
- Lee, Jang, et al. (2025) is an interesting study, because the authors use entirely automatically generated data, and then study how language models analyze this artificial data. Currently, I do not have a good understanding what the findings in the paper mean - because neither the data nor the annotations are human-made or naturally occurring. I admit that the data is generated based on human data though, but it is not clear to me if the findings therefore generalize to data as it occurs in the wild naturally. A similar criticism also applies to a lot of studies we do, in which we elicit data from humans in non-natural experimental environments. I think the question how much such analysis allow interesting insights is still an open research question.

- Muhammad et al. (2025) is not just another emotion data set. It is a corpus for many languages, and many of which did not receive enough attention yet. The corpus is manually annotated, contains many domains and various genres. It contains intensity and categorical labels.
- Duong et al. (2025) is the first work that I am aware of that annotates emotion expressions for bodily reactions. We did also find in Casel, Heindl, and Klinger (2021) that a substantial number of emotion expressions use body descriptions, so it is really nice to see this work. The authors also rely on automatic annotation with best-worst scaling, as proposed by Bagdon et al. (2024).

Appraisals in Emotion Analysis

- Tak et al. (2025) builds on top of our Crowd-enVent corpus to study the cognitive evaluation process taking place in emotion event processing. While our corpus only provided emotion and appraisal annotation and predictions (Troiano, Oberländer, and Klinger (2023)) the authors of this paper really focus on understanding how LLMs process emotions and if that process is aligned with human processing. To do so, they build on top of the idea of mechanistic interpretability, by probing the model. A very impressive idea in this paper is to make use of the model understanding to then intervene on the cognitive evaluation process to study the relation to the emotion category. I like appraisals and the authors use our data, so I am biased, but this paper goes the extra mile to bring together LLM introspection methods with psychological concepts.
- Yeo and Jaidka (2025) build on top of appraisals, which they consider to be a fundament for the interpretation of implicitly expressed emotions, to curate a data set focused on the Theory of Mind. They focus therefore not so much on the analysis of emotions from one particular perspective, but on the interpretation of an emotion in a person as a private state. I think this is also quite related to various work on empathy. While I really like the idea, this paper suffers a bit from the lack of a related work section (due to it being a short paper, but still, the context of this work is a bit opaque for me).
- Debnath, Graham, and Conlan (2025) train an appraisal predictor on our appraisal data set Crowd-enVent and automatically label dialogue data to study the information flow in dialogues. The paper therefore brings together event-centered emotion analysis (Klinger (2023)) and emotion recognition in conversations (Pereira, Moniz, and Carvalho (2024)). They do so in a multi-task learning setup, which may also benefit from the emotion labels in the conversation data.

Personality

- Wei et al. (2025) ensure that a dialogue, guided by an LLM is consistent regarding the emotion and the personality. They do so by modeling the emotion and personality transitions with a Markov chain. What is not clear for me in this paper is if personality and emotions are handled differently according to the fact that one is a state and the other are traits.
- Lim et al. (2025) show how agents in text-based games change their behaviour based on different personality traits.

- Hartley et al. (2025) study how LLMs change their risk-taking behaviour based on differing personality traits given as conditions. This work is related to our work on measuring regulatory focus theory (RFT, promotion or prevention orientation), but we did only build classifiers (Velutharambath, Sassenberg, and Klinger (2023)). The authors here do use personality conditions for guiding the behaviour of an agent. Bringing RFT and such studies together could be an interesting step in future work.

Other

- Wu et al. (2025) may be the first paper on music information retrieval I have seen at ACL conferences. They authors align sheet music, audio recordings, performance data and multilingual text for an improved retrieval process.
- Quensel, Falk, and Lapesa (2025) study subjective factors of argument strengths. Their work aggregates various aspects such as emotions, hedging and storytelling in a joint analysis. The emotion labels stem from a domain transfer of a predefined corpus. Next to our work ((Greschner and Klinger (2025))[https://aclanthology.org/2025.nlp4dh-1.52/]) this is one of the few studies that do not consider binary emotionality but distinguish various emotion categories.
- Menis Mastromichalakis et al. (2025) advcocate for not removing harmful information from historic sources; but instead automatically contextualize the information, such that it is better understood. I find this is an interesting perspective on offensive language processing.
- Pramanick et al. (2025) is a meta-study on the research field of NLP. The authors show empirically that the focus on language shifts towards more computational methods, people care more about human-centric studies, and that there is a steady increase in methods and data sets.
- Russell, Karpinska, and Iyyer (2025) probably has the best title in this conference, because it makes it very easy to summarize the main result: “People who frequently use ChatGPT for writing tasks are accurate and robust detectors of AI-generated text”
- Sicilia and Alikhani (2025) also study theory of mind (as mentioned further above), but with a focus on uncertainty prediction. The authors propose a benchmark to evaluate the uncertainty in participants in a dialogue. Therefore, the prediction is really not about the language model, but of a second order. Very interesting idea and a new twist to uncertainty prediction!
- Corso, Pierri, and De Francisci Morales (2025) propose data and methods to find conspiracy theories on TikTok. An interesting task and setup. What remains is a study what the properties of these conspiracy theories are, and if also novel instances can be found. Otherwise, the task might not focus on properties of the instances, but only on similarities.
- F. Chen et al. (2025) ask people to judge the own perceived empathy in a story, without clearly defining the task for the annotators. This is an interesting idea, because it leaves the decision what “empathy” actual means to the annotators. Maybe it is related to stance or opinion in this setup.
- Jin et al. (2025) is the first work I have seen that does study argument quality with a clear perspectivism angle - people with different backgrounds assess arguments

differently. Unfortunately, the persona descriptions are automatically generated, and the assessment and rational generation also seems to only be automatic. It is not clear to me if there is human annotation from these various personas is involved.
- Yang and Jin (2025) perform book-long evaluations automatically, but with the help of human assigned scores. The setup is quite interesting: the authors automatically structure book reviews into a structured representation; and then, they develop methods to automatically assess these scores from the book alone. This is challenging because of the text length, and the authors propose various approaches for aggregation into a shorter representation.
- Cahyawijaya et al. (2025) is a paper that is not exactly related to my main interests - it's about a data set to develop vision-language models. The interesting aspect for me here is that the authors evaluate different ways to collect the data: they crowdsource, crawl or generate. Therefore, this paper is quite related to the corpus we publish at the same conference. Our paper is called "Donate or Create". While both terms in our case refer to crowdsourcing, there is an interesting overlap in methodology (Bagdon et al. (2025)). The authors of this paper, however, do also evaluate automatic data generation, which is something we did not do (yet). By the way, it's also the first paper which has enough authors such that the abstract continues on the second page ;-).
- Bavaresco et al. (2025) is a very nice exception from the many papers that ask "can LLMs do X" by studying the same question in a systematic manner, across many tasks. I think this is a very natural but very well carried out study that consolidates various ideas that came up in recent work. I assume this will be one of the mostly highly cited papers in this conference.
- Y. Chen and Eger (2025) describes results that come from the same project as Greschner and Klinger (2025). The authors of this paper do automatically generate non-emotional arguments and emotional arguments with language models, to setup a human annotation study in a controlled manner.

The whole proceedings are available in the ACL Anthology.

Venue and Place

The conference took place in Vienna - a city I recently visited for KONVENS, so my pressure to do sightseeing was not too strong. The conference was north of the Danube, in an area I have not seen so far, and it was mostly a modern concrete building with more high concrete buildings around. What was really nice is that I could cycle every day from the hotel over a bridge to the venue. Further, there was a beach/river promenade-like area with some restaurants around; where one could also go swimming. This was quite nice.

The social event took place in the conference venue, probably the only possible decision with a conference of this size. I still think that such conference dinners should not serve meat, given how many animals they alone are responsible then to kill, but with this opinion I seem to be quite alone. The vegetarian food quality was good, though.

Next to the unavoidable (and probably expected-by-many) Waltz session, the DJ had a sax player and two singers; and they were playing Electro Swing. I did unfortunately not learn who that was, but if any of you knows, please tell me. I like this type of music

quite a lot and was very happy about this; such parties do not take place in areas in which I live. Dancing was, however, not possible for me - the floor was moving so strongly that I couldn't stay in this area without fear ;-). (nothing happened though)

Figure 4: Social Event

Bibliography

Bagdon, Christopher, Aidan Combs, Carina Silberer, and Roman Klinger. 2025. "Donate or Create? Comparing Data Collection Strategies for Emotion-Labeled Multimodal Social Media Posts." In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 17307–30. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-long.847/.

Bagdon, Christopher, Prathamesh Karmalkar, Harsha Gurulingappa, and Roman Klinger. 2024. "'You Are an Expert Annotator': Automatic Best–Worst-Scaling Annotations for Emotion Intensity Modeling." In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, edited by Kevin Duh, Helena Gomez, and Steven Bethard, 7924–36. Mexico City, Mexico: Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.439.

Barz, Christina, Melanie Siegel, Daniel Hanss, and Michael Wiegand. 2025. "Understanding Disagreement: An Annotation Study of Sentiment and Emotional Language in Environmental Communication." In *Proceedings of the 19th Linguistic Annotation Workshop (LAW-XIX-2025)*, edited by Siyao Peng and Ines Rehbein, 1–20. Vienna, Austria: Association for Computational Linguistics. https://aclantholo

gy.org/2025.law-1.1/.

Bavaresco, Anna, Raffaella Bernardi, Leonardo Bertolazzi, Desmond Elliott, Raquel Fernández, Albert Gatt, Esam Ghaleb, et al. 2025. "LLMs Instead of Human Judges? A Large Scale Empirical Study Across 20 NLP Evaluation Tasks." In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 238–55. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-short.20/.

Cahyawijaya, Samuel, Holy Lovenia, Joel Ruben Antony Moniz, Tack Hwa Wong, Mohammad Rifqi Farhansyah, Thant Thiri Maung, Frederikus Hudi, et al. 2025. "Crowdsource, Crawl, or Generate? Creating SEA-VL, a Multicultural Vision-Language Dataset for Southeast Asia." In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 18685–717. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-long.916/.

Casel, Felix, Amelie Heindl, and Roman Klinger. 2021. "Emotion Recognition Under Consideration of the Emotion Component Process Model." In *Proceedings of the 17th Conference on Natural Language Processing (KONVENS 2021)*, edited by Kilian Evang, Laura Kallmeyer, Rainer Osswald, Jakub Waszczuk, and Torsten Zesch, 49–61. Düsseldorf, Germany: KONVENS 2021 Organizers. https://aclanthology.org/2021.konvens-1.5/.

Chen, Francine, Scott Carter, Tatiana Lau, Nayeli Suseth Bravo, Sumanta Bhattacharyya, Kate Sieck, and Charlene C. Wu. 2025. "Empathy Prediction from Diverse Perspectives." In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 8959–74. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-long.439/.

Chen, Yanran, and Steffen Eger. 2025. "Do Emotions Really Affect Argument Convincingness? A Dynamic Approach with LLM-Based Manipulation Checks." In *Findings of the Association for Computational Linguistics: ACL 2025*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 24357–81. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.findings-acl.1250/.

Corso, Francesco, Francesco Pierri, and Gianmarco De Francisci Morales. 2025. "Conspiracy Theories and Where to Find Them on TikTok." In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 8346–62. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-long.408/.

Debnath, Alok, Yvette Graham, and Owen Conlan. 2025. "An Appraisal Theoretic Approach to Modelling Affect Flow in Conversation Corpora." In *Proceedings of the 29th Conference on Computational Natural Language Learning*, edited by Gemma Boleda and Michael Roth, 233–50. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.conll-1.16/.

Du, Quanqi, and Veronique Hoste. 2025. "Another Approach to Agreement Measurement and Prediction with Emotion Annotations." In *Proceedings of the 19th Linguistic Annotation Workshop (LAW-XIX-2025)*, edited by Siyao Peng and Ines Rehbein, 87–102. Vienna, Austria: Association for Computational Linguistics. https:

//aclanthology.org/2025.law-1.7/.

Duong, Phan Anh, Cat Luong, Divyesh Bommana, and Tianyu Jiang. 2025. "CHEER-Ekman: Fine-Grained Embodied Emotion Classification." In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 1118–31. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-short.88/.

Greschner, Lynn, and Roman Klinger. 2025. "Fearful Falcons and Angry Llamas: Emotion Category Annotations of Arguments by Humans and LLMs." In *Proceedings of the 5th International Conference on Natural Language Processing for Digital Humanities*, edited by Mika Hämäläinen, Emily Öhman, Yuri Bizzoni, So Miyagawa, and Khalid Alnajjar, 628–46. Albuquerque, USA: Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.nlp4dh-1.52.

Greschner, Lynn, Amelie Wührl, and Roman Klinger. 2025. "QoLAS: A Reddit Corpus of Health-Related Quality of Life Aspects of Mental Disorders." In *ACL 2025*, edited by Dina Demner-Fushman, Sophia Ananiadou, Makoto Miwa, and Junichi Tsujii, 201–16. Viena, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.bionlp-1.18/.

Hartley, John, Conor Brian Hamill, Dale Seddon, Devesh Batra, Ramin Okhrati, and Raad Khraishi. 2025. "How Personality Traits Shape LLM Risk-Taking Behaviour." In *Findings of the Association for Computational Linguistics: ACL 2025*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 21068–92. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.findings-acl.1085/.

Hofmann, Jan, Cornelia Sindermann, and Roman Klinger. 2025. "Prompt-Based Personality Profiling: Reinforcement Learning for Relevance Filtering." In *Proceedings of the 1st Workshop for Research on Agent Language Models (REALM 2025)*, edited by Ehsan Kamalloo, Nicolas Gontier, Xing Han Lu, Nouha Dziri, Shikhar Murty, and Alexandre Lacoste, 1–16. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.realm-1.1/.

Jin, Bojun, Jianzhu Bao, Yufang Hou, Yang Sun, Yice Zhang, Huajie Wang, Bin Liang, and Ruifeng Xu. 2025. "A Multi-Persona Framework for Argument Quality Assessment." In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 12148–70. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-long.593/.

Klinger, Roman. 2023. "Where Are We in Event-Centric Emotion Analysis? Bridging Emotion Role Labeling and Appraisal-Based Approaches." In *Proceedings of the Big Picture Workshop*, edited by Yanai Elazar, Allyson Ettinger, Nora Kassner, Sebastian Ruder, and Noah A. Smith, 1–17. Singapore: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.bigpicture-1.1.

Lee, Jaewook, Yeajin Jang, Oh-Woog Kwon, and Harksoo Kim. 2025. "Does the Emotional Understanding of LVLMs Vary Under High-Stress Environments and Across Different Demographic Attributes?" In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 23196–210. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-long.1130/.

Lee, Jaewook, Woojin Lee, Oh-Woog Kwon, and Harksoo Kim. 2025. "Do Large Language Models Have 'Emotion Neurons'? Investigating the Existence and Role."

In *Findings of the Association for Computational Linguistics: ACL 2025*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 15617–39. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.findings-acl.806/.

Li, Jiahui, and Roman Klinger. 2025. "IPrOp: Interactive Prompt Optimization for Large Language Models with a Human in the Loop." In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, edited by Jin Zhao, Mingyang Wang, and Zhu Liu, 276–85. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-srw.18/.

Li, Jiayi, Yingfan Zhou, Pranav Narayanan Venkit, Halima Binte Islam, Sneha Arya, Shomir Wilson, and Sarah Rajtmajer. 2025. "Can Third Parties Read Our Emotions?" In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 21478–99. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-long.1042/.

Lim, Seungwon, Seungbeen Lee, Dongjun Min, and Youngjae Yu. 2025. "Persona Dynamics: Unveiling the Impact of Persona Traits on Agents in Text-Based Games." In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 31360–94. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-long.1515/.

Menis Mastromichalakis, Orfeas, Jason Liartis, Kristina Rose, Antoine Isaac, and Giorgos Stamou. 2025. "Don't Erase, Inform! Detecting and Contextualizing Harmful Language in Cultural Heritage Collections." In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 21836–50. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-long.1060/.

Muhammad, Shamsuddeen Hassan, Nedjma Ousidhoum, Idris Abdulmumin, Jan Philip Wahle, Terry Ruas, Meriem Beloucif, Christine de Kock, et al. 2025. "BRIGHTER: BRIdging the Gap in Human-Annotated Textual Emotion Recognition Datasets for 28 Languages." In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 8895–8916. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-long.436/.

Palma, Dario Di, Alessandro De Bellis, Giovanni Servedio, Vito Walter Anelli, Fedelucio Narducci, and Tommaso Di Noia. 2025. "LLaMAs Have Feelings Too: Unveiling Sentiment and Emotion Representations in LLaMA Models Through Probing." In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 6124–42. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-long.306/.

Papay, Sean, Roman Klinger, and Sebastian Padó. 2025. "Regular-Pattern-Sensitive CRFs for Distant Label Interactions." In *Proceedings of the 1st Joint Workshop on Large Language Models and Structure Modeling (XLLM 2025)*, edited by Hao Fei, Kewei Tu, Yuhui Zhang, Xiang Hu, Wenjuan Han, Zixia Jia, Zilong Zheng, et al., 26–35. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.xllm-1.4/.

Pereira, Patrícia, Helena Moniz, and Joao Paulo Carvalho. 2024. “Deep Emotion Recognition in Textual Conversations: A Survey.” *Artificial Intelligence Review* 58 (1): 10. https://doi.org/10.1007/s10462-024-11010-y.

Pramanick, Aniket, Yufang Hou, Saif M. Mohammad, and Iryna Gurevych. 2025. “The Nature of NLP: Analyzing Contributions in NLP Papers.” In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 25169–91. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-long.1224/.

Quensel, Carlotta, Neele Falk, and Gabriella Lapesa. 2025. “Investigating Subjective Factors of Argument Strength: Storytelling, Emotions, and Hedging.” In *Proceedings of the 12th Argument Mining Workshop*, edited by Elena Chistova, Philipp Cimiano, Shohreh Haddadan, Gabriella Lapesa, and Ramon Ruiz-Dolz, 126–39. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.argmining-1.12/.

Russell, Jenna, Marzena Karpinska, and Mohit Iyyer. 2025. “People Who Frequently Use ChatGPT for Writing Tasks Are Accurate and Robust Detectors of AI-Generated Text.” In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 5342–73. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-long.267/.

Schäfer, Johannes, Aidan Combs, Christopher Bagdon, Jiahui Li, Nadine Probol, Lynn Greschner, Sean Papay, et al. 2025. “Which Demographics Do LLMs Default to During Annotation?” In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 17331–48. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-long.848/.

Sicilia, Anthony, and Malihe Alikhani. 2025. “Evaluating Theory of (an Uncertain) Mind: Predicting the Uncertain Beliefs of Others from Conversational Cues.” In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 8007–21. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-long.395/.

Tak, Ala N., Amin Banayeeanzade, Anahita Bolourani, Mina Kian, Robin Jia, and Jonathan Gratch. 2025. “Mechanistic Interpretability of Emotion Inference in Large Language Models.” In *Findings of the Association for Computational Linguistics: ACL 2025*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 13090–120. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.findings-acl.679/.

Troiano, Enrica, Laura Oberländer, and Roman Klinger. 2023. “Dimensional Modeling of Emotions in Text with Appraisal Theories: Corpus Creation, Annotation Reliability, and Prediction.” *Computational Linguistics* 49 (1): 1–72. https://doi.org/10.1162/coli_a_00461.

Velutharambath, Aswathy, Kai Sassenberg, and Roman Klinger. 2023. “Prevention or Promotion? Predicting Author’s Regulatory Focus.” Edited by Leon Derczynski. *Northern European Journal of Language Technology* 9. https://doi.org/10.3384/nejlt.2000-1533.2023.4561.

Wei, Yangbo, Zhen Huang, Fangzhou Zhao, Qi Feng, and Wei W. Xing. 2025. “MECoT: Markov Emotional Chain-of-Thought for Personality-Consistent Role-Playing.” In

Findings of the Association for Computational Linguistics: ACL 2025, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 8297–8314. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.findings-acl.435/.

Wu, Shangda, Guo Zhancheng, Ruibin Yuan, Junyan Jiang, SeungHeon Doh, Gus Xia, Juhan Nam, Xiaobing Li, Feng Yu, and Maosong Sun. 2025. "CLaMP 3: Universal Music Information Retrieval Across Unaligned Modalities and Unseen Languages." In *Findings of the Association for Computational Linguistics: ACL 2025*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 2605–25. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.findings-acl.133/.

Yang, Dingyi, and Qin Jin. 2025. "What Matters in Evaluating Book-Length Stories? A Systematic Study of Long Story Evaluation." In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 16375–98. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.acl-long.799/.

Yeo, Gerard Christopher, and Kokil Jaidka. 2025. "Beyond Context to Cognitive Appraisal: Emotion Reasoning as a Theory of Mind Benchmark for Large Language Models." In *Findings of the Association for Computational Linguistics: ACL 2025*, edited by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 26517–25. Vienna, Austria: Association for Computational Linguistics. https://aclanthology.org/2025.findings-acl.1359/.

[Download this post as PDF]