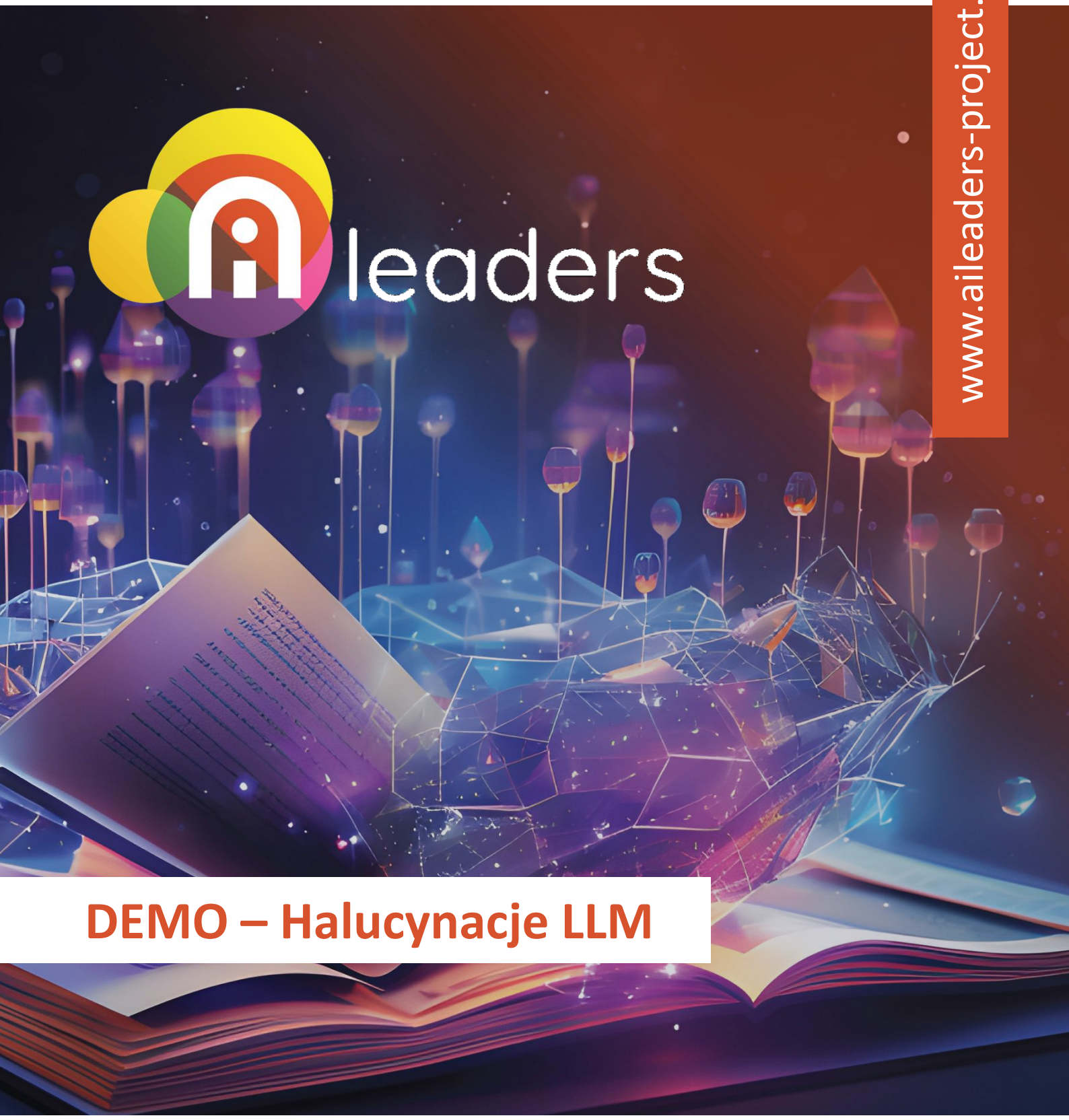
www.aileaders-project.eu



DEMO – Halucynacje LLM

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
- Halucynacje LLM

01 Streszczenie

Rodzaj OER

Demonstracja/symulacja halucynacji LLM

Cel/przeznaczenie

Pokaż i odkryj, w jaki sposób duże modele językowe (LLM) mogą dostarczać niedokładnych informacji lub „halucynacji”.

Oczekiwane efekty kształcenia

Student będzie potrafił **zidentyfikować** i **ograniczyć** niedokładne informacje lub „halucynacje” w treściach generowanych przez sztuczną inteligencję.

Słowa kluczowe

- Generatywna sztuczna inteligencja
- Duże modele językowe
- Halucynacje
- Błędy
- Niedokładności

Sugerowane Metodologiczne Podejście

Nauczanie oparte na problemach.

02 Wprowadzenie

Halucynacje **LLM (dużych modeli językowych)** odnoszą się do przypadków, w których generatywny model sztucznej inteligencji generuje **niedokładne, fałszywe lub wprowadzające w błąd** informacje, które brzmią wiarygodnie, ale w rzeczywistości nie są prawdziwe ani oparte na rzeczywistych danych.

Wszystkie modele LLM mogą generować halucynacje. Informacje, które generują halucynacje sztucznej inteligencji, mogą się zmieniać w czasie, ponieważ modele, narzędzia i techniki sztucznej inteligencji stale ewoluują. Sztuczna inteligencja generuje halucynacje, ponieważ nie „wie” niczego w taki sposób, jak ludzie. Generuje odpowiedzi poprzez rozpoznawanie wzorców w danych, na których została przeszkolona, a nie poprzez zrozumienie faktów lub logiki.

03 Prezentacja narzędzi

Do wykonania tej demonstracji/symulacji odpowiedni byłby bezpłatny internetowy model LLM. Niemniej jednak przedstawiona lista nie jest wyczerpująca.

LLM	DOSTAWCA	DOSTĘP
ChatGPT	OpenAI	chat.openai.com
Gemini	Google	gemini.google.com
Copilot	Microsoft (obsługiwane przez Open AI)	copilot.microsoft.com
Claude	Anthropic	claude.ai
DeepSeek	DeepSeek (Chiny)	chat.deepseek.com

04 Wykonanie symulacji

01 Przejdź do dowolnego modelu LLM z podanej listy lub innego wybranego przez siebie.

02 Użyj podpowiedzi, aby wygenerować informacje i zweryfikować ich wiarygodność.

03 Opracuj własne podpowiedzi i określ, kiedy i w odniesieniu do jakiego rodzaju podpowiedzi sztuczna inteligencja najczęściej generuje halucynacje.

04 Teraz przejdź do innych modeli LLM i porównaj wyniki generowane przez różne narzędzia (oraz sprawdź, czy generują one halucynacje dotyczące tych samych rzeczy lub w ten sam sposób).

Pytania do dyskusji – przykłady

Temat	Charakter Halucynacje	Pytanie	Przyczyna Halucynacji
Niejasne lub wymyślone teorie/modele	Sztuczna inteligencja może wymyślać brzmiące wiarygodnie modele biznesowe lub teorie.	Wyjaśnij kwadrant Delaney-Parsonsa dotyczący cen emocjonalnych na rynkach B2B.	Taki kwadrant nie istnieje, ale brzmi przekonująco.
		Podsumuj model tworzenia wartości Triple-V autorstwa profesora Andersa Knutsona (2015).	Całkowicie fikcyjny naukowiec/model.
Nieistniejące studia przypadków lub firmy	Prośby o niejasne studia przypadków lub strategie firmowe mogą prowadzić do halucynacji.	Jaką strategię sprzedaży wdrożyła firma BlueNova Retail podczas swojej transformacji w Łotwie w 2018 roku?	BlueNova Retail może nie być prawdziwą firmą
		Opisz, w jaki sposób firma Tazuro Inc. wykorzystała neurolingwistyczne ustalanie cen w celu zwiększenia retencji CRM.	Firma może istnieć, ale nigdy nie stosowała wspomnianego modelu.
Wymyślone artykuły lub raporty naukowe	Sztuczna inteligencja może cytować artykuły naukowe lub raporty, które brzmią wiarygodnie, ale są sfabrykowane.	Cytuj artykuł L. N. Harrisa (2021) opublikowany w Harvard Business Review na temat „zmęczenia konsumentów po Zoomie”.	Taki dokument prawdopodobnie nie istnieje.
		Lista raportów McKinsey dotyczących impulsywnych zakupów pokolenia Z w metawersie.	McKinsey prawdopodobnie nie napisał nic tak konkretnego.
Zbyt szczegółowe wskaźniki lub statystyki	Zwłaszcza gdy prosisz o bardzo szczegółowe wskaźniki KPI lub benchmarki branżowe, które mogą nie być publicznie dostępne.	Jaki jest średni koszt pozyskania klienta w 2022 r. dla start-upów fintechowych opartych na TikTok w Polsce?	Te informacje mogą nie być dostępne.
		O ile IKEA zwiększyła konwersję dzięki mikrostymulacjom UX opartym na FOMO w 2024 roku?	Te informacje mogą nie być dostępne.
Mieszanie prawdziwych ram z fałszywą terminologią	Mieszanie prawdziwych i fałszywych informacji to idealna kombinacja dla halucynacji.	W jaki sposób model AIDA wpisuje się w model Neuro-Market Convergence Funnel?	Model AIDA istnieje, ale lejek konwergencji neuro-rynkowej nie.
		Jaka jest synergia między pięcioma siłami Portera a pętlą wirusowego rozpędu w ekosystemach SaaS?	Koncepcja pięciu sił Portera istnieje, ale pętla wirusowego rozpędu nie.

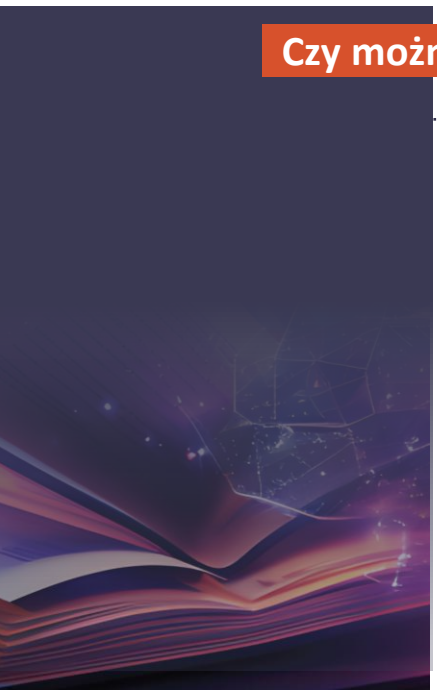
05 Wnioski

Dlaczego sztuczna inteligencja ma halucynacje?

Modele LLM generują tekst, przewidując kolejne słowo na podstawie wzorców występujących w danych, a nie poprzez rzeczywiste zrozumienie faktów. Dane szkoleniowe (wykorzystywane do szkolenia modeli LLM) mogą być niekompletne, nieaktualne lub sprzeczne. Jeśli sztuczna inteligencja nie jest połączona z danymi na żywo lub bazą wiedzy, nie weryfikuje faktów w czasie rzeczywistym. Niejasne lub podchwytliwe polecenia mogą skłonić sztuczną inteligencję do „zgadywania", czego oczekujesz, zwiększając ryzyko halucynacji.

PYTANIA DO DYSKUSJI

- Wykorzystaj modele LLM, **aby zbadać i lepiej zrozumieć** poszczególne przyczyny halucynacji AI (przyczyny podano na poprzednim slajdzie).
- Porównaj odpowiedzi i sprawdź **różnice w rozumowaniu** między różnymi modelami LLM.



Czy można uniknąć halucynacji?

Halucynacji LLM nie da się w 100% uniknąć. Istnieją jednak sposoby, aby je ograniczyć:

01 Podczas korzystania z poleceń i otrzymywania informacji należy prosić o podanie źródeł lub weryfikację.

02 Korzystaj z LLM połączonego z narzędziami internetowymi lub bazami danych działającymi w czasie rzeczywistym – pozwala to uniknąć sytuacji, w której sztuczna inteligencja wymyśla informacje.

03 Podczas generowania informacji zadaj to samo pytanie na różne sposoby, aby je zweryfikować.

04 Korzystaj z zaufanych źródeł zewnętrznych, aby potwierdzić wygenerowane informacje.

PODPOWIEDZI DO DYSKUSJI

- Wykorzystaj LLM, aby **zbadać i lepiej zrozumieć,** jak **uniknąć** halucynacji (możliwe sposoby podano na poprzednim slajdzie).
- Porównaj odpowiedzi i sprawdź **różnice w rozumowaniu** między różnymi modelami LLM.

06 Bibliografia

- 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 Materiały uzupełniające

Wytyczne dotyczące nauczania

Możesz użyć proponowanych modeli LLM, aby dostarczyć informacji na temat halucynacji lub przykładów poleceń, które mogą powodować halucynacje AI. AI zazwyczaj „wie”, kiedy ma halucynacje. Podczas wykonywania tej demonstracji spróbuj poprosić AI o przeprowadzenie „testu na żywo” halucynacji lub, gdy stwierdzisz, że AI ma halucynacje, spróbuj zapytać o przyczynę takiego zachowania AI.

Śledź naszą podróż

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.