PROMPT: A Multilingual, LLM-powered pipeline for Narrative Trajectories

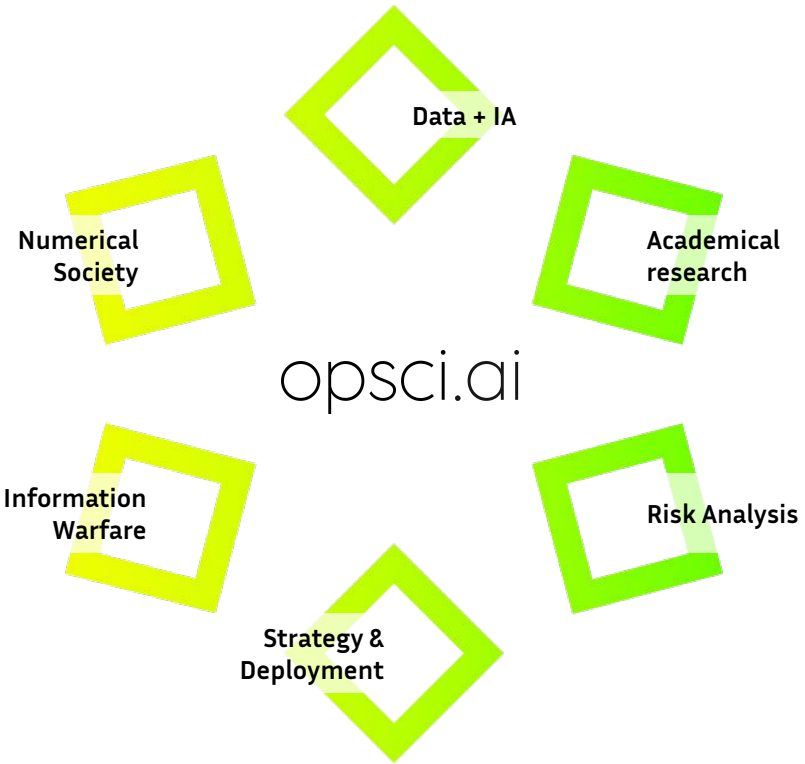
A word about Opsci

A private-sector research lab for LLMs

- **Open LLM France:** training of an open-{weight, science, data} LLM for generala uses, with emphasis on educational prospects

- **Audit LLM**: auditing performances, bias and regulatory frameworks for LLMs

- **PROMPT:** see following slides.



Disinformation detection: an overview of the european landscape.

Disinformation in Europe is a widely investigating subject :

- Vera.ai
- AI4Trust
- AI4All
- GLOWIN
- European Fact-Checking Standards Network
- FANDANGO
- Hoaxbusters
- AFP Factual
- EDMO (European Digital Media Observatory)

The plan is **not** to do “yet another project”.

PROMPT: Predictive Research On Misinformation & Propagation Trajectories

Premises:

- **What:** A tool with metrics and explanations for narrative trajectories across social media platforms (X, Facebook, Tiktok, Bluesky?, Wikipedia,...)
- **For who:** Fact-checkers and journalists
- **How:** Comparison with known fake news, analysis of evolutions (e.g. early accelerations and coordination markers), topic modeling, cross-platform analysis...
- **On which subjects:** War between Ukraine & Russia, LGBTQIA+ rights, EU Elections
- **Specifics:** emphasis on social media (short texts with diluted context) & 6 different countries (France, Italy, Romania, Lithuania, Latvia, Estonia) + English / Russian

opsci.ai

1506 UNIVERSITÀ DEGLI STUDI DI URBINO CARLO BO

HUN REN CENTRE FOR SOCIAL SCIENCES

ROMÂNIA

ORIZZONTI POLITICI

RĪGA STRADIŅŠ UNIVERSITY

WIKIMEDIA FRANCE

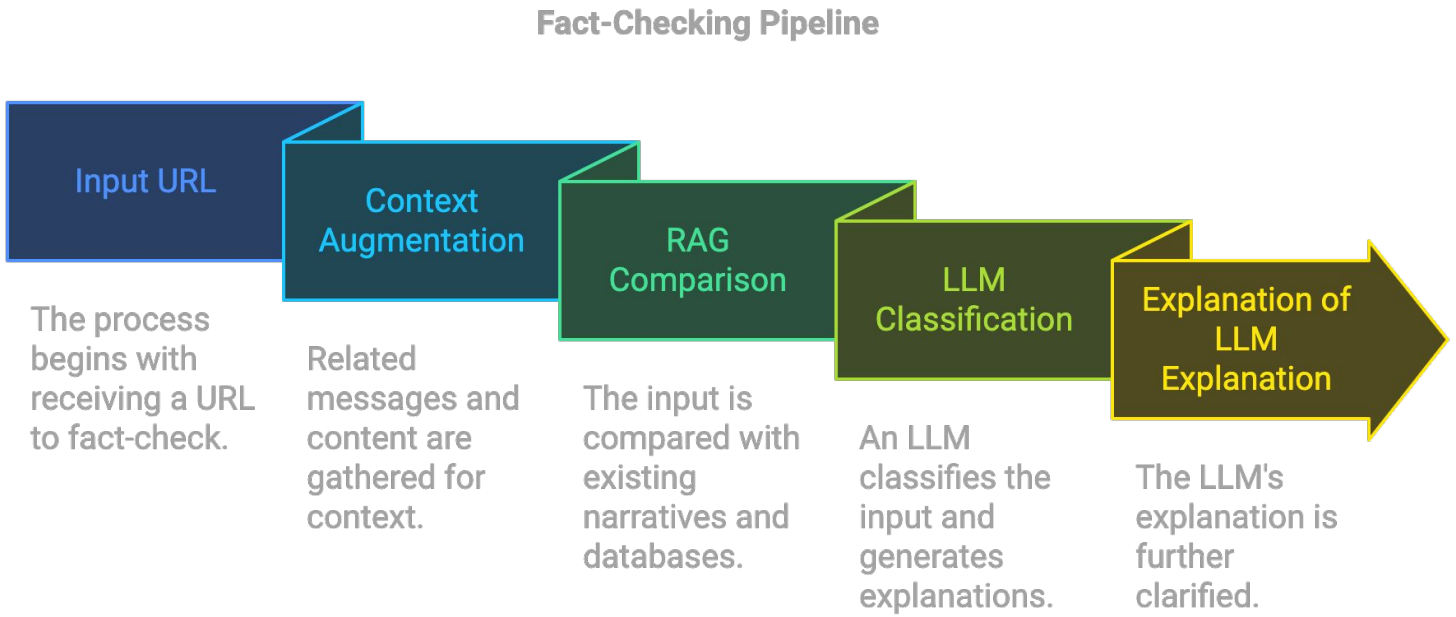
LES SURLIGNEURS

Erich-Brost-Institut für internationalen Journalismus

re: baltica

Pipeline keypoints:

- **RAG:** embeddings (fine-tuned), knowledge graphs, additional metrics ...
- **Explanations**: first and second level (influence functions, mechanistic interpretability), Mixture of Experts (MoE), example of “souk”.



Semantico-axiological matrix: (in collaboration with Riga Stradiņš University)

- Pre-existing works: DISARM, ABCDE, "Decoding Antisemitism" Lexicon...
- Go from **general rhetorical devices** to **detailed markers** of disinformation
- Discussions with journalists & political science researchers for **common ground**

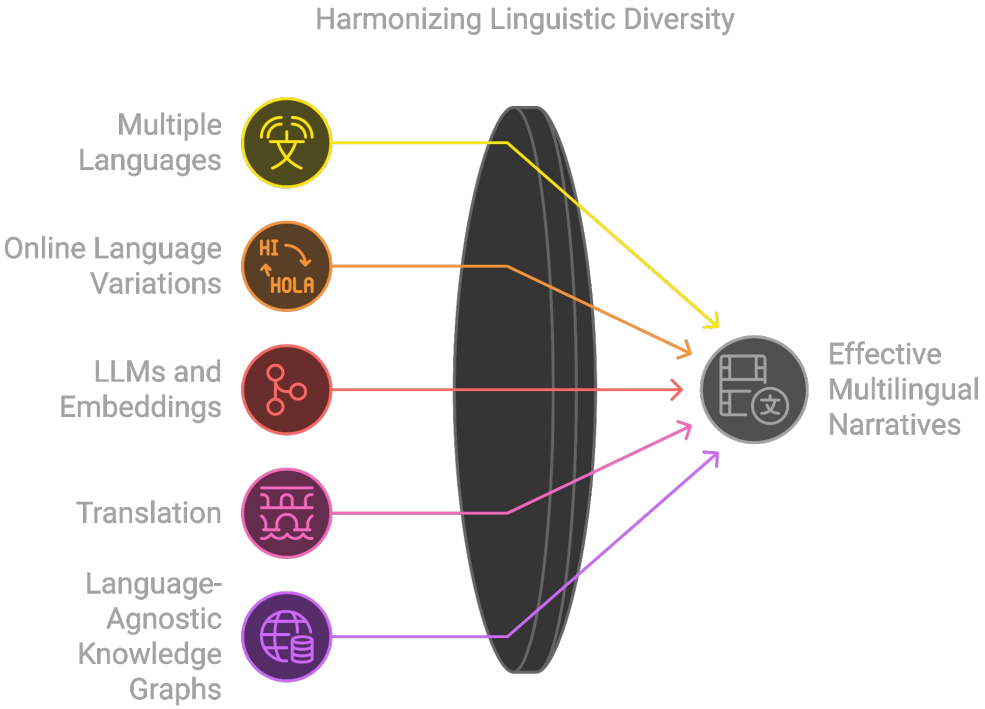
	A	B	C	D	E	G	H	I	J	L	M
1			WHAT								
2	Date	Narrative	Link	Topic	Subtopic	Disinfo typology	Facticity / truthfulness	Verifiability	Format	Platform	Geographical S
3	Legend: Green = closed list of items, dropdown menu Yellow = open field	Short description of the object of dis/misinformation		Big issue	Specific subject of disinfo/misinfo					Social network or media type	Typology
4	July 2024	Brigitte Macron is a man	https://x.c	LGBTQIA+	Representative's personal life	Conspiracy theory	FALSE	no	Video + short text + AI generated audio	Youtube; X	national

(Subset of) Upcoming challenges

Multi-lingual narratives:

- Most European-centric narratives are in **multiple languages**.
- The language used **online** is not the same as the language commonly used -> classical translation is not as effective.

- LLMs and embeddings become **more efficient** in this regard, but translation can sometimes be useful (e.g. for code-switching).
- Knowledge Graph are **language-agnostic** -> but issue of transliteration, need to remove the multiplicity of entities.



```
According to the Actantial Model by Greimas with the actant label set ["Sender", "
    Receiver", "Subject", "Object", "Helper", "Opponent"], the actants are defined as
    follows:

* Subject: The character who carries out the action and desires the Object.
* Object: The character or thing that is desired.
* Sender: The character who initiates the action and communicates the Object.
* Receiver: The character who receives the action or the Object.
* Helper: The character who assists the Subject in achieving its goal.
* Opponent: The character who opposes the Subject in achieving its goal.

Based on this Actantial Model and the actant label set, please recognize the actants
    in the given article.

Article: {{ article }}

Question: What are the main actants in the text? Provide the answer in the following
    JSON format: {"Actant Label": ["Actant Name"]}. If there is no corresponding
    actant, return the following empty list: {"Actant Label": []}.

Answer:
```

J. Elfes, arxiv: 2409.06540

Multi-modal narratives:

- Usage of social media involves **more than just text**.
- Text-image **relationships**: repetition, complementarity, illustration, commentary...

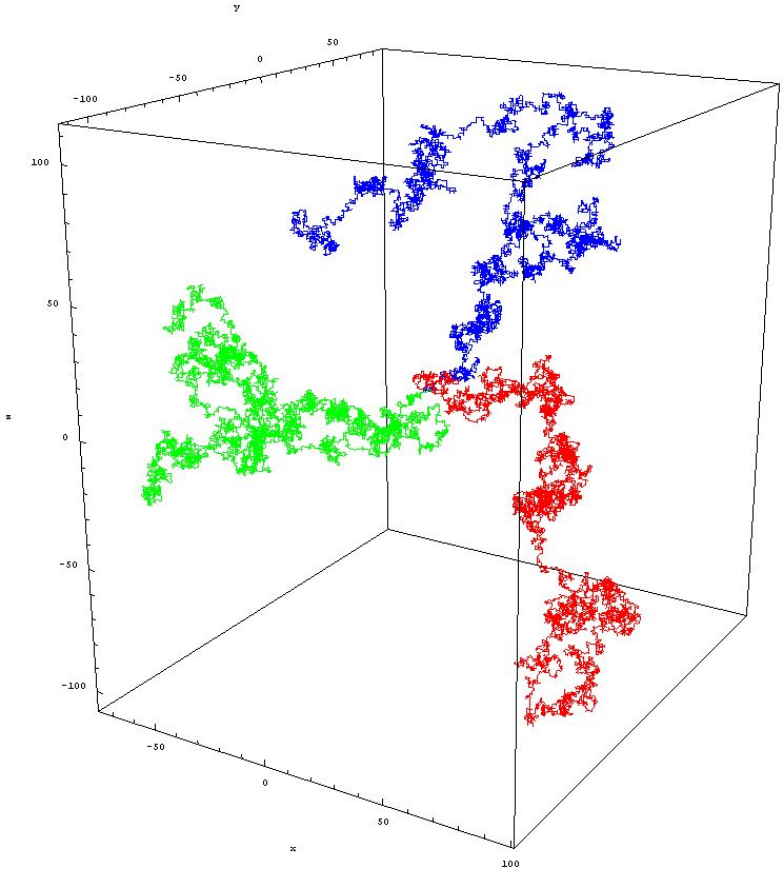
- Some solutions: embeddings (concatenation, bge-m3), llms (Mistral Large), knowledge graphs (Actant properties)...

Cross-platform narratives

- Linked with multi-lingual narratives, platforms can have their own language.
- Temporality and diffusion into the Western sphere (e.g. “ideological visa”)
- Identifiability of communities and actors -> Identify sub-graphs in order to cross-reference actors?

To go even further:

- Dynamic & longitudinal analysis of narratives: decomposition of narratives into “frames and concepts” (@Michal)
- Coordination -> So far, mostly time-based with the 60s rule or based on membership to some lists.
- “Effort of postage” for medium-sized actors based on stochastic processes (e.g. influencers)?
- Fine tuning of embeddings of narratives according to topic and platforms.



Annexes

Related works:

- Chirkova, Nadezhda, David Rau, Hervé Déjean, Thibault Formal, Stéphane Clinchant, and Vassilina Nikoulina. "**Retrieval-Augmented Generation in Multilingual Settings.**" arXiv, July 1, 2024. http://arxiv.org/abs/2407.01463.
- Feng, Zhangchi, Dongdong Kuang, Zhongyuan Wang, Zhijie Nie, Yaowei Zheng, and Richong Zhang. "**EasyRAG: Efficient Retrieval-Augmented Generation Framework for Automated Network Operations.**" arXiv, October 15, 2024. https://doi.org/10.48550/arXiv.2410.10315.
- Elfes, Jan. "**Mapping News Narratives Using LLMs and Narrative-Structured Text Embeddings.**" arXiv, September 10, 2024. http://arxiv.org/abs/2409.06540.
- Liu, Bo, Li-Ming Zhan, Zexin Lu, Yujie Feng, Lei Xue, and Xiao-Ming Wu. "**How Good Are LLMs at Out-of-Distribution Detection?**," n.d.
- Kusupati, Aditya, Gantavya Bhatt, Aniket Rege, Matthew Wallingford, Aditya Sinha, Vivek Ramanujan, William Howard-Snyder, et al. "**Matryoshka Representation Learning.**" arXiv, February 8, 2024. http://arxiv.org/abs/2205.13147.
- Giglietto, Fabio, Nicola Righetti, Luca Rossi, and Giada Marino. "**Coordinated Link Sharing Behavior as a Signal to Surface Sources of Problematic Information on Facebook.**" In International Conference on Social Media and Society, 85–91. Toronto ON Canada: ACM, 2020. https://doi.org/10.1145/3400806.3400817.
- "**Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet**"

Thank you for your attention!

JUSTIN@OPSCI.AI
STRAT

JORDAN@OPSCI.AI
OPS

ERWAN@OPSCI.AI
DATA

CLEMENT@OPSCI.AI
IA