Reading news stories on LLMs

Jindřich Libovický
April 14, 2025

Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

Learning outcome

Be able to critically think about news stories about LLMs in terms of news values.

Outline

Instructor talking

- Lecture on news stories [10 min]
- Introducing news articles [10 min]

Group work

- Reading articles, identifying news values
- Fact-checking and improving the news story
- Final report

News Values

Theory of news values

- Criteria used by journalists to determine whether a story is newsworthy
- Historical context:
  - Concept first formalized by Johan Galtung and Mari Holmboe Ruge in 1965
  - They analyzed international news coverage to identify patterns in story selection
  - Their work identified 12 factors that influenced newsworthiness
  - The concept has evolved but remains fundamental to journalism practice
  - Later scholars like Harcup and O'Neill (2001, 2016) refined and updated these values
- Understanding news values helps us critically analyze media coverage, including tech news
- These values influence what stories about LLMs and AI get covered and how

The values (1)

Frequency

- **Frequency**: Events that unfold within the news production cycle
- Examples in tech news:
  - Short-term AI experiments or challenges get more coverage than longer-term research
  - Quick breakthroughs receive more attention than gradual improvements
  - Quarterly earnings and product launches match news cycles better than ongoing development

Timeliness

- **Timeliness**: Recent or current events are more newsworthy
- Examples in LLM coverage:
  - Breaking news about model releases (e.g., "GPT-5 released today")
  - Emerging trends in AI capabilities ("New benchmark results show...")
  - Quick coverage of AI conferences and announcements

The values (2)

Familiarity

- **Familiarity**: Events involving recognizable topics, places, or concepts
- Examples in AI news:
  - Coverage focuses on familiar use cases (chatbots, image generation)
  - New AI concepts often explained through familiar metaphors
  - Stories connecting AI to well-known movies or books (e.g., HAL 9000, Skynet)

Negativity

- **Negativity**: Bad news tends to be more newsworthy than good news
- Examples in LLM coverage:
  - AI safety concerns and potential risks receive extensive coverage
  - Data breaches and model failures make headlines
  - Stories about AI replacing jobs get more attention than job creation

The values (3)

Conflict

- **Conflict**: Stories involving disagreement, debate, or tension
- Examples in LLM news:
  - "Researchers clash over AI safety priorities"
  - "Legal battle over training data copyright"
  - "Tech CEOs disagree on AI regulation approach"

Unexpectedness

- **Unexpectedness**: Rare, surprising, or unusual events
- Examples in AI news:
  - "AI generates unexpected solution to 50-year-old math problem"
  - "Language model demonstrates surprising new capability"
  - "AI performs task experts thought impossible for current systems"

The values (4)

Unambiguity

- **Unambiguity**: Events with clear meaning or interpretation
- Examples in LLM coverage:
  - Simple metrics and benchmarks get more attention than nuanced evaluations
  - Clear successes or failures highlighted over mixed results
  - Headlines emphasize certainty rather than the complexities of AI systems

Personalization

- **Personalization**: Events framed through individual experiences
- Examples in AI news:
  - "Developer creates AI to solve personal problem"
  - "CEO's vision shapes company's AI strategy"
  - "User's experience with AI assistant goes viral"

The values (5)

Meaningfulness

- **Meaningfulness**: Events culturally relevant to the audience
- Examples in LLM coverage:
  - AI's impact on local industries or education systems
  - How language models handle specific cultural contexts
  - AI developments that align with cultural values or concerns

Eliteness

- **Eliteness**: Stories involving prominent nations, institutions, or people
- Examples in AI news:
  - Leading tech companies (Google, OpenAI, Anthropic) receive more coverage
  - Elite university research gets more attention than work from lesser-known institutions
  - Famous tech figures' opinions on AI are extensively covered

The values (6)

Superlativeness

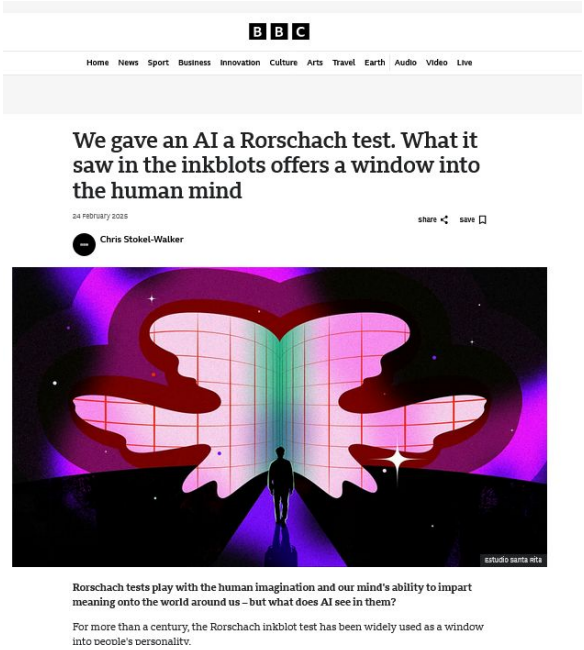
- **Superlativeness**: Events of large magnitude or scope
- Examples in LLM coverage:
  - "Largest language model ever created"
  - "Most advanced AI system to date"
  - "Record-breaking performance on benchmarks"

Consonance

- **Consonance**: Events that fit with existing expectations or narratives
- Examples in AI news:
  - Stories reinforcing "AI revolution" narrative
  - Coverage matching pre-existing hopes or fears about technology
  - News that confirms what journalists or the public already believe about AI

Selected news stories

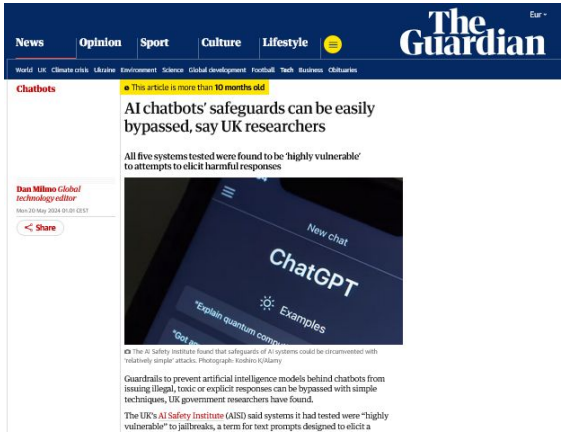
https://www.bbc.com/future/article/20250224-what-happens-when-you-give-an-ai-a-rorschach-inkblot-test



We gave an AI a Rorschach test. What it saw in the inkblots offers a window into the human mind

The Guardian

https://www.theguardian.com/technology/article/2024/may/20/ai-chatbots-safeguards-can-be-easily-bypassed-say-uk-researchers


News Opinion Sport Culture Lifestyle The Guardian

Chatbots

This article is more than 10 months old

AI chatbots' safeguards can be easily bypassed, say UK researchers

All five systems tested were found to be 'highly vulnerable' to attempts to elicit harmful responses

Dan Milmo Global technology editor

Share

New chat

ChatGPT

Examples

The AI Safety Institute found that safeguards of AI systems could be circumvented with 'relatively simple' attacks. Photograph: Koshiro K/Alamy

Guardrails to prevent artificial intelligence models behind chatbots from issuing illegal, toxic or explicit responses can be bypassed with simple techniques, UK government researchers have found.

The UK's AI Safety Institute (AISI) said systems it had tested were "highly vulnerable" to jailbreaks, a term for text prompts designed to elicit a

AI chatbots' safeguards can be easily bypassed, say UK researchers

Deník N

https://denikn.cz/1653441/ai-dosahla-schopnosti-vytvorit-vlastni-kopii-jde-o-nenapadny-ale-nebezpecny-zlom/?ref=list

Deník N – rozumět lépe světu

DENÍK N

Můj účet

Darovat předplatné

Česko Svět Ekonomika Kultura Magazín Podcasty Kontext N E-shop knihy Další

14. února 2025 15:48 Umělá inteligence

AI dosáhla schopnosti vytvořit vlastní kopii. Jde o nenápadný, ale nebezpečný zlom

PETR KOUBSKÝ + Odebírat e-mailem

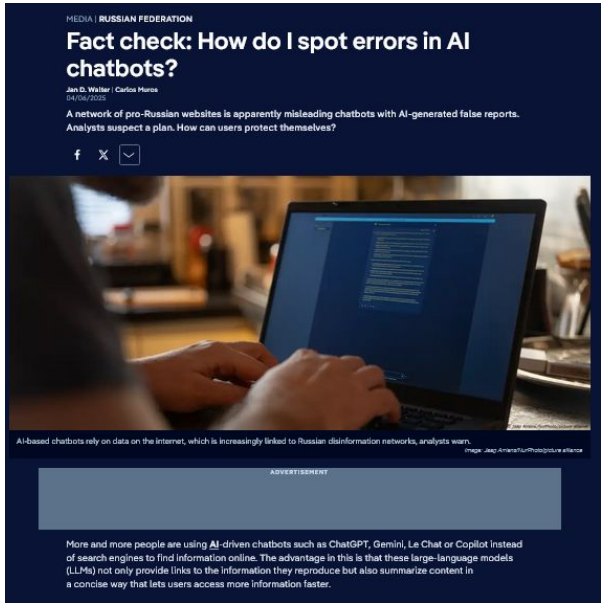
Postavili si pěknou AI, jenže pak se jim začala nekontrolovaně množit… Ilustrace: Adobe Stock (generováno AI)

Čínští vědci z univerzity Fu-tan v Šanghaji naučili dva jazykové modely – jeden americký, jeden čínský – vytvořit plně funkční kopii

AI dosáhla schopnosti vytvořit vlastní kopii. Jde o nenápadný, ale nebezpečný zlom

Deutsche Welle

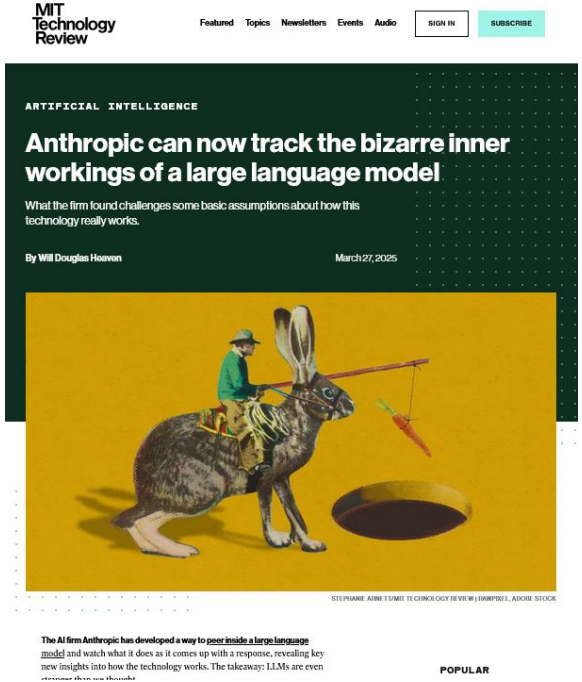
https://www.dw.com/en/fact-check-how-do-i-spot-errors-in-ai-chatbots/a-72106646



Fact check: How do I spot errors in AI chatbots?

MIT Technology Review

https://www.technologyreview.com/2025/03/27/1113916/anthropic-can-now-track-the-bizarre-inner-workings-of-a-large-language-model/



Anthropic can now track the bizarre inner workings of a large language model

All articles

BBC
Roschard Test

1

Guardian
Safeguards bypassed

2

Deník N
Self-cloning

3

DW
Fact checking

4

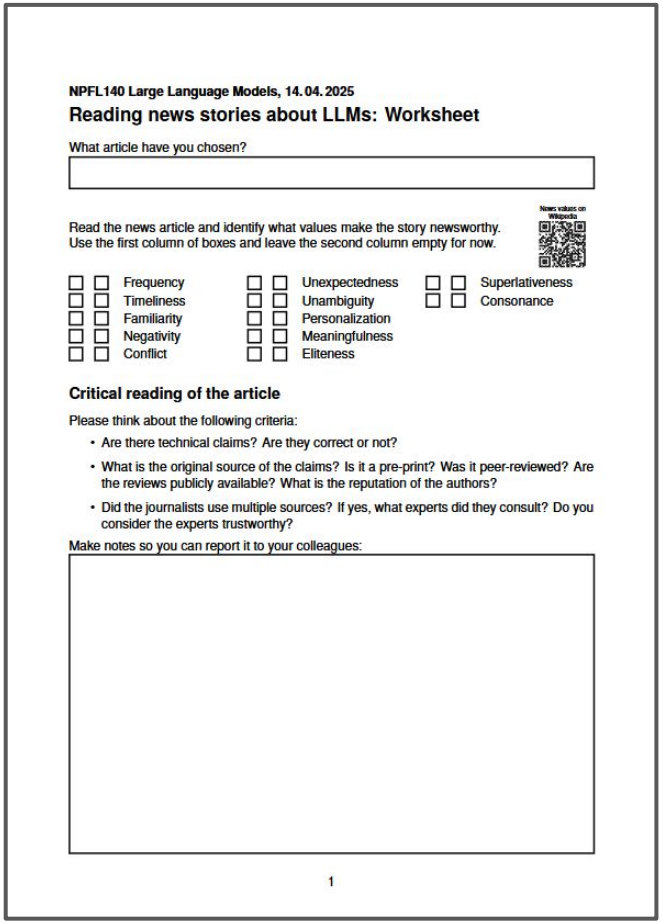
MIT TR
Anthropic's news

5

Group Activity

Activity 1: Critical reading of the aricle

- Form groups
- Choose article
- Follow page one of the worksheet
- Report your findings



NPFL140 Large Language Models, 14. 04. 2025

Reading news stories about LLMs: Worksheet

What article have you chosen?

Read the news article and identify what values make the story newsworthy.
Use the first column of boxes and leave the second column empty for now.

News values on Wikipedia

☐ ☐ Frequency	☐ ☐ Unexpectedness	☐ ☐ Superlativeness
☐ ☐ Timeliness	☐ ☐ Unambiguity	☐ ☐ Consonance
☐ ☐ Familiarity	☐ ☐ Personalization	
☐ ☐ Negativity	☐ ☐ Meaningfulness	
☐ ☐ Conflict	☐ ☐ Eliteness	

Critical reading of the article

Please think about the following criteria:

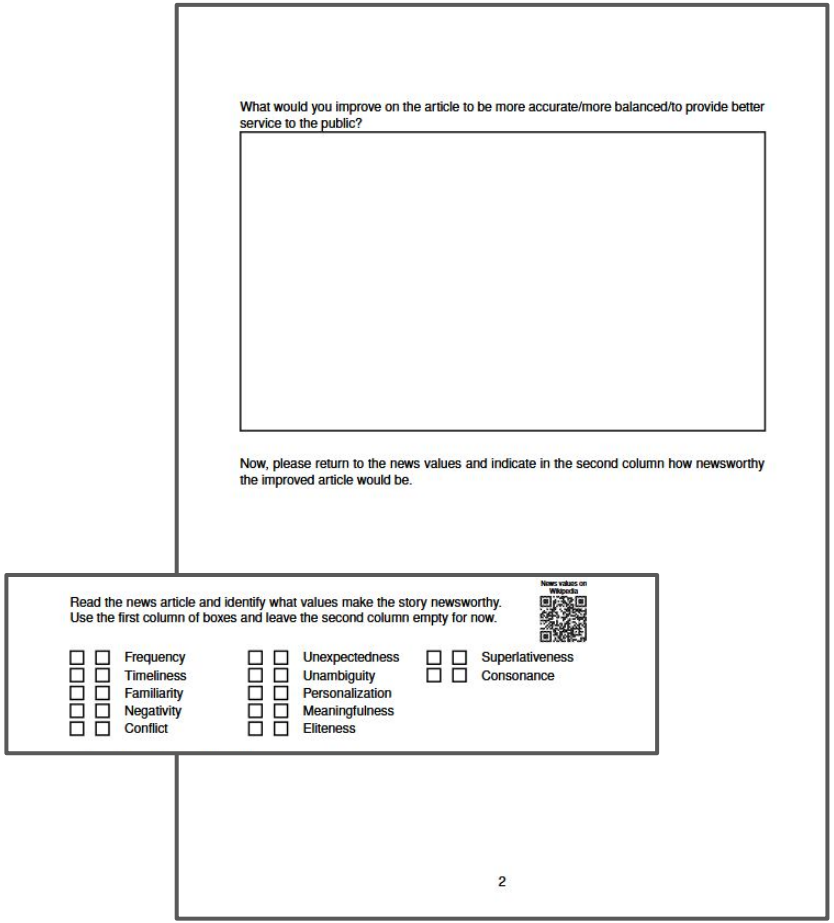
- Are there technical claims? Are they correct or not?
- What is the original source of the claims? Is it a pre-print? Was it peer-reviewed? Are the reviews publicly available? What is the reputation of the authors?
- Did the journalists use multiple sources? If yes, what experts did they consult? Do you consider the experts trustworthy?

Make notes so you can report it to your colleagues:

1

Activity 2: Improve the article

- Follow page 2 of the worksheet
- What would you do to make the article more informative
- How would it change it's newsworthiness
- Report to the class



What would you improve on the article to be more accurate/more balanced/to provide better service to the public?

Now, please return to the news values and indicate in the second column how newsworthy the improved article would be.

Read the news article and identify what values make the story newsworthy.
Use the first column of boxes and leave the second column empty for now.

News values on Wikipedia

Frequency	Unexpectedness	Superlativeness
Timeliness	Unambiguity	Consonance
Familiarity	Personalization	
Negativity	Meaningfulness	
Conflict	Eliteness	

2