

DiSCourse Seminar

The Digital Science Center and the Department of Political Science would like to invite you to the following talk:

Hauke Licht
University of Innsbruck

Metadata-aware transformer fine-tuning: Mitigating regression-to-mean bias in text classification

Social science researchers increasingly rely on Transformer classifiers fine-tuned on human-annotated data for comparative analysis. Applying these classifiers to new, diverse contexts (e.g., different countries, languages, political parties, or time periods) can lead to systematic biases, particularly regression-to-the-mean bias. This bias occurs when classifiers predict label classes based on their overall distribution in the training data, regardless of subgroup variations. Our study demonstrates that classifiers tend to systematically under- or overestimate label prevalence in subgroups depending on their representation in the training set. To address this, we propose incorporating subgroup metadata during fine-tuning to reduce regression-to-the-mean bias. Our cross-sectional, multilingual, and temporal experiments validate the effectiveness of this approach, improving subgroup performance parity and cross-lingual equivalence. Our findings aim to enhance the reliability and validity of text classification in comparative research.

About the speaker

[Hauke Licht](#) is an Assistant Professor of Computational Political Science at the Department of Political Science and the Digital Science Center of the University of Innsbruck. He develops and applies computational methods for the comparative study of political rhetoric. His particular interest lies in the role of rhetorical strategies in democratic representation, electoral competition, and legislative politics.

Date, Time, Place:

Friday, 14 March 2025, 12:00 (CET), hybrid

Participants are invited to join the event at the Digital Science Center, Innrain 15, Open Space Area (1st floor) *or* online via [Big Blue Button](#).