

Large Language Models and Generative AI for Conflict Research

Organizers

Vito D'Orazio
West Virginia University
vito.dorazio@mail.wvu.edu

Javier Osorio
University of Arizona
josorio1@arizona.edu

Patrick Brandt
University of Texas at Dallas
pbrandt@utdallas.edu

Latifur Khan
University of Texas at Dallas (Computer Science)
lkhan@utdallas.edu

Our team has a record of generating advanced cyberinfrastructure tailored for political science, with a particular focus on conflict (Brandt et al. 2025, Osorio et al. 2025, Li et al. 2024, Osorio et al. 2024, Hu et al. 2024, Alsarra et al. 2024). With the support of various National Science Foundation (NSF) grants, our research team developed ConflIBERT, a domain specific large language model (LLM) specialized on political conflict, armed violence, and international relations generally. Initially developed to process text in English, our team advanced ConflIBERT's capabilities to process text in Spanish and Arabic, thus providing tools that foster more inclusive and connected academic communities around the world (Hu et al. 2022, Alsarra et al. 2023, Yang et al. 2023). We have been applying this cyberinfrastructure to advance research in political science topics such as political violence, terrorism, organized criminal violence, congressional hearings, political interactions in the UN Security Council, among others.

In addition to making our cyberinfrastructure free and publicly available, we are developing a set of training materials with the support of NSF to help the broader political science community learn how to use these computational tools for their own research projects. Our collaborations with various research groups allow us to bring together an increasingly numerous group of graduate students and early career faculty interested in applying advanced computational tools to political science research. Thanks to the support of our current NSF grant, we would be able to extend the activities of this workshop beyond the initial auspices of the Peace Science Meeting. More information about our work is available at the links below:

Project: <https://eventdata.utdallas.edu/>

Github: <https://github.com/eventdata/conflibert>

HuggingFace: <https://huggingface.co/eventdata-utd>

Focus and Format

Focus

The rise of massive volumes of electronic information paired with big data analytics, machine learning, and natural language processing paved the way for the ongoing revolution using Large Language Models (LLMs) and Artificial Intelligence (AI) to transform the way scientists conduct research. Conflict scholars, like many others, are just beginning to harness the potential that LLMs and AI offer to advance research. At the same time, these computer science methods are rapidly evolving. As a field of research, it is critical that we adopt and contribute to the development of these computational tools for both research and education.

Integrating the latest LLM and AI developments enables Peace Science scholars to expand their research frontier, develop new domain-specific ontologies, generate new databases by processing large amounts of unstructured electronic documents, analyze massive volumes of data to gain deeper insights into political behavior, and increase our understanding of complex systems. Integrating these innovative methodologies in peace science will advance the quality and scope of research in this field, foster multi-disciplinary collaborations, and increase our ability to inform the policy sector.

Much of the attention that LLMs have received is due to their abilities to generate text. However, their research potential also lies in their abilities to extract and analyze text. Since a lot of our conflict data and Peace Science research comes from text sources, LLMs provide us with enormous opportunities to move the field forward. They provide advantages over traditional tools for analyzing text in terms of performance and cost, as well as new opportunities beyond what previously existed.

LLMs and AI are quickly becoming an essential part of research pipelines and it is important that faculty and students learn to use these tools to conduct their research. Furthermore, these technologies are also becoming an important part of our classrooms. Beyond research, it is important for faculty and students to become familiar with the strengths and limitations of different models and methods. Beyond academia, these are also in-demand skills for graduate and undergraduate students inclined to join the government sector or industry.

Format

We propose a full day workshop consisting of two sections. The morning session will include an introduction to LLMs and AI for conflict research, a technical presentation on how these models work, and a hands-on technical tutorial on how to develop and use them. The introduction will provide a meta-analysis, integrating our team's research and that of others, and discussing opportunities and issues for conflict research. The technical presentation will cover the theory of LLMs and focus on discussing questions like the differences in extractive and generative models. The tutorial will guide participants through the development and application of our ConflIBERT model for different tasks including binary document classification and named entity recognition. It will use Colab, so will be free for all participants. The morning session will be led by the workshop organizers.

The afternoon session will include research presentations by workshop participants. While some participants have noted they wish to attend and not present, enough have indicated they would like to present for us to fill up an afternoon session with research talks and working paper discussion. Topics include conflict data projects and potential applications of LLMs to these efforts, experiments with the performance of different models for different data collection tasks, and more standard research where generative AI has been incorporated into a research design to answer a research question. We will have a group of researchers with expertise in this area who will provide feedback and offer guidance.

Workshop Participants

The team of organizers recently led the APSA Virtual Research Group on “Advancing the Use of Computational Tools in Political Science” and attracted the attendance of about 40 participants. Given the research profile of the organizers, most participants conduct research on topics related to political conflict and peace. This first workshop became the seminal effort to build a broader community of researchers and students using LLMs and AI for political science research. We are confident that enabling a PSSI workshop on “Large Language Models and Generative AI for Conflict Research” will allow this research community to grow by attracting more participants, thus increasing the critical mass of researchers using computational tools to advance conflict research.

We have extended a personalized invitation to over 40 researchers, inviting them to participate in this workshop, offering them the opportunity to present, and encouraging them to invite their graduate students. The table below lists everyone who responded ‘yes’ to our invitation to attend.

Participants		
Damian Boldt	FSU / Univ. of Virginia	dboldt@fsu.edu
Alex Braithwaite	Univ. of Arizona	abraith@arizona.edu
Jeff Carter	Appalachian State	carterjt2@appstate.edu
Chad Clay	Univ. of Georgia	kcclay@uga.edu
Maya Dalton	Penn State	mad6821@psu.edu
Cassy Dorff	Vanderbilt	cassy.dorff@vanderbilt.edu
Chris Fariss	Univ. of Michigan	cjf0006@gmail.com
Madeline Fleishman	Maryland / Univ. of Virginia	mfleish@umd.edu
Kevin Galambos	Rhodes College	galambosk@rhodes.edu
Michael Kenwick	Rutgers	michael.kenwick@rutgers.edu

Steven Landis	Notre Dame	slandis@nd.edu
Valerie Li	Penn State	valerie.li@psu.edu
Caleb Lucas	Indiana Univ.	lucascaj@iu.edu
Angela Nichols	Florida Atlantic	nicholsa@fau.edu
Simon Polichinel von der Maase	PRIO	simmaa@prio.org
Ben Radford	UNC-Charlotte / Pitt	benjamin.radford@charlotte.edu
Cecilia Cavero Sanchez	Penn State	ccaveros@psu.edu
Emmanuel Teitelbaum	George Washington	ejt@gwu.edu
Scott Wolford	UT - Austin	swolford@austin.utexas.edu
Thorin Wright	Arizona State	Thorin.Wright@asu.edu

Anticipated Products and Output

The goal of this workshop is to generate a scientific community to advance the adoption, adaptation, and application of LLMs and AI in conflict research. We will share research, insights, and best practices into these critical methodological tools. The workshop will encourage networking, foster collaborations for future research, and facilitate discussion on recent trends and challenges related to integrating these tools for conflict research. The organizing team will then rely on its current NSF grant to cultivate and sustain this research group beyond the auspices of the Peace Science meeting by fostering research collaborations, providing technical training, and pursuing a special issue.

References

2025. "ConflIBERT: A Language Model for Political Conflict." Patrick T. Brandt, Sultan Alsarra, Vito D'Orazio, Dagmar Heintze, Latifur Khan, Shreyas Meher, Javier Osorio, Marcus Sianan. Revise and resubmit. <https://arxiv.org/abs/2412.15060>

2025. "The Devil is in the Details: Assessing the Effects of Machine-Translation on LLM Performance in Domain-Specific Texts." Javier Osorio, Afraa Alshammari, Naif Alatrush, Dagmar Heintze, Amber Converse, Sultan Alsarra, Latifur Khan, Patrick T. Brandt and Vito D'Orazio. *20th Machine Translation Summit (MT Summit 2025)*, forthcoming.

2024. "ConflLPC: Logits and Parameter Calibration for Political Conflict Analysis in Continual Learning." Xiaodi Li, Niamat Zawad, Patrick T. Brandt, Javier Osorio, Vito D'Orazio, Latifur Khan. *2024 IEEE International Conference on Big Data (BigData)*, pp 6320-6329.

2024. "Keep it Local: Comparing Domain-Specific LLMs in Native and Machine Translated Text using Parallel Corpora on Political Conflict." Javier Osorio, Sultan Alsarra, Amber Converse, Afraa Alshammari, Dagmar Heintze, Naif Alatrush, Latifur Khan, Patrick T. Brandt, Vito D'Orazio, Niamat Zawad, Mahrusa Billah. *Proceedings of the International Conference on Foundation and Large Language Models (FLLM2024)*, forthcoming.

2024. "Leveraging Codebook Knowledge with NLI and ChatGPT for Zero-Shot Political Relation Classification." Yibo Hu, Erick Skorupa Parolin, Latifur Khan, Patrick Brandt, Javier Osorio, and Vito D'Orazio. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp 583-603.

2024. "Extractive Question Answering for Spanish and Arabic Political Text." Sultan Alsarra, Parker Whitehead, Naif Alatrush, Luay Abdeljaber, Latifur Khan, Javier Osorio, Patrick Brandt and Vito D'Orazio. *2024 International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, pp 144-153.

2023. "ConfliBERT-Spanish: A Pre-Trained Spanish Language Model for Political Conflict and Violence." Wooseong Yang, Sultan Alsarra, Luay Abdeljaber, Niamat Zawad, Zeinab Delaram, Javier Osorio, Latifur Khan, Patrick T. Brandt, Vito D'Orazio. *7th IEEE Congress on Information Science and Technology (CiSt)*, pp 287-292.

2023. "ConfliBERT-Arabic: A Pre-trained Arabic Language Model for Politics, Conflicts and Violence." Sultan Alsarra, Luay Abdeljaber, Wooseong Yang, Niamat Zawad, Latifur Khan, Patrick Brandt, Javier Osorio, Vito D'Orazio. *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing (RANLP)*, pp 99-108.

2022. "ConfliBERT: A Pre-trained Language Model for Political Conflict and Violence." Yibo Hu, Mohammad Saleh Hosseini, Erick Skorupa Parolin, Javier Osorio, Latifur Khan, Patrick Brandt, Vito D'Orazio. *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2022)*, pp 5469-5482.