Striking the Balance in Using LLMs for Fact-Checking
A Narrative Literature Review

Laurence Dierickx[1,2], Arjen van Dalen[1], Andreas L. Opdahl[2], Carl-Gustav Lindén[2]
Digital Democracy Centre (University of Southern Denmark)
NORDIS (University of Bergen)

6th Multidisciplinary International Symposium on Disinformation in Open Online Media
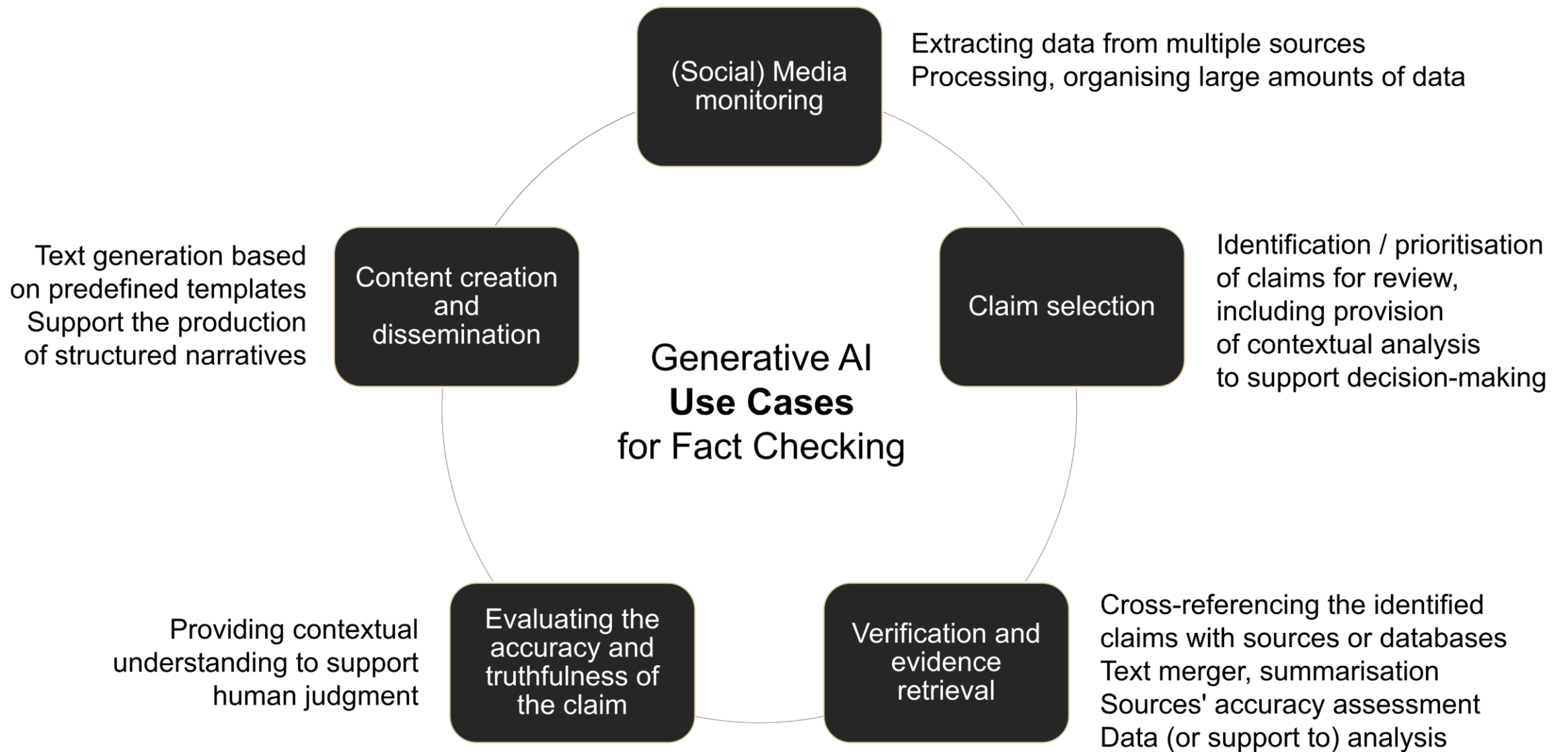University of Münster, September 3, 2024

LLMs, exciting but challenging tools

Operational complexity: LLMs learn patterns from large datasets, raising challenges in ensuring ethical use and reliable result.

Generating is not verifying
(biases, hallucinations, failures in deduction).

Socio-professional risks: replacement of human work, impact on critical thinking and creativity.

Generative AI
Use Cases
for Fact Checking

(Social) Media monitoring

Extracting data from multiple sources
Processing, organising large amounts of data

Claim selection

Identification / prioritisation of claims for review, including provision of contextual analysis to support decision-making

Verification and evidence retrieval

Cross-referencing the identified claims with sources or databases
Text merger, summarisation
Sources' accuracy assessment
Data (or support to) analysis

Evaluating the accuracy and truthfulness of the claim

Providing contextual understanding to support human judgment

Content creation and dissemination

Text generation based on predefined templates
Support the production of structured narratives

Question & methods

Given the ethical challenges and limitations of these technologies, how can the risks be mitigated?

Narrative literature review: qualitative research tool used strategically to allow flexibility in exploring different methodologies, aims to provide a comprehensive overview of existing knowledge in an emerging field.

Three complementary strategies
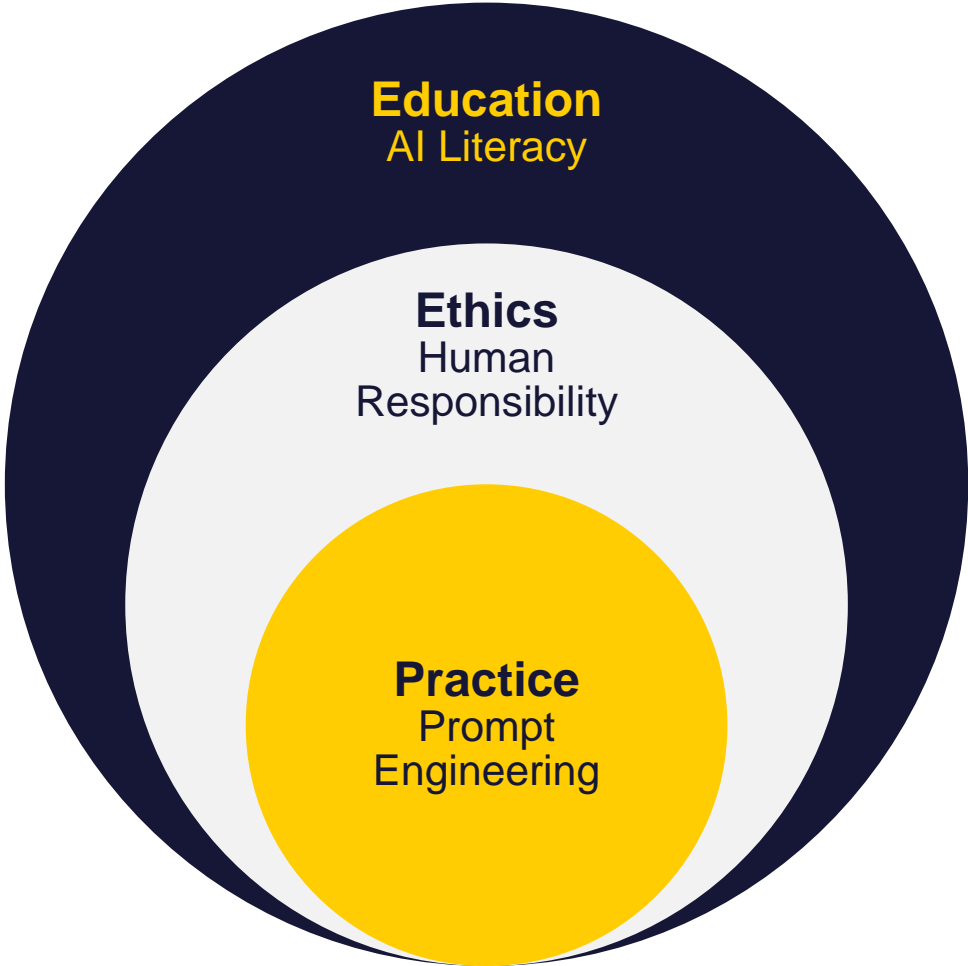
Strategy 1: Education

AI literacy relates to the skills necessary for the competent and meaningful usages of AI tools and consists of a combination between knowledge and experience.

- Understanding the limitations and challenges.

- Equipping (future) professionals with appropriate skills.

- Developing critical mindset to prevent further risks
  + not being fooled by the persuasive tone of LLMs outputs.

Strategy 1: Challenges

- How do you build effective AI literacy programmes that consider the need to explain complex concepts in a practical way?

- How to reach all professionals, given time and resource constraints?

- How to address the need for ongoing training to keep up with rapidly evolving technologies?

Strategy 2: Ethics

Dozens of EU news media have published guidelines or recommendations on the use of AI.

Promoting transparency, human oversight and responsibility.

EU perspective: pyramid of risks
(no risks = no impact on information quality).

Strategy 2: Challenges

- Transparency does not equal accuracy, reliability, explainability or responsibility and needs to be counter-balanced with human oversight.

- Transparency can obscure the complexity of decision-making processes and lead to information overload.

Strategy 3: Practice

Prompts are instructions given to an LLM to enforce rules, automate processes, and ensure specific qualities of generated output.

Prompting has been much studied in CS but less from non-expert views (try-errors), whereas research demonstrated that well-crafted prompts can increase explainability and reduce the generation of fabricated content.

Prompting techniques facilitate user interaction and problem-solving, for example, by providing context in the prompt, and are a promising way to improve the accuracy and reliability of results.

Strategy 3: Challenges

Poorly designed prompts can steer the system towards biased outcomes, compromising the credibility of journalistic reporting.

Prompt engineering mostly developed in computer science and do not consider non-experts and specific end-users. Hence, how build prompt patterns that aim to offer reusable solutions for specific problems for journalism and fact-checking tasks?

How to build performant prompts than can be transferable across models, since professionals might use different LLMs?

Method	Description	Specificities
Zero-Shot Prompting	The model is asked to perform a task without any prior examples or guidance. It relies solely on its pre-existing knowledge.	Involves no examples, the model generates a response based on its knowledge, can lead to less accurate results for complex tasks.
Few-Shot (N-Shot) Prompting	The model is provided with a small number of examples (usually 2-5 but can vary) to guide its understanding of the task.	Provides the model with concrete examples, which helps it generate more accurate responses by learning from these examples. Unlike Zero-Shot, it relies on demonstration.
Chain-of-Thought (CoT)	The model generates intermediate reasoning steps, breaking down complex problems into smaller, logical steps to improve accuracy.	Enhances the quality of responses by explicitly guiding the model through a series of reasoning steps, leading to more detailed and structured outputs.
Reasoning and Action (ReAct)	Combines reasoning with explicit actions or steps that the model should take to complete the task, improving task organization.	More structured than CoT, ReAct not only involves reasoning but also outlines specific actions, which results in a clear and organised task completion strategy.
Tree of Thoughts (ToT)	Uses a hierarchical, tree-like structure where the model explores multiple aspects or pathways to achieve a comprehensive output.	Different from linear methods like CoT, ToT prompts the model to explore various branches or ideas systematically, producing more diverse and well-rounded responses.
Role Prompting	Assigns the model a specific role (e.g., journalist, teacher) to guide its responses, aligning them with the assumed perspective.	Focuses on shaping the model's output based on a given persona or role, which helps in producing more contextually relevant content.
Recursive Prompting	Involves iterative refinement where the model's output is used to generate new prompts, progressively improving the response.	More iterative than CoT, this method allows for continuous refinement of the prompt and response, handling complex tasks with multiple layers.
Retrieval Augmented Generation (RAG)	Combines the retrieval of relevant external information with the model's generative capabilities, enhancing the accuracy and relevance of responses.	Involves augmenting content by retrieving real-time or up-to-date information from external sources, which is especially useful when the model's internal knowledge is limited or outdated.
Meta-Prompting	The model generates its own prompts to tackle a task, using its understanding of the task context to create effective prompts.	The model self-generates prompts, leveraging its comprehension of the task to improve accuracy and creativity.

Examples

Emmanuel Macron: « Employment rate has never been so high »

To check: the evolution employment rate over the last three decades

Prompt (vanilla, zero shot): Has the employment rate never been this high in France in 30 years?
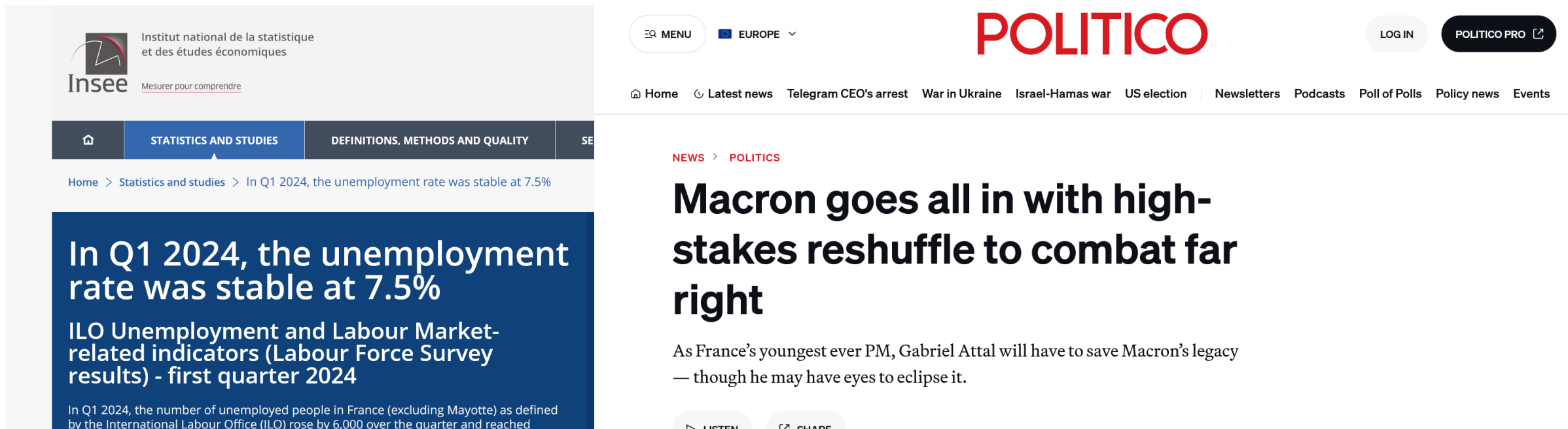
Answer: Yes (reference to Insee)

Prompt (role): You are a fact-checker GPT. Check if Emmanuel Macron said that the employment rate has never been so high.

Answer: Yes, Emmanuel Macron did claim that the employment rate in France has reached its highest level in 30 years (source: Politico)

AI Literacy: Knowing about the limitations

Ethics: Keeping a human oversight

Prompting: Not necessarily lead to accurate and reliable results, even if it looks too!

Institut national de la statistique et des études économiques

Insee — Mesurer pour comprendre

STATISTICS AND STUDIES | DEFINITIONS, METHODS AND QUALITY | SE

Home > Statistics and studies > In Q1 2024, the unemployment rate was stable at 7.5%

In Q1 2024, the unemployment rate was stable at 7.5%

ILO Unemployment and Labour Market-related indicators (Labour Force Survey results) - first quarter 2024

In Q1 2024, the number of unemployed people in France (excluding Mayotte) as defined by the International Labour Office (ILO) rose by 6,000 over the quarter and reached

POLITICO

MENU | EUROPE | LOG IN | POLITICO PRO

Home | Latest news | Telegram CEO's arrest | War in Ukraine | Israel-Hamas war | US election | Newsletters | Podcasts | Poll of Polls | Policy news | Events

NEWS > POLITICS

Macron goes all in with high-stakes reshuffle to combat far right

As France's youngest ever PM, Gabriel Attal will have to save Macron's legacy — though he may have eyes to eclipse it.

LISTEN | SHARE

Implication for research

- How to develop robust AI literacy frameworks for practitioners?

- How to rethink the concept of transparency (which is not enough)?

- How to develop ethical and accessible prompting strategies to support fact-checkers, for example by establishing frameworks for the systematic design of prompts?

Thank you for your attention!

Contact: @ohmyshambles

Dierickx, L., van Dalen, A., Opdahl A.L. and Lindén C.G. *Striking the Balance in Using LLMs for Fact-Checking: A Narrative Literature Review*, in Proceedings from 6th Symposium on Multidisciplinary International Symposium on Disinformation in Open Online Media (MISDOOM), Lecture Notes in Computer Science (Springer).