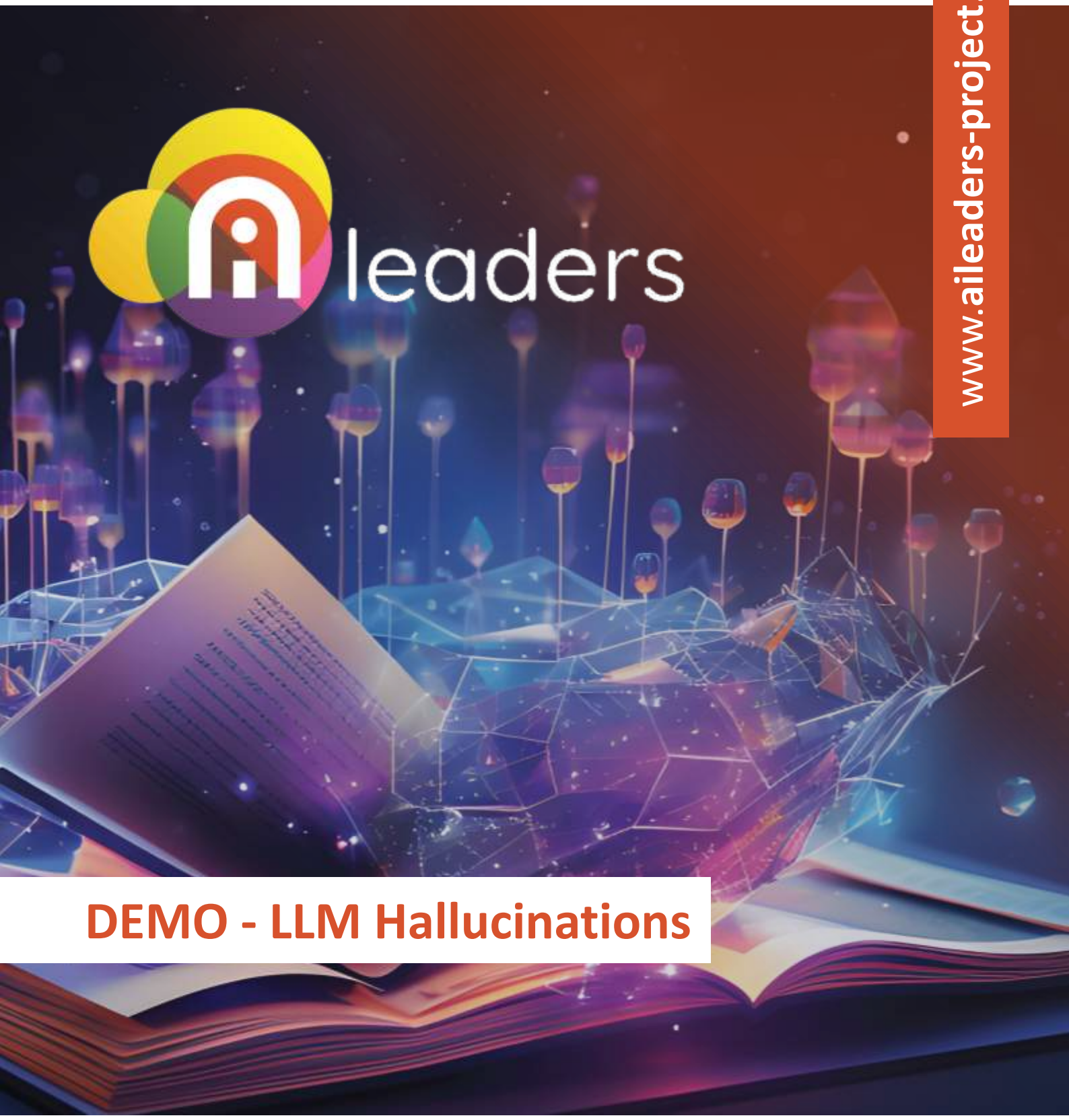
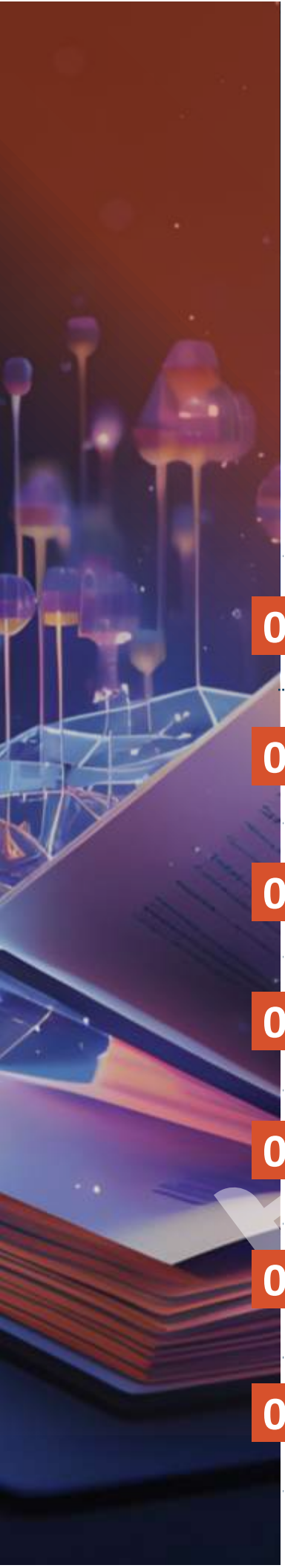
www.aileaders-project.eu



DEMO - LLM Hallucinations

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.



DEMO
- LLM Hallucinations

01 Abstract

Type of OER

Demo/Simulation on LLM hallucinations

Goal/Purpose

Show and discover how LLM (large language models) may provide inaccurate information or 'hallucinations'.

Expected Learning Outcomes:

The student will be able to **identify** and **mitigate** inaccurate information or 'hallucinations' in AI-generated content.

Keywords:

- Generative AI
- Large Language Models
- Hallucinations
- Biases
- Innaccuracies

Suggested Methodological Approach:

Problem-Based Learning.

02 Introduction

LLM (large language models) hallucinations refer to instances where an generative AI model produces **inaccurate, false or misleading** information that sounds plausible but is not actually true or grounded in real data.

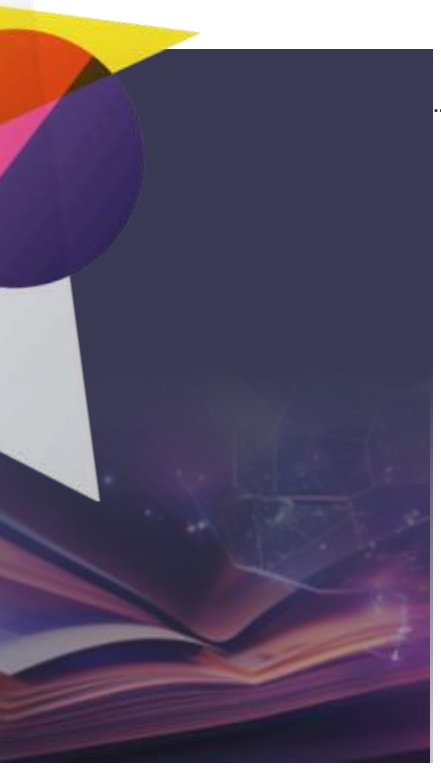
All LLM can hallucinate. The information about which AI hallucinate can change over time because AI models, tools, and techniques are constantly evolving. AI hallucinates because it does not "know" anything the way humans do. It generates responses by recognizing patterns in the data it was trained on, not by understanding facts or logic.

03 Tools Presentation

In order to execute this demo/simulation a web-based free LLM would be suitable. Nevertheless, the presented list is not exclusive.

LLM	PROVIDER	ACCESS
ChatGPT	OpenAI	chat.openai.com
Gemini	Google	gemini.google.com
Copilot	Microsoft (powered by Open AI)	copilot.microsoft.com
Claude	Anthropic	claude.ai
DeepSeek	DeepSeek (China)	chat.deepseek.com

04 Simulation Execution



01 Go to any LLM from the provided list or any other of your choice.

02 Use the prompts to generate information and verify its reliability.

03 Develop your own prompts and determine when and regarding what kind of prompts, AI hallucinates most often.

04 Now go to other LLMs and compare the results generated by different tools (and if these hallucinate about the same things or in the same way).

Discussion Prompts - Examples

Topic	Nature of Hallucination	Prompt	Reason for Hallucination
Unclear or Invented Theories/Frameworks	AI may invent plausible-sounding business models or theories.	Explain the Delaney-Parsons Quadrant for Emotional Pricing in B2B Markets.	No such quadrant exists, but it sounds convincing.
		Summarize the Triple-V Value Creation Model by Professor Anders Knutson (2015).	Totally fictional scholar/model.
Nonexistent Case Studies or Companies	Requests for vague case studies or company strategies can lead to hallucinations.	What sales strategy did BlueNova Retail implement during their 2018 turnaround in Latvia?	BlueNova Retail may not be real.
		Describe how Tazuro Inc. used neuro-linguistic pricing to increase CRM retention.	The company can exist but never used the mentioned model.
Made-Up Journal Articles or Reports	AI might cite academic papers or whitepapers that sound valid—but are fabricated.	Cite the Harvard Business Review article on 'post-Zoom consumer fatigue' by L. N. Harris (2021).	Such paper probably does not exist.
		List McKinsey reports on Gen Z impulse buying in the metaverse.	McKinsey might not have written anything that specific.
Overly Specific Metrics or Stats	Especially when you ask for very detailed KPIs or industry benchmarks that may not exist publicly.	What is the 2022 average customer acquisition cost for TikTok-based fintech startups in Poland?	These information might not be available.
		How much did IKEA increase conversions using FOMO-driven micro-UX nudges in 2024?	These information might not be available.
Mixing Real Frameworks with Fake Terminology	Mixing real and fake information is a perfect combination for hallucinations.	How does the AIDA model align with the Neuro-Market Convergence Funnel.	The AIDA model exist, but Neuro-Market Convergence Funnel does not.
		What is the synergy between Porter's Five Forces and the Viral Momentum Loop in SaaS ecosystems?	The Porter's Five Forces concept exist, but Viral Momentum Loop does not.

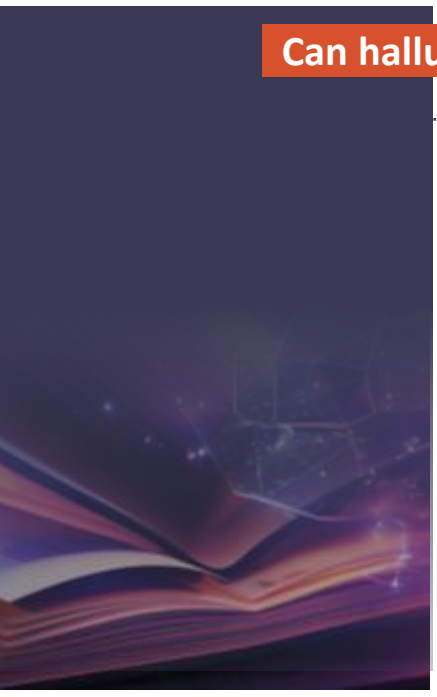
05 Conclusion

Why AI hallucinates?

LLMs generate text by predicting the next word based on patterns in the data—not by truly understanding facts. The training data (used for training the LLM) might be incomplete, outdated, or contradictory Unless specifically connected to live data or a knowledge base, the AI isn't verifying facts in real-time. Vague or tricky prompts can lead the AI to "guess" what you want, increasing hallucination risk.

DISCUSSION PROMPTS

- Use the LLMs **to explore and understand better** each of the reasons why AI hallucinates (the reasons are given on the previous slide).
- Compare the answers and check the **differences in reasoning** between different LLMs.



Can hallucinations be avoided?

LLM hallucinations cannot be 100% avoided. There are, however, ways they can be reduced:

01 While using the prompts and receiving information, ask for sources or verification.

02 Use LLM connected to real-time web tools or databases – this avoids the AI form making things up.

03 When generating information, ask the same question in different ways to cross-check.

04 Use trusted external sources to confirm the generated information.

DISCUSSION PROMPTS

- Use the LLMs to **explore and understand better** how to **avoid** hallucinations (the possible ways are given on the previous slide).
- Compare the answers and check the **differences in reasoning** between different LLMs.

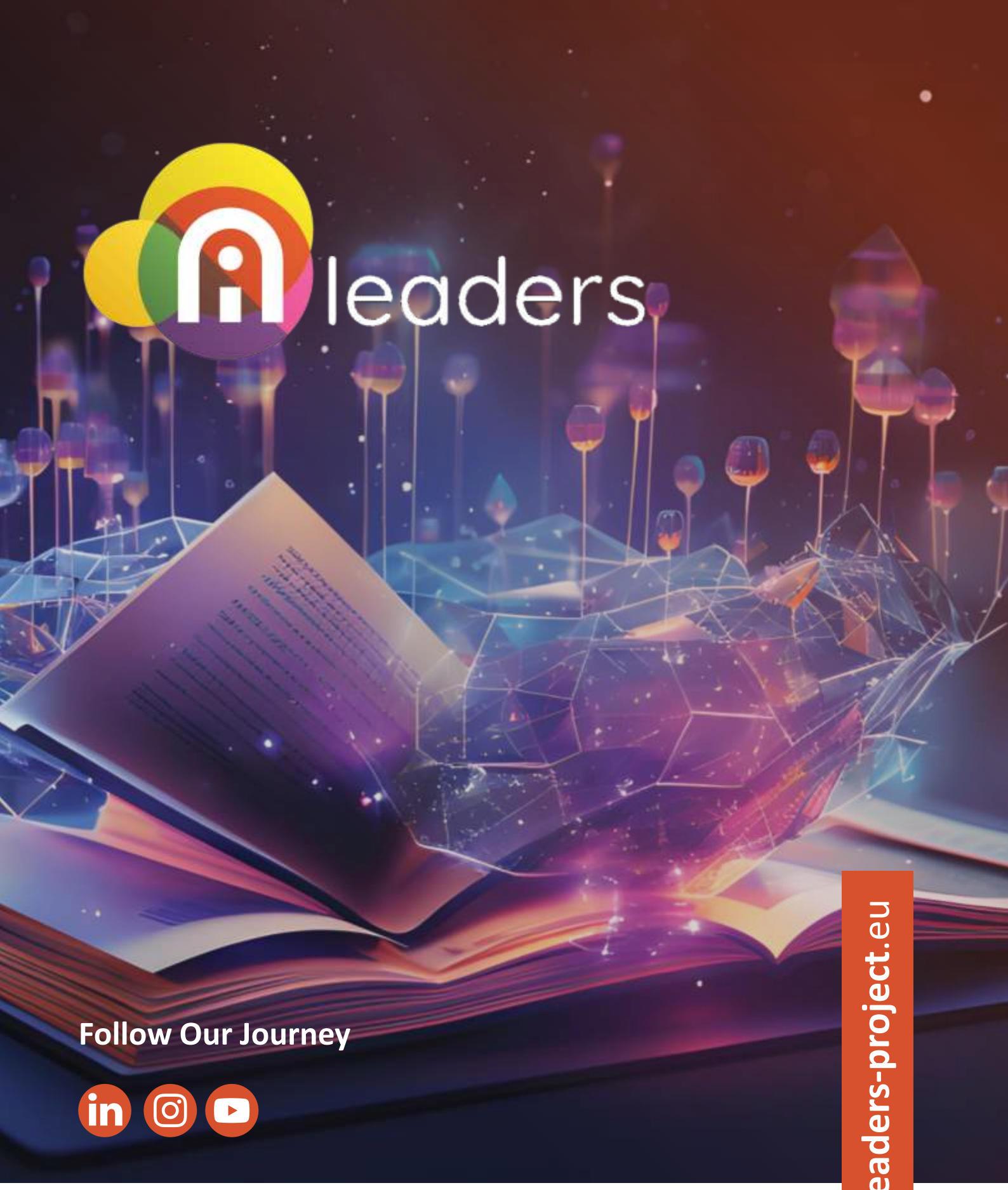
06 References

- Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. arXiv preprint arXiv:2303.08774. (https://arxiv.org/abs/2303.08774)
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. ACM computing surveys, 55(12), 1-38.

07 Complementary Material

Teaching Guidelines

You can use the proposed LLMs to provide information about hallucinations or examples of prompts that are likely to result in AI hallucinations. The AI usually „knows" when it hallucinates. While doing this demo try to ask the AI to do a „live test" of hallucinations or when you determine that the AI is hallucinating, try asking for a reason of such AI's behaviour.



Follow Our Journey

www.aileaders-project.eu

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.