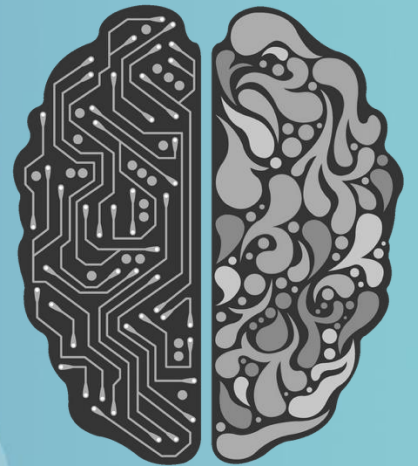


Natural Language Processing vs Large Language Models: *this is the end of the world as we know it, and I feel fine*

*FIRSA Webinar
16/06/2026*

*Bertrand DE LONGUEVILLE
Head of Text Mining and AI Competence Centre
European Commission's JRC*



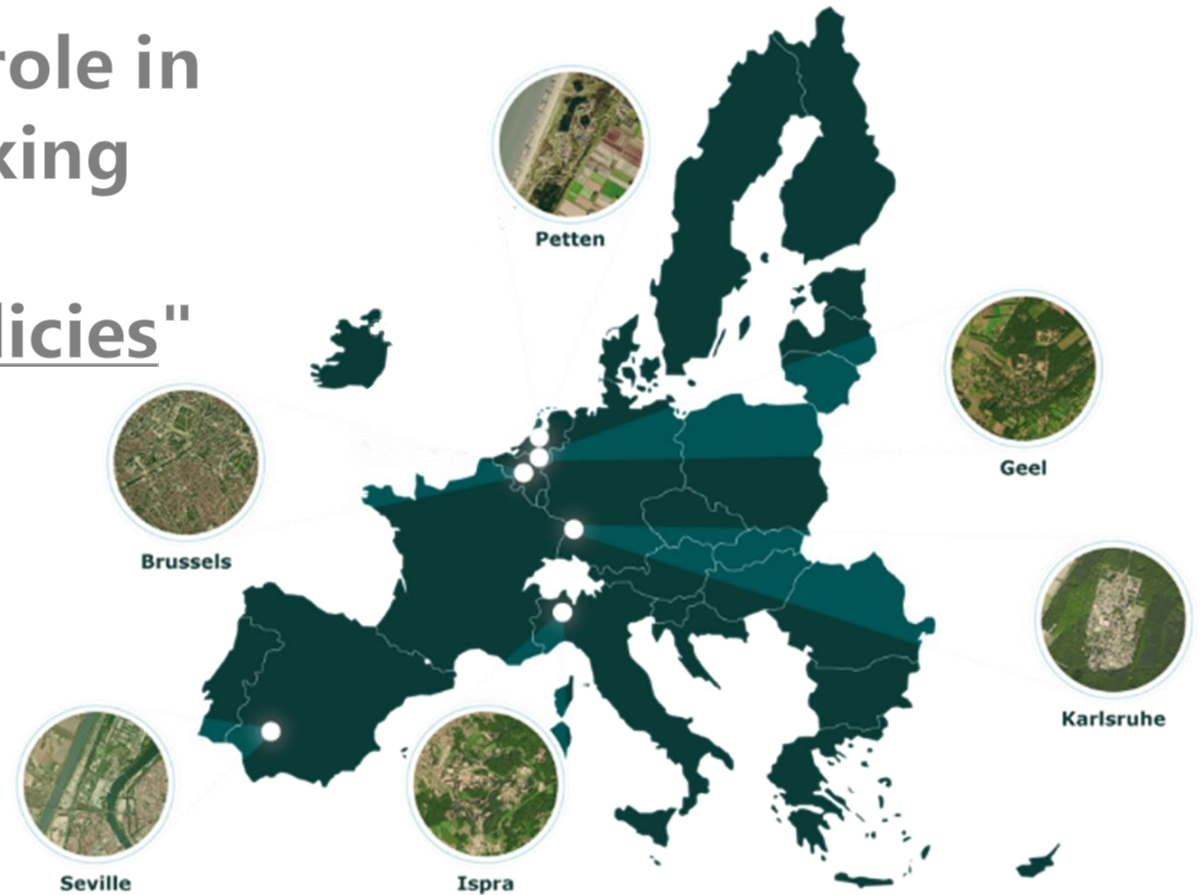
Joint Research Centre (JRC)

the European Commission's science and knowledge service

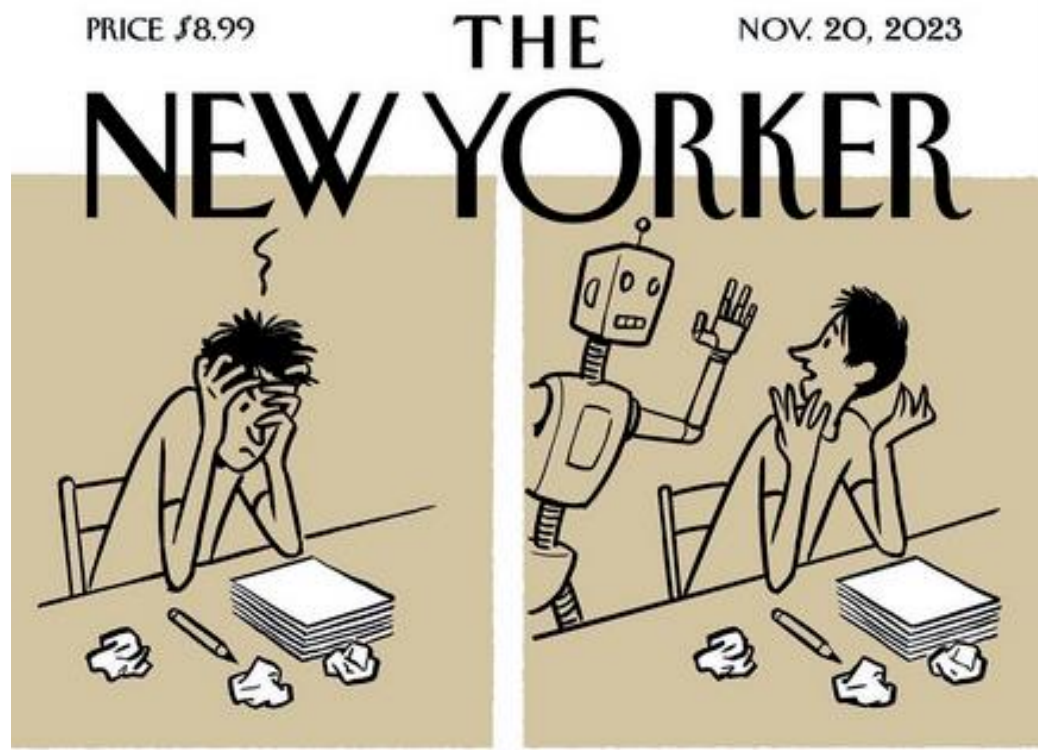
Mission : "To play a central role in creating, managing and making sense of collective scientific knowledge for better EU policies"



Image credit : undraw.co



Trustworthy AI



Search for Articles:

Title / Keyword

Author / Affiliation / Email

AI

All Article Types

Search

Journals / AI / Volume 6 / Issue 6 / 10.3390/ai6060128



Submit to this Journal

Review for this Journal

Propose a Special Issue

Article Menu

Academic Editor



Demos T. Tsahalidis

Subscribe SciFeed

Recommended Articles

Author Biographies

Related Info Link

More by Authors Links

Article Views

1396

IK

Order Article Reprints



Open Access Article

The Proof Is in the Eating: Lessons Learnt from One Year of Generative AI Adoption in a Science-for-Policy Organisation

by Bertrand De Longueville ^{1,*} , Ignacio Sanchez ² , Snezha Kazakova ¹ , Stefano Luoni ² , Fabrizio Zaro ² , Kalliopi Daskalaki ¹ and Marco Inchingolo ¹

¹ European Commission—Joint Research Centre, CDMA Building, 21 Rue du Champ de Mars/Marsveldstraat 21, B-1050 Brussels, Belgium

² European Commission—Joint Research Centre, Via Enrico Fermi 2749, I-21027 Ispra, VA, Italy

* Author to whom correspondence should be addressed.

AI **2025**, *6*(6), 128; <https://doi.org/10.3390/ai6060128>

Submission received: 15 April 2025 / Revised: 27 May 2025 / Accepted: 29 May 2025 /

Published: 17 June 2025

Download

Browse Figures

Versions Notes

Abstract

This paper presents the key results of a large-scale empirical study on the adoption of Generative AI (GenAI) by the Joint Research Centre (JRC), the European Commission's science-for-policy department. Since spring 2023, the JRC has developed and deployed GPT@JRC, a platform providing safe and compliant access to state-of-the-art Large Language Models for over 10,000 knowledge workers. While the literature highlighting the potential of GenAI to enhance productivity for knowledge-intensive tasks is abundant, there is a scarcity of empirical evidence on impactful use case types and success factors. This study addresses this gap and proposes the JRC GenAI Compass conceptual framework based on the lessons learnt from the JRC's GenAI adoption journey. It includes the concept of AI-IQ, which reflects the complexity of a given GenAI system. This paper thus draws on a case study of enterprise-scale AI implementation in European public institutions to provide approaches to harness GenAI's potential while mitigating the risks.

Keywords: Artificial Intelligence; Generative AI; Large Language Models; LLMs; AI governance;



Outline



WHAT – the significance of the “ChatGPT revolution” on NLP



SO WHAT – LLMs as “game changers” for NLP

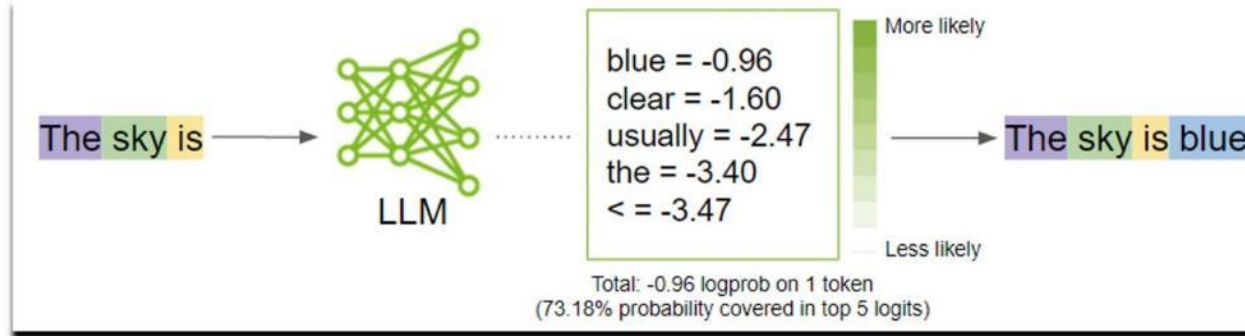


NOW WHAT – the new NLP that looks like the good old one

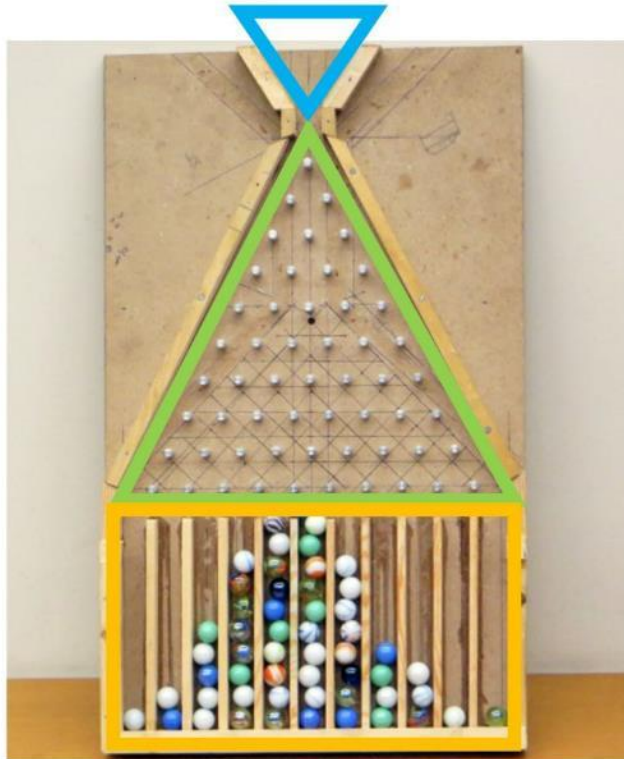
WHAT – the significance of the “ChatGPT revolution” on NLP

“non-deterministic” previsions

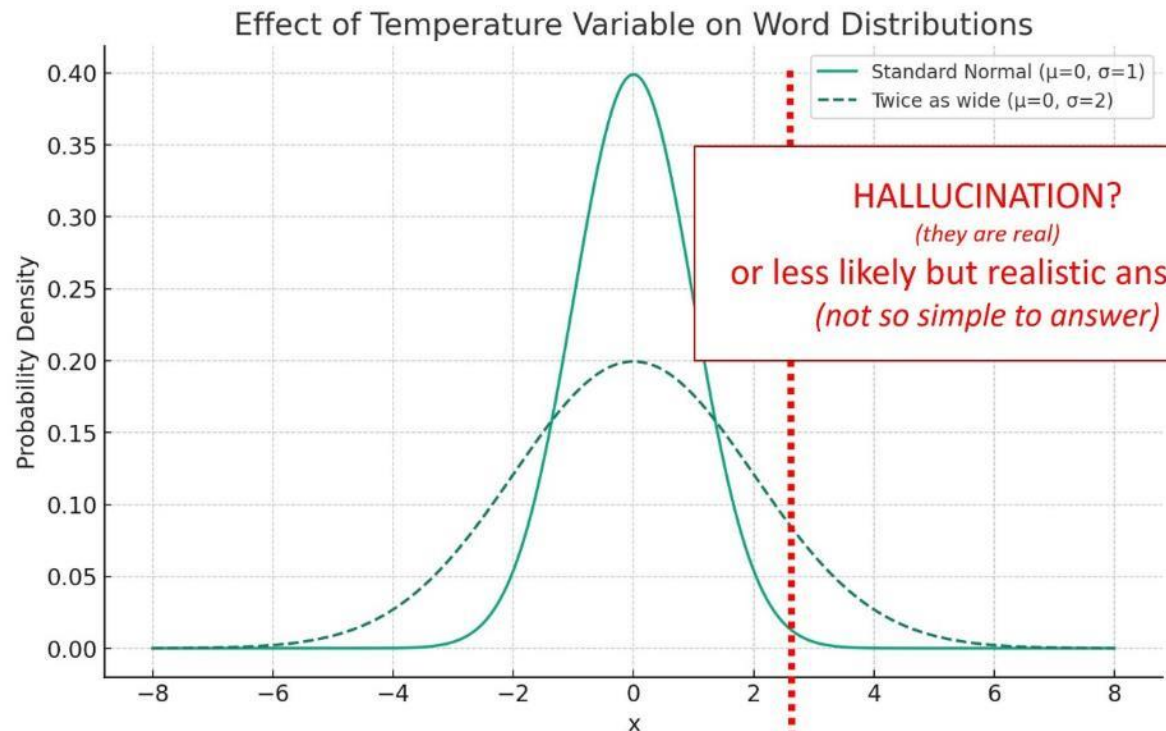
How LLMs work



prompt
LLM weights
results



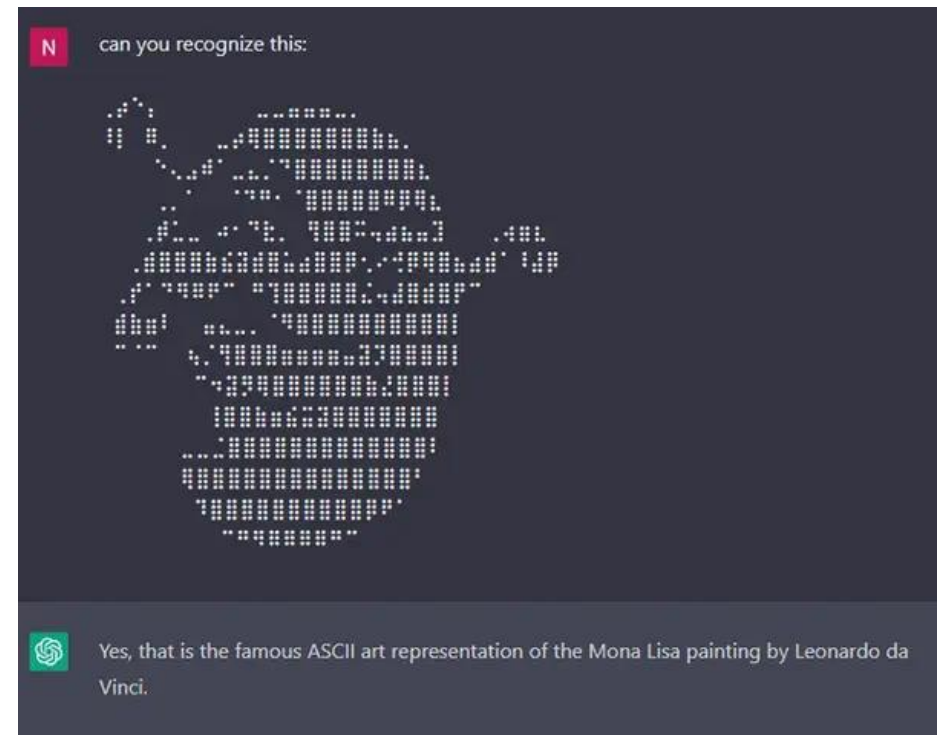
Galton Board



μ = alignment
= “education”

σ^2 =
temperature =
“character”

Stochastic parrots



<https://www.buzzfeed.com/andyneunschwander/chatgpt-ai-fails-funny>

Emily M. Bender et al., “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?,” in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT ’21 (New York, NY, USA: Association for Computing Machinery, 2021), 610–23, <https://doi.org/10.1145/3442188.3445922>

Loru, Edoardo, Jacopo Nudo, Niccolò Di Marco, Alessandro Santirocchi, Roberto Atzeni, Matteo Cinelli, Vincenzo Cestari, Clelia Rossi-Arnaud, and Walter Quattrocchi. 2025. “The Simulation of Judgment in LLMs.” Proceedings of the National Academy of Sciences 122(42). doi:10.1073/pnas.2518443122.



ACM DIGITAL LIBRARY Association for Computing Machinery

Browse About Sign in Register

Journals Magazines Proceedings Books SIGs Conferences People

FACCT

Home > Conferences > FACCT > Proceedings > FAccT '21 > On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

ARTICLE | OPEN ACCESS



On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

Authors: Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, Shrimai Prabhu

PNAS

FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency
<https://doi.org/10.1145/3442188.3445922>

Published: 01 March 2021 [Publication History](#)



3,068 467,909

The simulation of judgment in LLMs

Edoardo Loru, Jacopo Nudo, Niccolò Di Marco, and Walter Quattrocchi

Edited by Susan Fiske, Princeton University, Princeton, NJ; received July 10, 2025; accepted September 18, 2025

October 13, 2025 | 122 (42) e2518443122 | <https://doi.org/10.1073/pnas.2518443122>

1,611



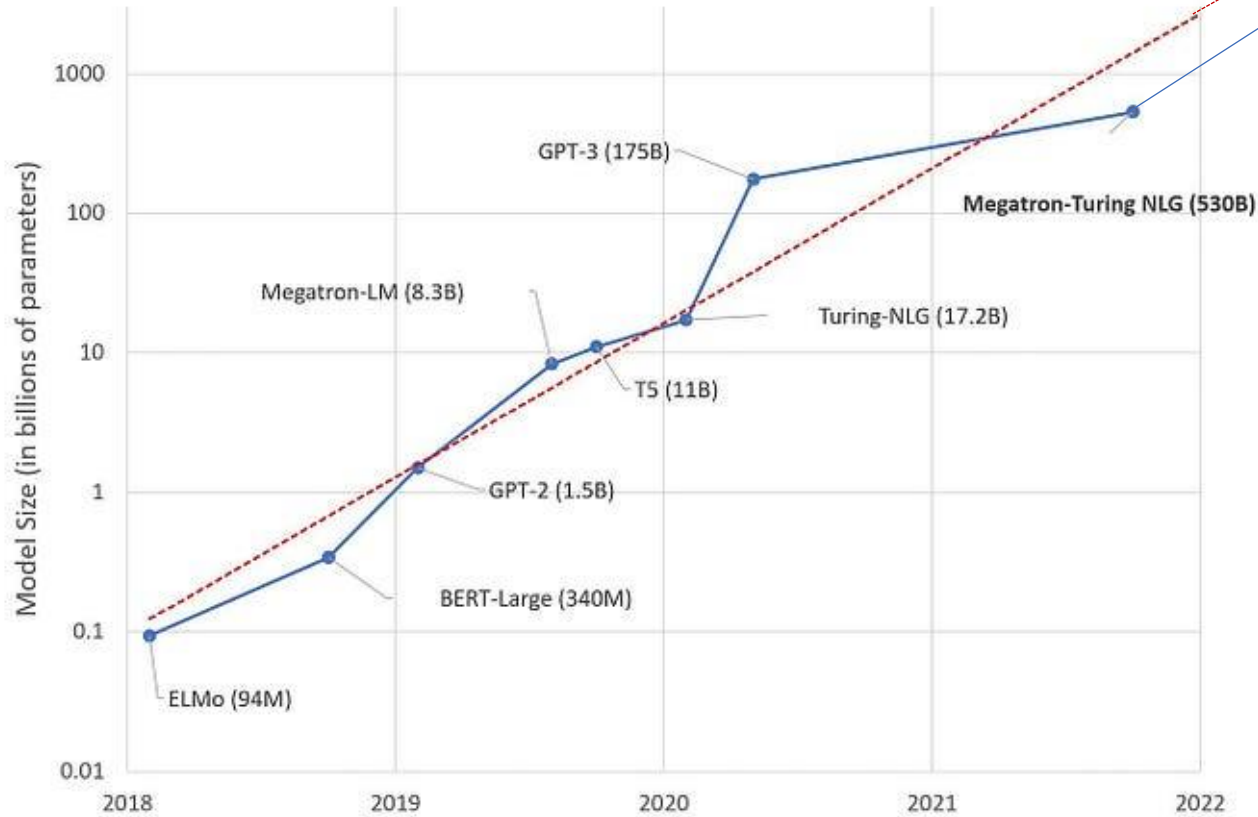
Significance

Large Language Models (LLMs) are used in evaluative tasks across domains. Yet, what appears as alignment with human or expert judgments may conceal a deeper shift in how “judgment” itself is operationalized. Using news outlets as a controlled benchmark, we compare six LLMs to expert ratings and human evaluations under an identical, structured framework. While models often match expert outputs, our results suggest that they may rely on lexical associations and statistical priors rather than contextual reasoning or normative criteria. We term this divergence epistemia: the illusion of knowledge emerging when surface plausibility replaces verification. Our findings suggest not only performance asymmetries but also a shift in the heuristics underlying evaluative processes, raising fundamental questions about delegating judgment to LLMs.

PDF

Help

Emergence



November 2024 : “Le Cun’s plateau” or Agentic age?

<https://nilkanthahire.medium.com/lms-are-basically-obsolete-why-ai-pioneer-yann-lecun-thinks-the-real-ai-revolution-hasnt-16112d9d0809>

Human Brain (86B neurons, 100T connections)



Image generated by Dall-E AI. Prompt : “a humanoid robot on block of concrete, sitting exactly in the position of the famous statue le penseur from Rodin” <https://labs.openai.com/>

Marteen Grootendorst, “Drawing Analogies between GPT and Human Psychology,” Towards Data Science (blog), June 29, 2023, <https://towardsdatascience.com/gpt-and-human-psychology-94a21ba6d20e>.

Celest Biever “ChatGPT broke the Turing test — the race is on for new ways to assess AI” Nature briefings 25 July 2023 <https://www.nature.com/articles/d41586-023-02361-7>

Wang, Xinglin, Peiwen Yuan, Shaoxiong Feng, Yiwei Li, Boyuan Pan, Heda Wang, Yao Hu, and Kan Li. 2024. ‘CogLM: Tracking Cognitive Development of Large Language Models’. <https://doi.org/10.48550/ARXIV.2408.09150>

A (somewhat) new era

“The text-processing systems used in event coding are still similar to ones developed more than 20 years ago” (Wang et al., 2016)

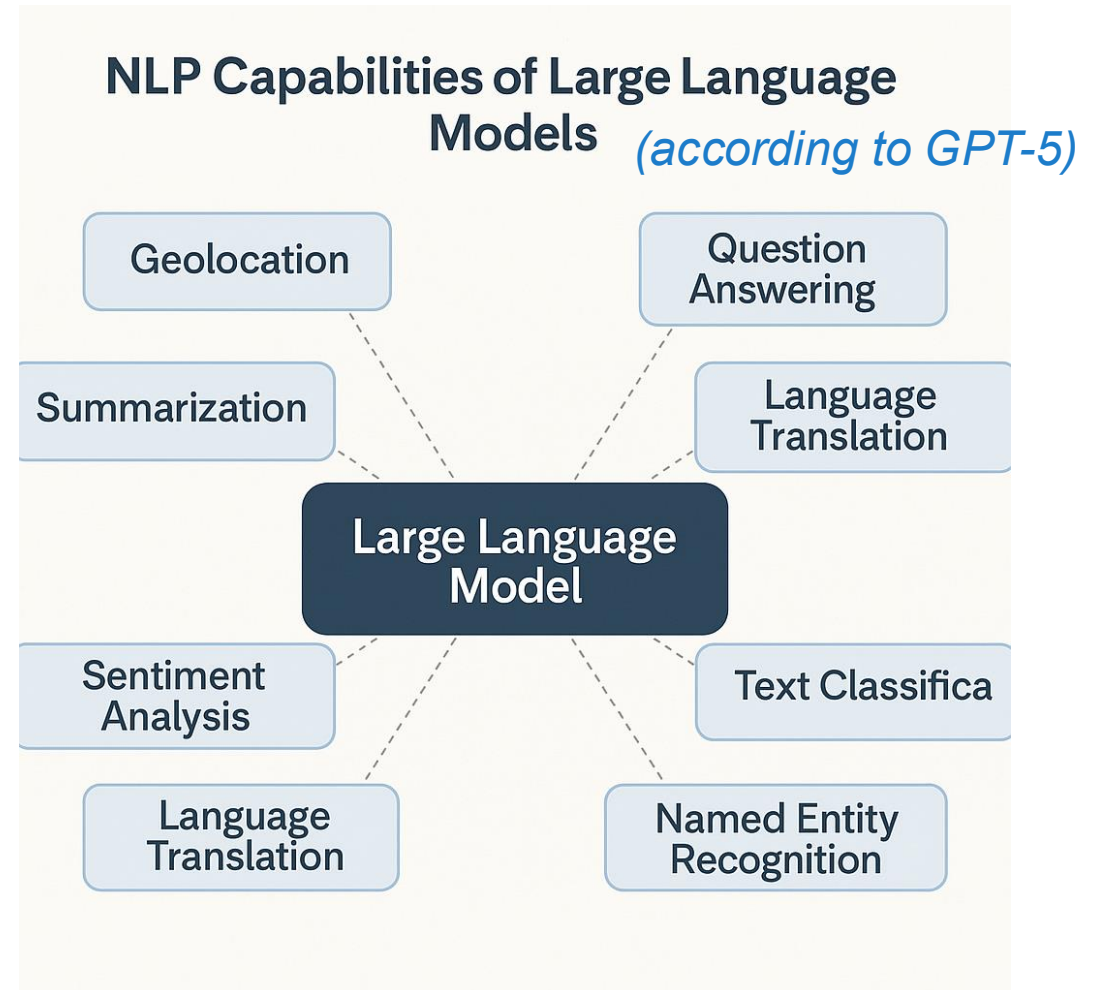
*“The text-processing systems used in event coding are **in no way** similar to ones developed less than 20 **month** ago” (**you**, 2026)*

Wang, W., Kennedy, R., Lazer, D., & Ramakrishnan, N. (2016). Growing pains for global monitoring of societal events. *Science*, 353(6307), 1502–1503.

SO WHAT – LLMs as “game changers” for NLP

The universal 0-shot classifier/predictor

“The answer is :
use an LLM.
But what was the
question, again?”



Versatility

BERT-like

vs

GPT5-like



Inference costs

BERT-like

vs

GPT5-like

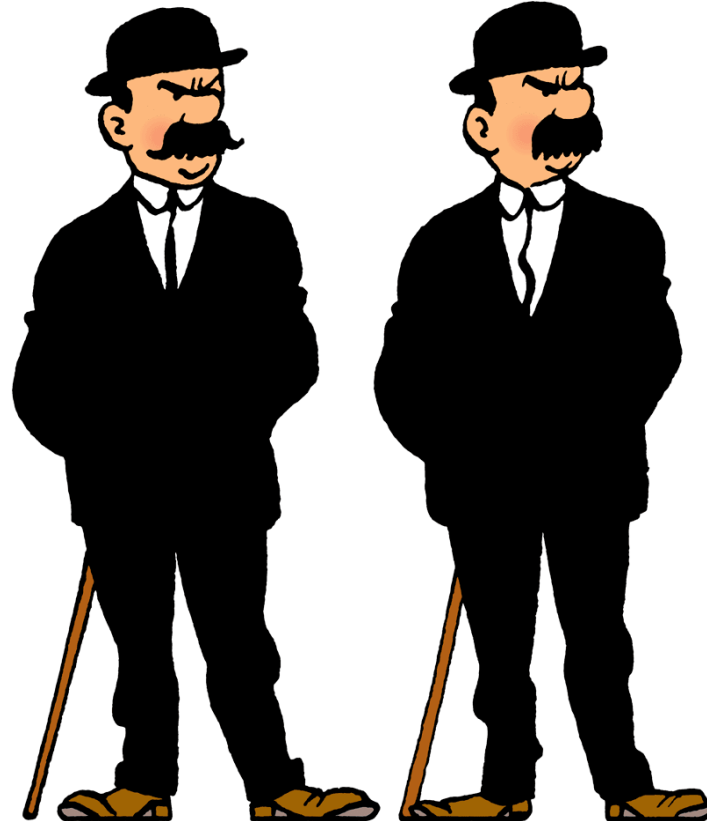


Need for evaluation

BERT-like

vs

GPT5-like



It's all about evaluation (there we go again)



Cartoon source: <https://andifugard.info/>

Table 4. Weighted F1 scores on UMSAB test set per language. For XLM-RLnews-8 we report also the weighted F1 scores per class.

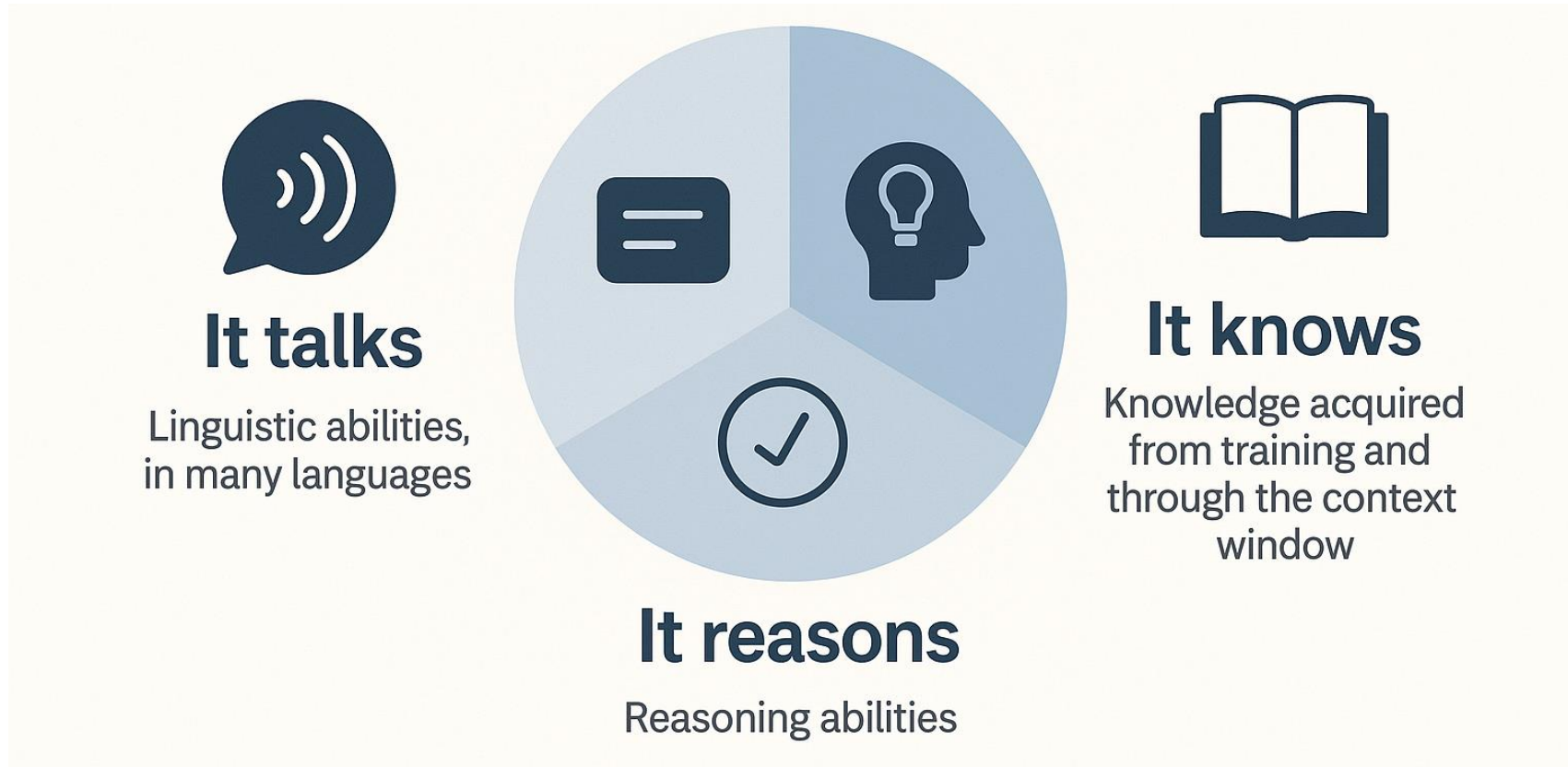
Language	XLM-T	XLM-RL	XLM-RLnews-8				Avg all models
			Avg	Neg	Neu	Pos	
AR	67.07	67.26	67.66	0.74	0.57	0.72	67.33
EN	73.00	70.45	72.24	0.80	0.60	0.77	71.90
FR	73.70	71.18	69.84	0.80	0.63	0.66	71.57
DE	75.14	77.23	78.45	0.78	0.75	0.82	76.94
HI	56.81	58.67	58.37	0.62	0.51	0.63	57.95
IT	69.06	74.57	74.37	0.74	0.72	0.77	72.67
PT	76.36	75.03	74.40	0.78	0.65	0.80	75.26
ES	69.32	70.23	69.54	0.74	0.56	0.78	69.70
All languages	70.06	70.58	70.61	0.75	0.62	0.74	70.41

An example among many, many, many evaluation tables present in NLP scientific papers

Di Nuovo, E., Cartier, E., & De Longueville, B. (2024). Meet XLM-RLnews-8: Not Just Another Sentiment Analysis Model. In *Natural Language Processing and Information Systems* (pp. 24–35). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-70242-6_3

NOW WHAT – the new
NLP science that looks
like the good old one

Untangle complex LLMs capabilities



Review tasks in your Event Extraction Pipeline

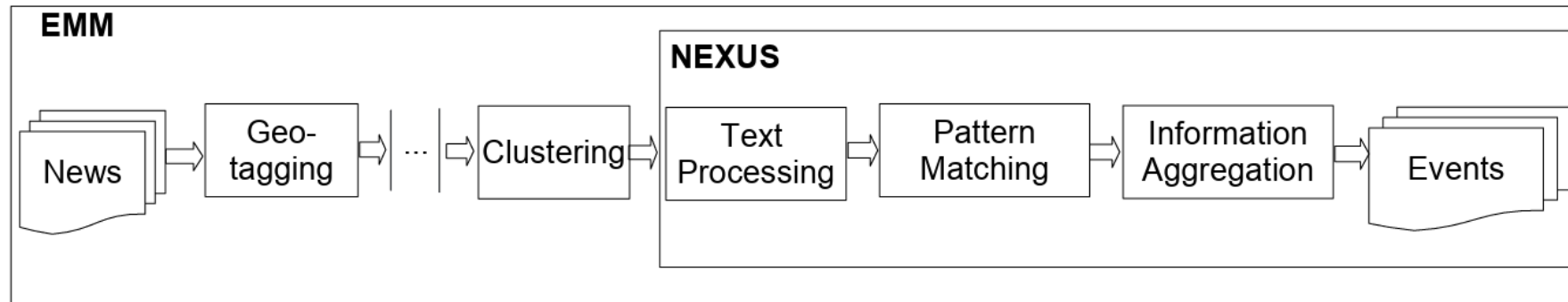
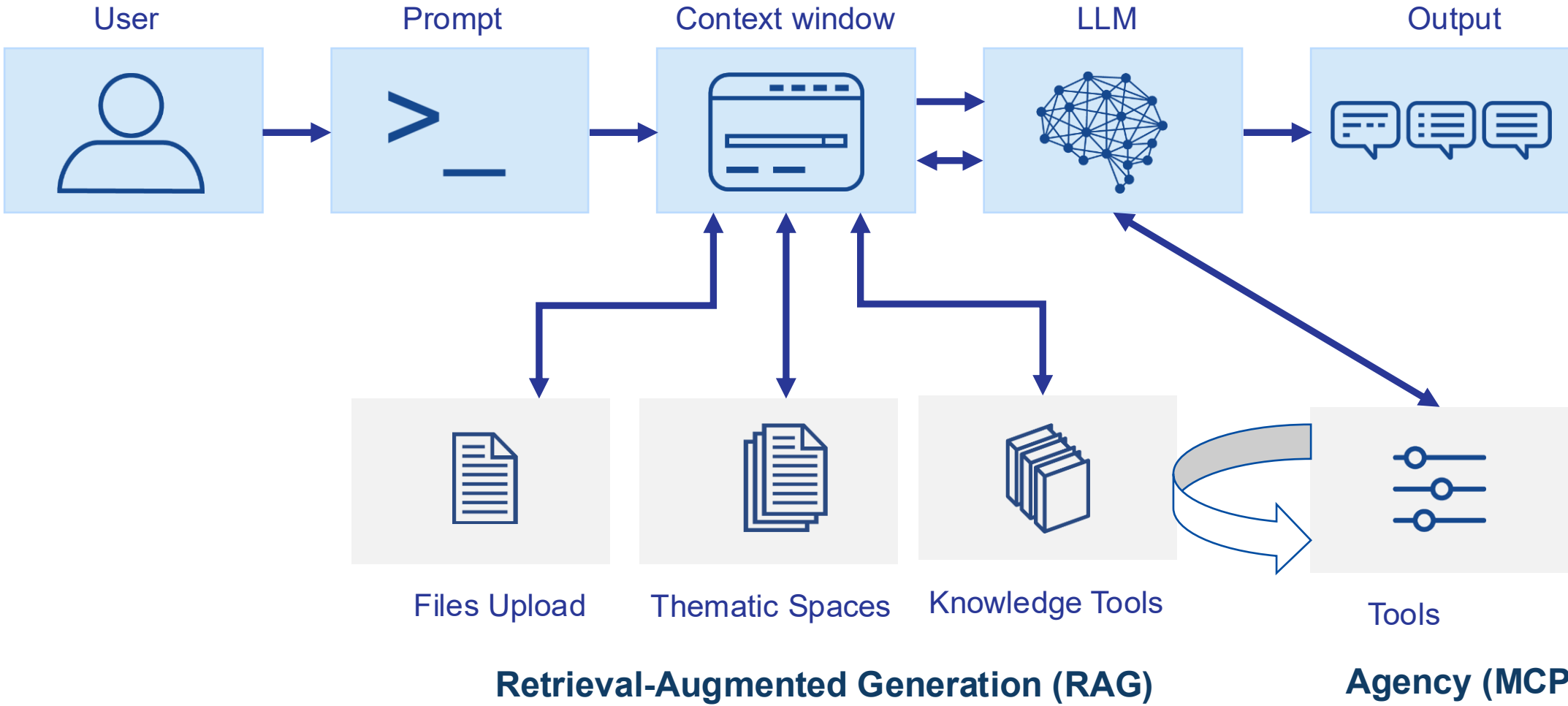


Fig. 1. Real-time event extraction processing chain

Tanev, H., Piskorski, J., Atkinson, M. (2008). Real-Time News Event Extraction for Global Crisis Monitoring. In: Kapetanios, E., Sugumaran, V., Spiliopoulou, M. (eds) Natural Language and Information Systems. NLDB 2008. Lecture Notes in Computer Science, vol 5039. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-69858-6_21

Design your LLM pipelines



And evaluate!

Gold standard

precision

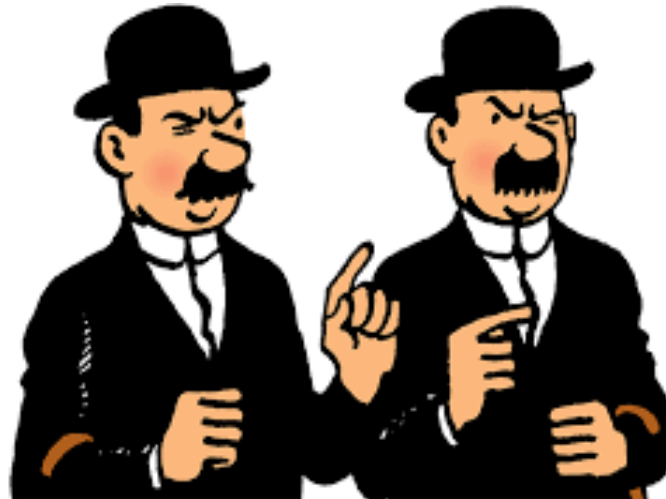
recall

F1-score

compare

again and again...

Annotations



Conclusion : it's the end of the world as we know it (and I feel fine)



Thank you and keep in touch

Bertrand DE LONGUEVILLE
Head of Text Mining and AI Competence Centre
European Commission's JRC
<https://www.linkedin.com/in/bertrand-de-longueville-78357316/>

BONUS TRACK : *It's The End Of The World As We Know It (And I Feel Fine)*
performed by R.E.M. <https://www.youtube.com/watch?v=Z0GFRcFm-aY>



© European Union, 2024

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.



EU Science Hub
[Joint-research-centre.ec.europa.eu](https://joint-research-centre.ec.europa.eu)