P217 / Natural language processing and large language models in policy research

Topic : Methods & Research Design

Chair : Nihit Goyal (Delft University of Technology)

Second Chair : Nayara Albrecht (University of Amsterdam)

Third Chair : Sebastian Sewerin (Swiss Federal Institute of Technology, Zurich)

GENERAL OBJECTIVES, RESEARCH QUESTIONS AND SCIENTIFIC RELEVANCE

The growing availability of digitised policy texts, combined with rapid advances in natural language processing (NLP), has opened new avenues for public policy research. These texts include executive orders, legislation, parliamentary debates, public consultation submissions, evaluation reports, media coverage, and social media. Further, large language models (LLMs) are increasingly being used for information extraction, summarisation, and annotation, alongside dictionaries, supervised techniques, and unsupervised learning.

These tools hold promise across three connected areas of the field. For policy process research, they enable large-scale measurement of attention, frames, narratives, coalitions, and diffusion. For policy design research, they make it possible to identify and compare instruments, instrument mixes, and design features across thousands of policies. For policy evaluation, they can help synthesise evaluation reports and analyse stakeholder and citizen feedback. Tasks that once required several annotators can now be attempted at scale.

However, the application of these techniques also faces challenges, such as distinct vocabularies, complex language, and information sparsity. Further, the use of proprietary models with stochastic outputs can create additional issues, and ethical questions about bias, costs, and data protection need to be considered.

Existing scholarship reflects this mix of promise and caution. Research on text as data has shown that automated methods can augment, but not replace, careful reading, and that their outputs must be validated against the research question at hand. More recent studies suggest that LLMs can perform well on many (simpler) annotation tasks, yet their performance varies across prompts, model versions, languages, and domains. There is also the danger of dataset generation being driven by what LLMs currently are good at (e.g. coding high-level policy design characteristics such as instrument use) rather than what would be needed from a conceptual or theoretical point of view (e.g. a comprehensive assessment of policy design characteristics).

The panel’s main research question is: under what conditions can NLP, including LLM-based approaches, produce reliable, valid, and theoretically relevant insights into policy processes, policy designs, and policy evaluation? We aim to bring together methodological, empirical, and critical contributions that facilitate the creation of a community, reusable outputs, and shared infrastructure for public policy research.

CALL FOR PAPERS

We invite papers that engage theoretically, methodologically, or empirically with the use of NLP in policy process, policy design, or policy evaluation research. Relevant methods range from dictionaries and topic models to transformer-based classifiers and large language models. Papers may address any policy domain or political setting, and may work with texts such as legislation, regulations, parliamentary debates, consultation submissions, evaluation reports, or media and social media content.

Questions of interest include, but are not limited to:
- How can NLP and LLMs measure core policy process concepts such as actors, events, institutions, and policy change?
- How can they identify policy instruments, instrument mixes, and design features to help compare designs across jurisdictions or over time?
- How can they support policy evaluation, for example by synthesising evaluation reports and evidence, analysing stakeholder or citizen feedback, or tracing and attributing policy impacts?
- How do pipeline setup, model choice, prompting strategies, and fine-tuning affect substantive inferences?

- What challenges arise with specialised vocabularies, long or complex documents, and multilingual corpora?
- What are the ethical, legal, and environmental implications of applying NLP to policy text?

We welcome empirical applications, methodological innovations, conceptual contributions, and critical reflections. We especially encourage submissions that compare methods on the same corpus, share validated codebooks or datasets, analyze understudied contexts (particularly in the Global South and non-English settings), and report negative or mixed results clarifying where NLP falls short. Papers that bridge two or more of the three areas, for example by linking design features to evaluation findings, are also welcome.

Paper proposals should state the research question, the corpus and methods used, and how the paper contributes to the panel's broader aim of building shared standards, reusable outputs, and a community of scholars working on NLP in policy research. We welcome submissions from doctoral students, early career researchers, and established scholars alike.