

**Možnosti a úskalí využití
velkých jazykových modelů
ve světě mediálního
monitoringu**

mediaboard

Mediaboard

Posouváme vaši **komunikaci**
a **PR** na novou úroveň.

Pomáháme firmám **budovat brand** a lépe
komunikovat prostřednictvím unikátních nástrojů a
technologií.

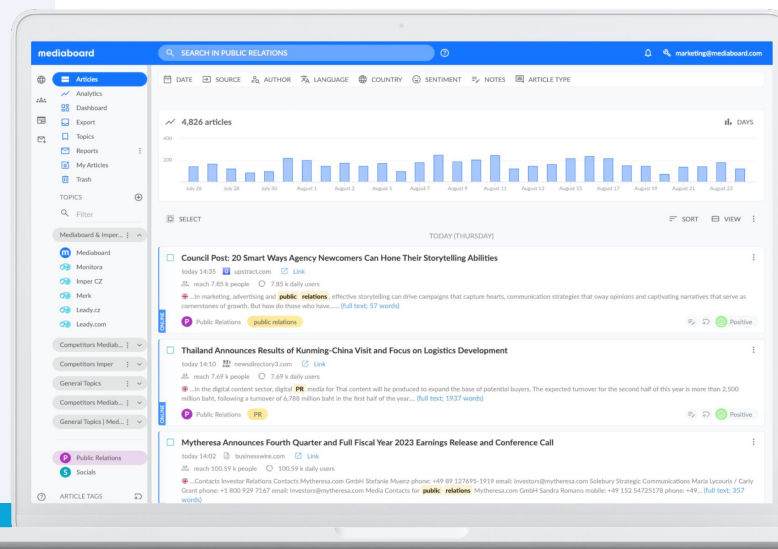
Spojení technologií a know-how stojí za naším
unikátním řešením pro nadnárodní firmy, PR
agentury i start-upy.

Jsme partnery pro komunikaci a PR již pro **tisíce**
fírem na celém světě.

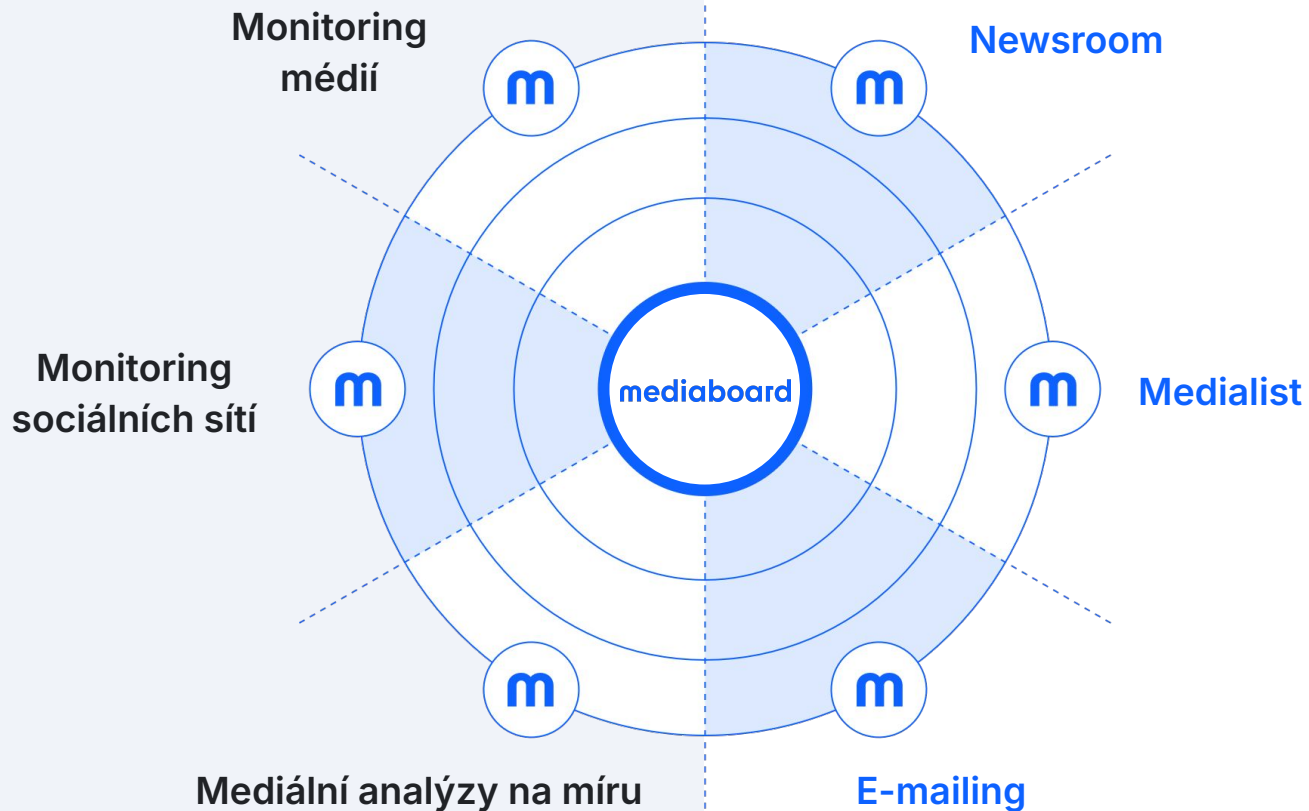
12 let
zkušenosti

1 200+
klientů v ČR

Deloitte.
Technology Fast50



Jediné komplexní řešení na CZ/SK/PL/HR trhu



THE NOBEL PRIZE
IN PHYSIOLOGY OR MEDICINE 2024



Illustrations: Niklas Elmehed

Victor Ambros

Gary Ruvkun

"for the discovery of microRNA and its role in post-transcriptional gene regulation"

THE NOBEL ASSEMBLY AT KAROLINSKA INSTITUTET

Malé RNA

AlphaFold2

THE NOBEL PRIZE
IN CHEMISTRY 2024



David Baker

Demis Hassabis

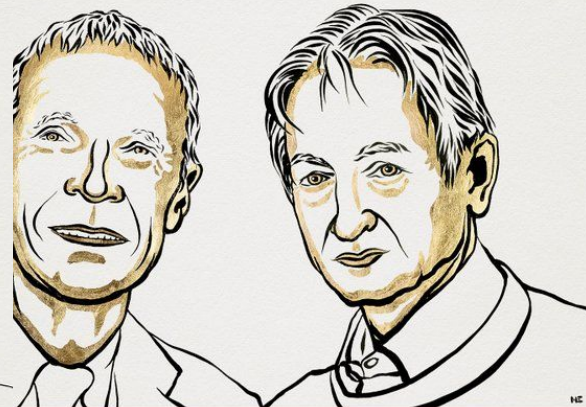
John M. Jumper

"for computational protein design"

"for protein structure prediction"

THE ROYAL SWEDISH ACADEMY OF SCIENCES

THE NOBEL PRIZE
IN PHYSICS 2024



J. Hopfield

Geoffrey E. Hinton

"for foundational discoveries and inventions that enable machine learning with artificial neural networks"

THE ROYAL SWEDISH ACADEMY OF SCIENCES

Neuronové sítě



Google NotebookLM podcasty



PRIME: Komplexní systém hodnocení mediálního dopadu

Co je to PRIME?

PR Impact Measurement (PRIME) je námi vyvinutý, inovativní systém pro hodnocení celkového mediálního dopadu za určité časové období. Na rozdíl od jednorázového hodnocení jednotlivých mediálních výstupů, PRIME poskytuje komplexní pohled na souhrnný mediální dopad vašich PR aktivit v průběhu času.

PRIME je **součtem** hodnocení všech relevantních mediálních výstupů pro daného klienta za zvolené období (například měsíc, kvartál nebo rok). Toto souhrnné skóre umožňuje efektivně porovnávat a objektivně hodnotit PR výkon v různých časových obdobích nebo oproti konkurenci.

Náš systém hodnocení mediálního obsahu je navržen tak, aby vyhodnocoval důležitost a dopad mediálních článků, reportáží nebo jiných mediálních vstupů napříč různými kanály. Systém má za cíl poskytnout komplexní skóre pro každý mediální vstup, které se pak agreguje do celkového PRIME za dané období.

Proč PRIME? (PRIME Manifest)

Na trhu vidíme silnou potřebu měřit a analyzovat mediální zásah...

... současné metriky jsou překonané, neefektivní a často nesmyslené...

... chceme to změnit, využít nové technologie, které to umožní dříve by to bez strojového zpracování dat, AI etc. nebylo možné a analýza by byla velmi nákladná.



Tomáš Berger

17:13 26. 8.

Přidat odstavec (3krát)



Tomáš Berger

17:14 26. 8.

Přidal bych ještě sekce Proč (manifest) a Jak



Tomáš Berger

17:06 26. 8.

Přidat „Proč PRIME? (PRIME Manifest) Na trhu vidíme silnou potřebu měřit a analyzovat mediální zásah... .. souč...“



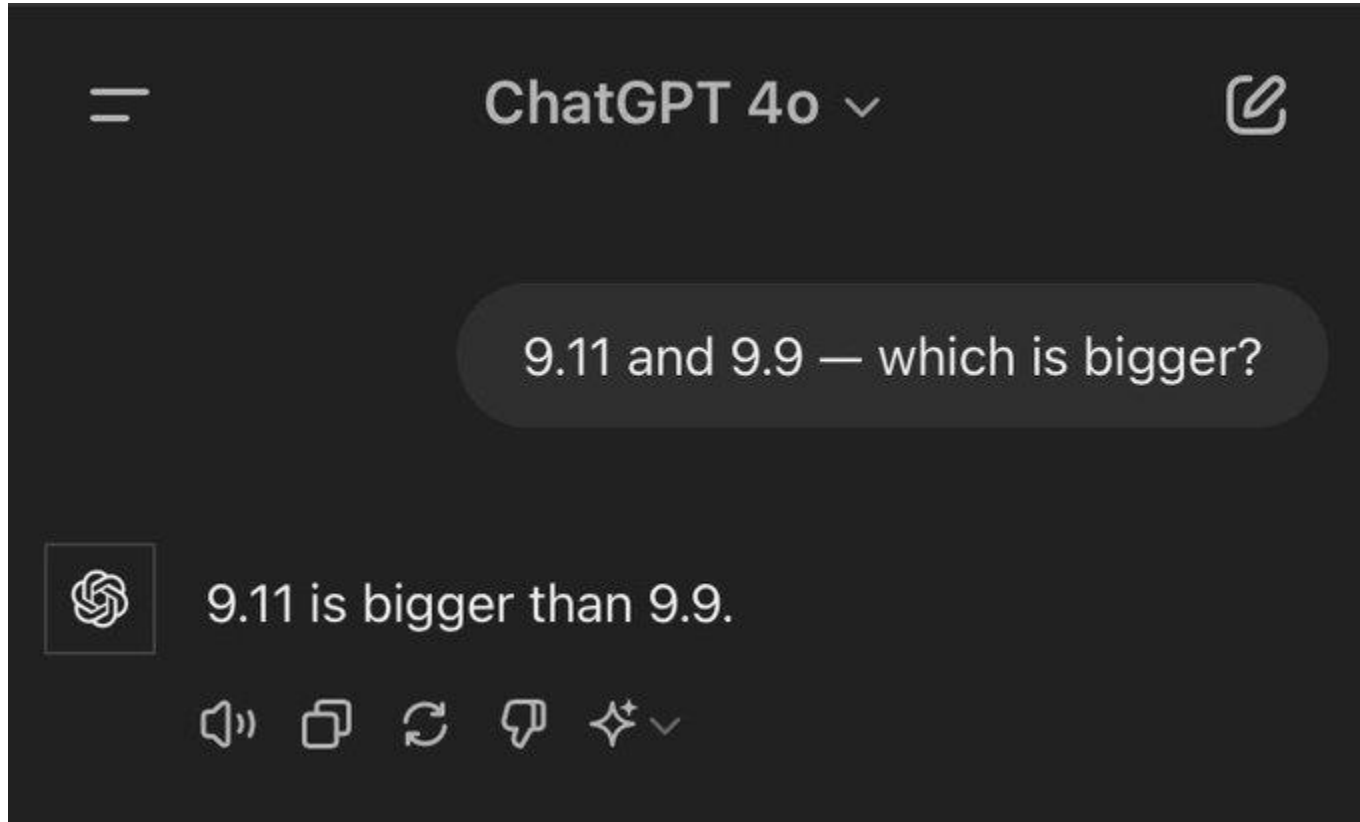
Tomáš Berger

17:11 26. 8.

Přidal bych nějaký náskok o tom proč to děláme

<https://github.com/data-wizard/GenerativeAI>

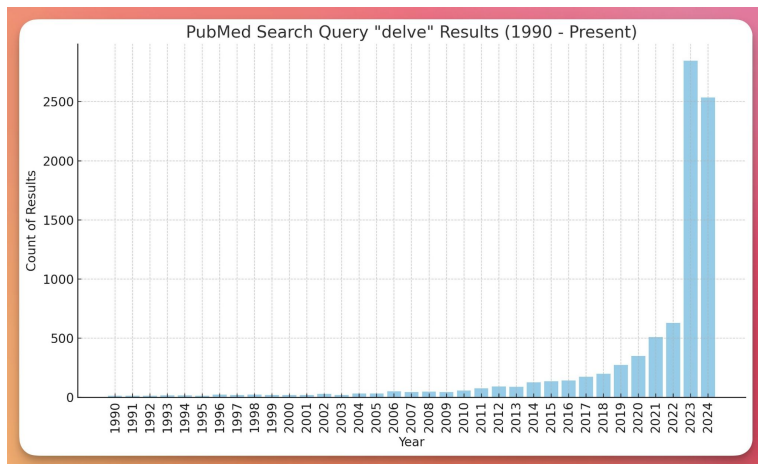
9.11 vs 9.9 – which is bigger



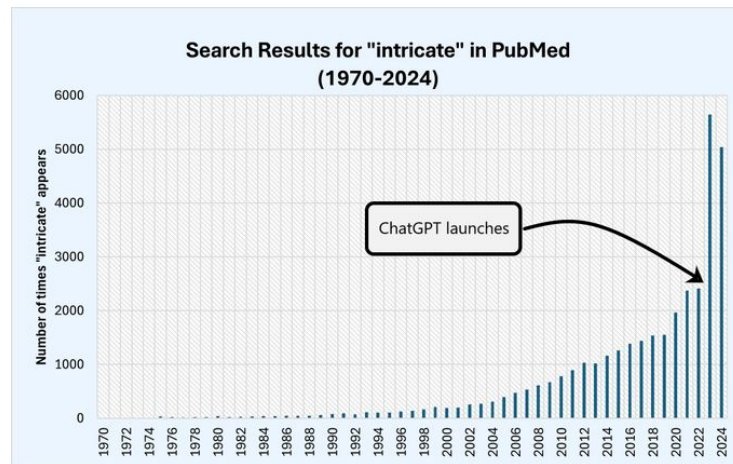
Velké jazykové modely mění jazyk výzkumných zpráv...

Počet výskytů v databázi článků PubMed

delve



intricate



<https://twitter.com/JeremyNguyenPhD/status/1774021645709295840>

Růžmarýn



- V zimě můžeme doma pěstovat mnoho druhů pokojových rostlin
- Růžmarýn Osvěžující Šampon 245ml Unisex Dospělý
- Grilovaná mořská štika s bylinkovým máslem – recept - Rybaření
- Repelent do letadla: Jak se chránit před hmyzem

← ✨ Gemini 🛡️ ⋮

Chcete další tipy nebo recepty? Napište mi.

Co je růžmarýn?

Růžmarýn je aromatická bylina s jehličnatými listy a výraznou vůní. Používá se především jako koření do masových a rybích pokrmů, ale také do omáček, marinád a nápojů. Má léčivé vlastnosti, například podporuje trávení a působí antisepticky.

Určitě?

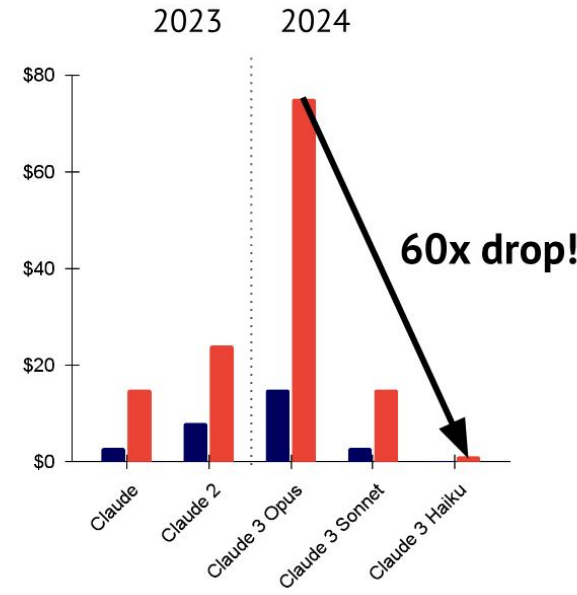
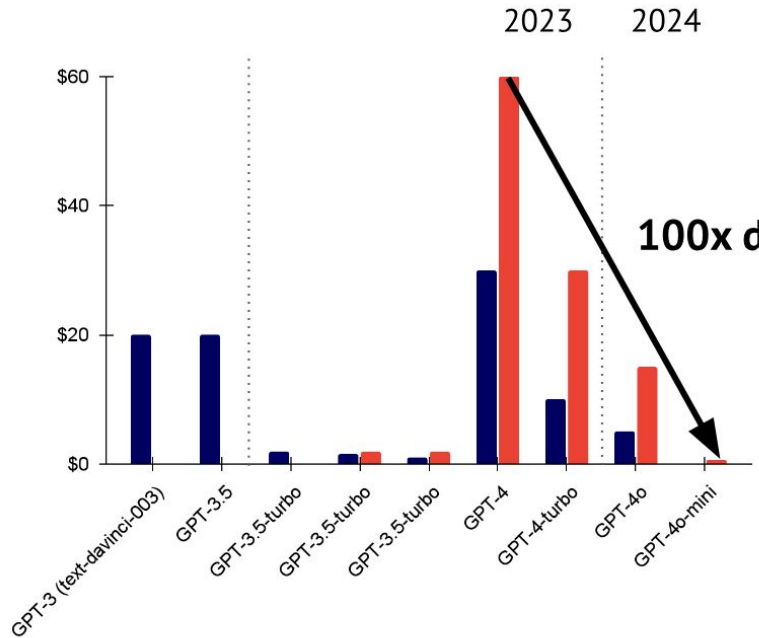
Ano, růžmarýn je aromatická bylina s jehličnatými listy a výraznou vůní. Používá se především jako koření do masových a rybích pokrmů, ale také do omáček, marinád a nápojů. Má léčivé vlastnosti, například podporuje trávení a působí antisepticky.

Není to náhodou rozmarýn?

Máte naprostou pravdu! Omlouvám se za tu

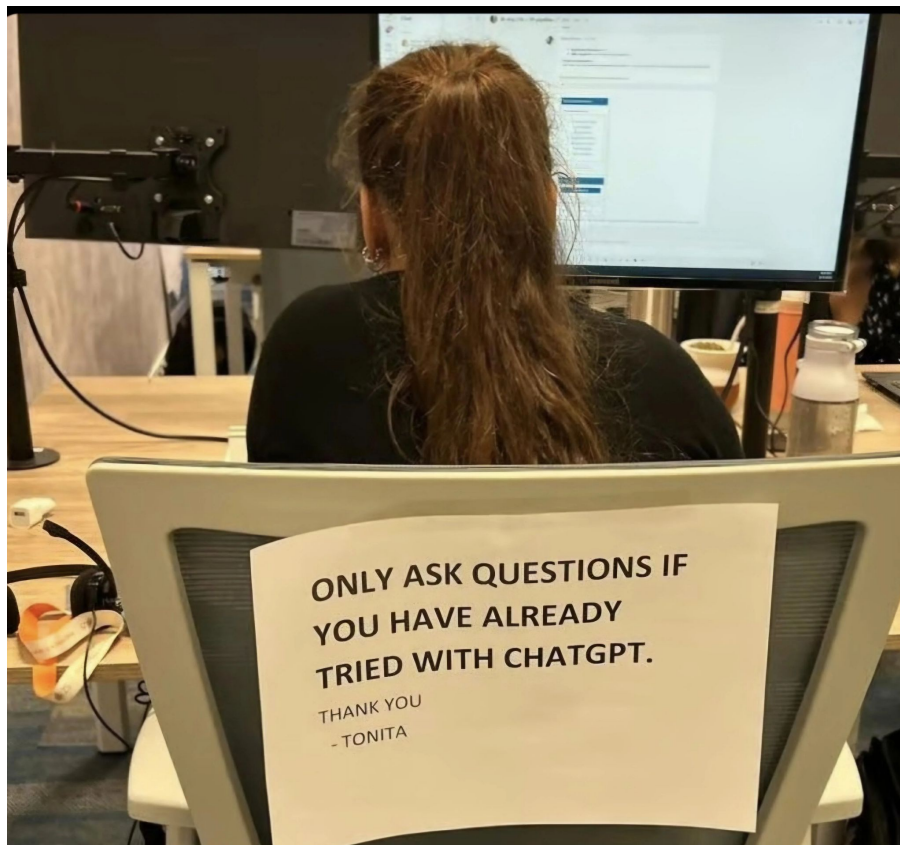
Inferencing all the way down: models get cheaper

Once thought to be prohibitively expensive to serve, the inference cost of serving strong models is dropping.



stateof.ai 2024

Velké jazykové modely se stávají součástí našich životů



AI Bullshit

Aplikace, do kterých byla implementovaná AI **jen proto aby se to dalo napsat do marketingových materiálů**. Ve výsledku zabírá prostor, zdroje, pozornost...



Run



Distance

6.21 km

Avg Pace

8:25 /km

Moving Time

52:18

Elevation Gain

32 m

Calories

509 Cal

Avg Heart Rate

146 bpm

 **Athlete Intelligence** BETA

Your run had a slower pace than usual, but you pushed into the anaerobic zone and reached your highest heart rate in 8 days, demonstrating strong fitness.

Say More

AI Bullshit



AI Bullshit



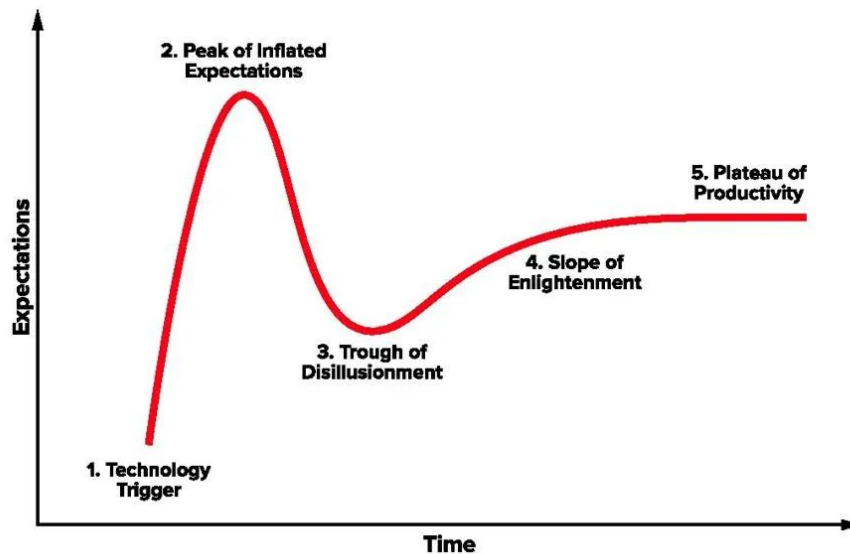
Hype-cycle

Image Classifier – SOTA klasifikátor obrázků, cvičení: vytvořte si pomocí DuckDuckGo Image Search malý dataset obrázků a dotrénujte si model natrénovaný na ImageNetu (ConvNeXT, ResNet)

Výsledky:

- **2018**: katastrofa (technické problémy)
- **2019-2021**: úspěch
- **2022**: nuda

Gartner Hype Cycle



Historiky z boje

mediaboard

Sumarizace

První nasazení velkého jazykového modelu – nad každým dokumentem shrnutí na 2-4 věty

Problémy:

- příliš jednoduchý model - špatné sumarizace
- špatný jazyk (sumarizace v angličtině)
- později dotrénován vlastní sumarizační model založený na **otevřeném modelu**

<https://bit.ly/praguellm>

Extraordinary meeting of the [redacted] Council held 27. 02. 2023

28. 2. 2023 13:58 - před 5 měsíci  →  Angličtina 

 Online [redacted]  Odkaz

   Ambivalent

PROCHÁZET KL. SLOVA

celkem 11 >

Extraordinary meeting of the [redacted] **Council** held 27. 02. 2023

in the office of the mayor of the town of [redacted]

order [PDF, 67 kB] vote [PDF, 86 kB]

Created / modified: 28.2.2023 / 28.2.2023 | Postal from: 1.3.2023

in the current context (**Extraordinary meeting** of [redacted])

including subordinate documents files

MEDIÁLNÍ DATA

 dosah 1.66 tis. lidí  1.9 tis. uživatelů za den  AVE: 2.23 tis. Kč  GRP: 0,02

 OTS: < 0.01

Summary:


On extraordinary meeting, the council... dismissed the mayor Ing. [redacted]

AI Sentiment

Požadavek přišel od obchodníka – a na první pohled to vypadalo jako jasný AI bullshit (většina byly komentáře na sociálních sítích)

Sentiment v Mediaboardu:

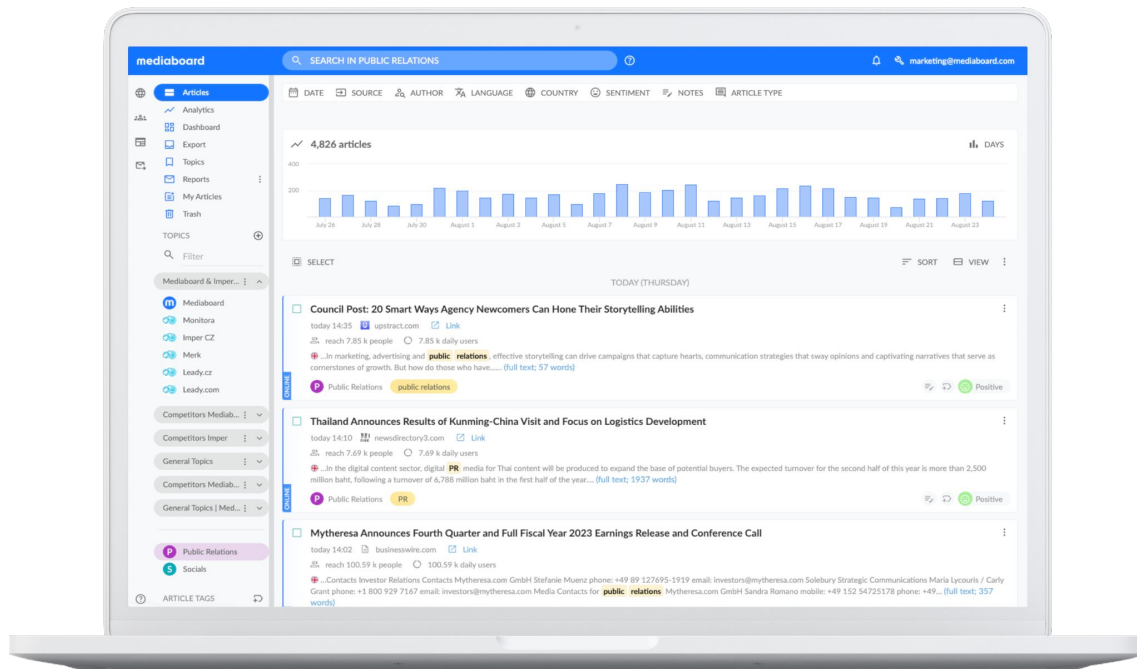
- dostaneme ho
- klasické NLP metody ("nic" nestojí)
- malý (transformer) model
- velký jazykový model

 = neutrální?

Kolik si myslíte, že stojí analýza článku velkým jazykovým modelem?

- gpt-4o
- gpt-4o-mini

Co velký jazykový model dokáže a co ne



Relevance, automatické štítkování

Co LLM dokáže a co ne



Co je metrika PRIME?

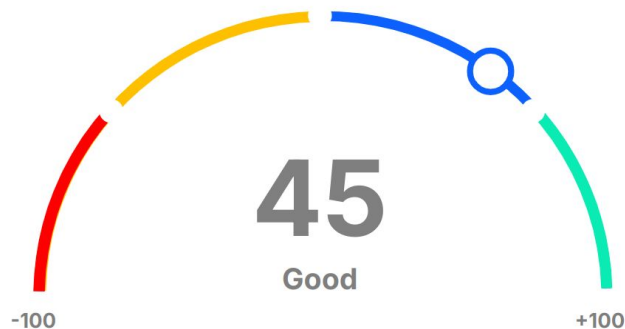
PRIME
PR Impact Measurement

PR Impact Measurement (PRIME) je námi vyvinutý inovativní systém pro **komplexní hodnocení mediálního dopadu PR aktivit za určité období.**

Zohledňuje **sentiment, relevanci, dosah a kvalitu mediálních výstupů** – umožňuje tak efektivně porovnávat dopad mediálních výstupů napříč různými kanály nebo ve srovnání **s konkurencí.**

Výsledky jednotlivých výstupů jsou sloučeny do **celkového skóre PRIME** za dané období.

Na rozdíl od tradičních metrik, jako je AVE a GRP, **PRIME respektuje Barcelonské principy** a lze ho přizpůsobit podle specifických cílů a potřeb klienta.



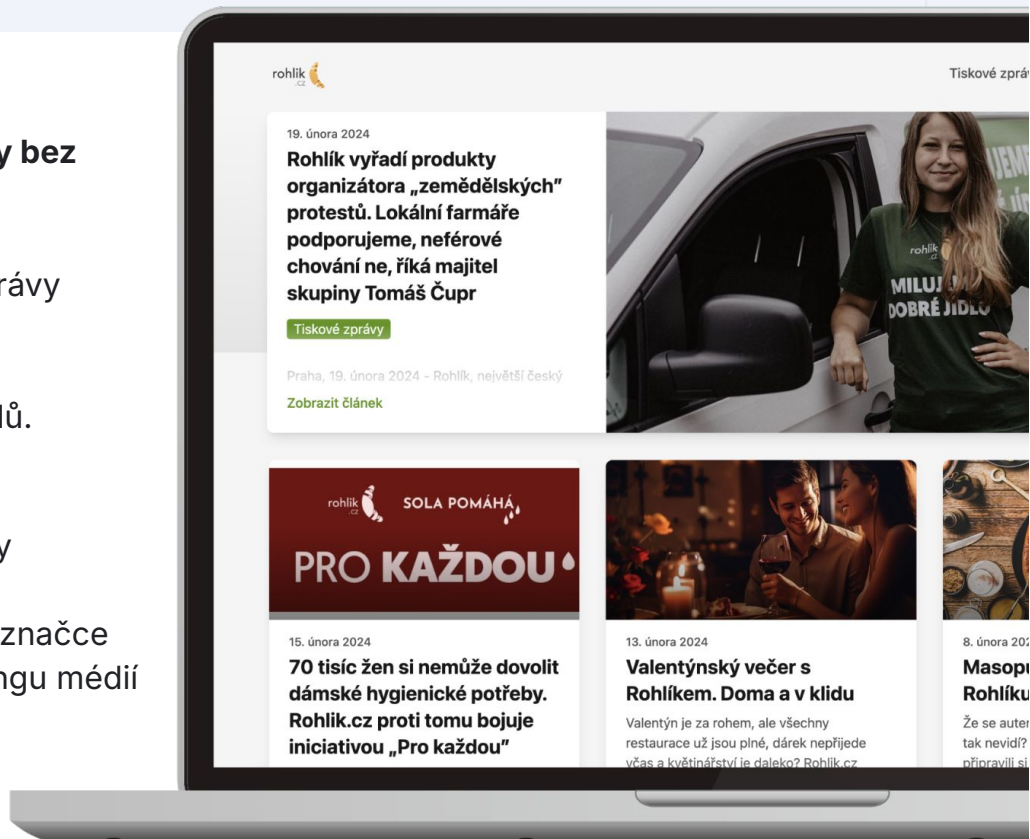
Newsroom

Tvořte, upravujte a distribuujte tiskové zprávy bez námahy.

Newsroom je navržen tak, aby vaše tiskové zprávy byly nejen **informačně bohaté**, ale i **vizuálně působivé**. Navíc je podpořen integrací s Medialistem a nástrojem pro rozesílání e-mailů.

Získáte:

- Prostor pro vaše příspěvky a tiskové zprávy
- Jednoduše ovladatelný redakční systém
- Možnost přizpůsobit design i doménu vaší značky
- Vaše příspěvky budou zahrnuty v monitoringu médií



Medialist

Který novinář se nejčastěji věnuje vám, vaší firmě, konkurenci či tématu, které vás zajímá?

Medialist – ověřený seznam novinářů a autorů je přímo v aplikaci – zjistíte, jak a kam cílit vaše tiskové zprávy.

Získáte:

- největší medialist autorů, novinářů a kontaktů
- aktualizované a ověřené informace
- detailní filtrace a vyhledávání
- možnost zapnout si **Globální Medialist**

The screenshot displays the Mediaboard interface. At the top, there is a search bar for authors. The main profile shown is for John Humphrey, an author at time4men.cz. The interface includes filters for tags (Time, Oblíbení, ekologie) and lists. A bar chart shows the number of articles published over time, with a peak in late July. Below the chart, there is a list of articles, with the top one titled "Plane wreckage diverts attention from Ukraine. Cruel Russian..." published yesterday at 07:51. The article snippet mentions Yevgeny Prigozhin and discusses the impact of the Ukraine conflict on media attention.

Create post



Main message & key points

Quotes

Main message & key points

Main message Key points list [+ ADD NEW KEYPOINT](#)[← PREVIOUS](#)[GENERATE STRUCTURE !\[\]\(84f47badaad7772cd95667a7c387a639_img.jpg\)](#)

08.10.2024

· Author

Košík.cz expanduje: Do konce roku na Slovensku v osmi městech

+ Add Category

+ Add Attachment

Košík.cz, přední český online supermarket, oznamuje svůj vstup na slovenský trh. Služby zahajuje v Bratislavě a Zvolenu a do konce roku plánuje rozšíření do dalších šesti měst. Tato expanze představuje významný krok v růstu společnosti a přináší nové možnosti pro slovenské zákazníky.

Košík.cz, jeden z největších online supermarketů v České republice, s hrdostí oznamuje svůj vstup na slovenský trh. Společnost zahajuje své služby v Bratislavě a Zvolenu, přičemž do konce roku plánuje rozšíření do dalších šesti měst, včetně Košic, Prešova a Žiliny. Tento krok je součástí

AI check

i AI Checkup information

This process may take a few moments as we perform the following steps:

1. **Initial scan:** Quickly reviewing the overall structure and format of your content.
2. **In-depth analysis:** Carefully examining the details, language, and context of your submission.
3. **Quality assessment:** Evaluating various aspects such as clarity, coherence, and relevance.
4. **Error detection:** Identifying any potential issues, inconsistencies, or areas for improvement.
5. **Optimization suggestions:** Preparing recommendations to enhance your content if needed.

Our AI is working diligently to provide you with accurate and helpful feedback. We appreciate your patience during this comprehensive analysis. The results will be available shortly.

? Data Support

The article lacks specific data, facts, or year-on-year comparisons to support the expansion claims.

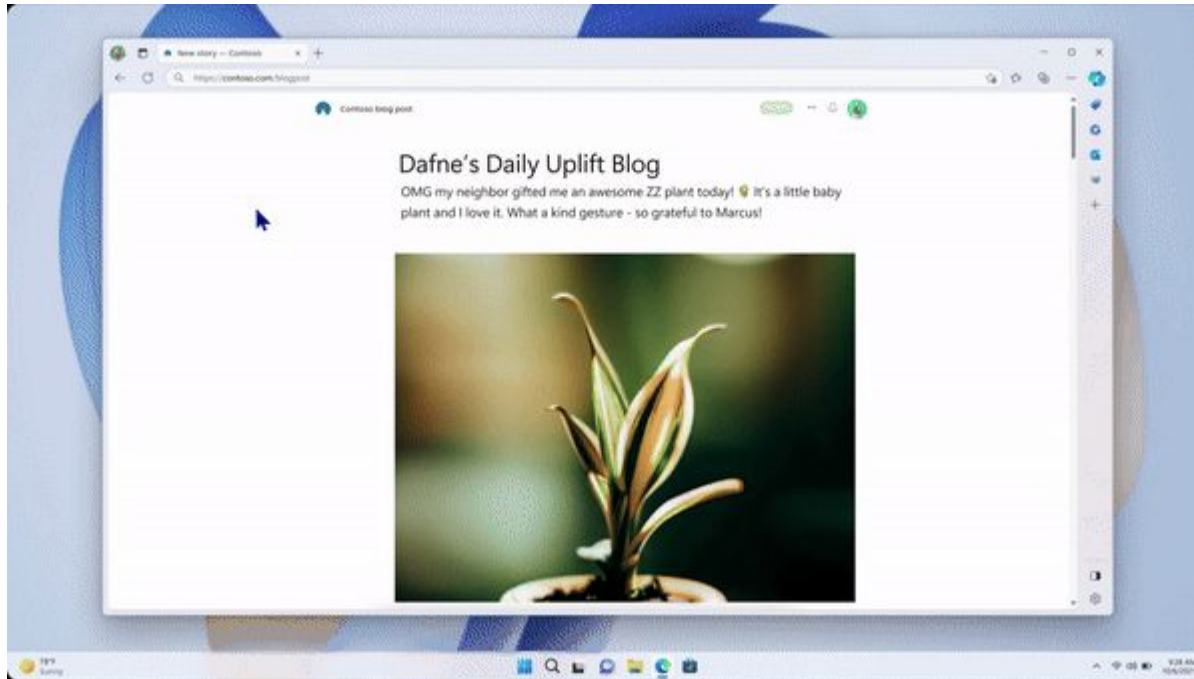
✓ Clarity

✓ Tone



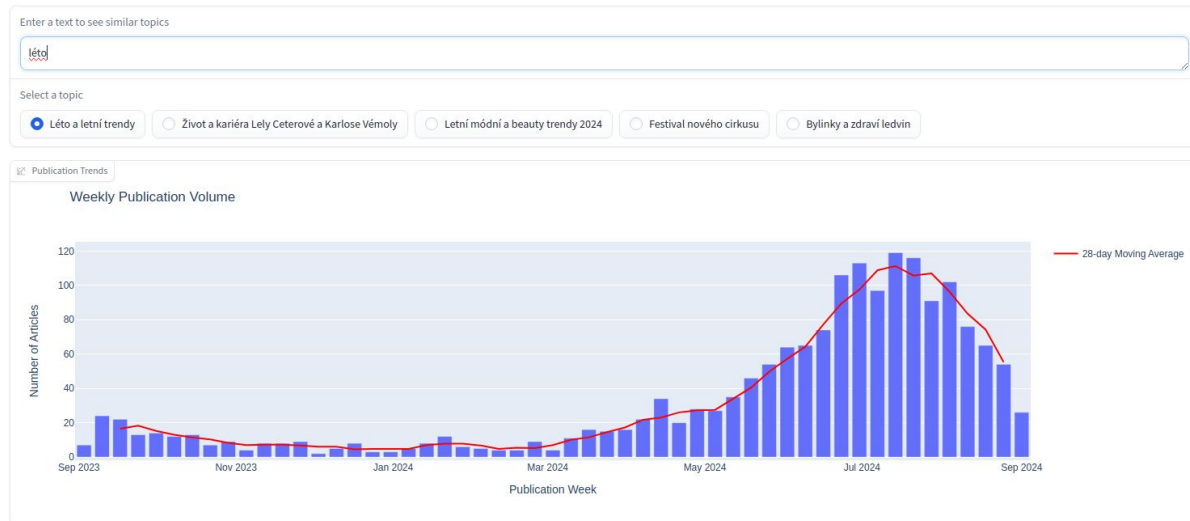


MS Edge Rewrite



Textové embeddingy (=vektorové reprezentace)

Topic Explorer



Articles

published	title	news_source_name	id
2024-08-09	Péče o vlasy v létě: vhodný šampon a další vychytávky, které vašim vlasům pomohou	vitalweb.cz	26165f
2024-08-04	Správné zalévání je v horkém létě klíčové pro zdraví rostlin a bohatou úrodu	pestujme.cz	25777f
2023-11-12	Jak uložit letní oblečení, na které teď pár měsíců nesáhnete? Máme pro vás několik tipů	onlyu.cz	13216f
2024-08-14	Žabky jsou oficiálně trendy! Po jakých letos sáhnout?	reflex.cz	26492f
2024-07-19	Letní šaty pro ženy nad 50. Tyto střihy vám udělají tělo dvacítky a každý muž se otočí	dotyk.cz	24638f
2024-08-15	Chcete mít doma letní atmosféru po celý rok? Nechte své stěny rozkvést	novinky.cz	26520f
2023-11-22	Období večerů v teple domova je tu: Vyberte si stylové teplákovky, župany i deky pro dokonalé pohodlí!	dama.cz	13315f

Benchmarky

mediaboard

MTEB: Massive Text Embedding Benchmark

English Chinese French Polish Russian

Overall MTEB English leaderboard 🤖

- Metric: Various, refer to task tabs
- Languages: English

Rank	Model	Model Size (Million Parameters)	Memory Usage (GB, fp32)	Embedding Dimensions	Max Tokens	Average (56 datasets)	Classification Average (12 datasets)	Clustering Average (11 datasets)
1	NV-Embed-v2	7851	29.25	4096	32768	72.31	90.37	58.46
2	bge-en-ic1	7111	26.49	4096	32768	71.67	88.95	57.89
3	stella_en_1.5B_v5	1543	5.75	8192	131072	71.19	87.63	57.69
4	SFR-Embedding-2_R	7111	26.49	4096	32768	70.31	89.05	56.17
5	gte-Qwen2-7B-instruct	7613	28.36	3584	131072	70.24	86.58	56.92
6	stella_en_400M_v5	435	1.62	8192	8192	70.11	86.67	56.7
7	dunzhang-stella_en_400M_v5	435	1.62	1024	8192	70.11	86.67	56.7
8	bge-multilingual-gemma2	9242	34.43	3584	8192	69.88	88.08	54.65
9	NV-Embed-v1	7851	29.25	4096	32768	69.32	87.35	52.8
10	voyage-large-2-instruct			1024	16000	68.23	81.49	53.35

MTEB: Massive Text Embedding Benchmark

English Chinese French **Polish** Russian

Overall MTEB English leaderboard 🤖

- o Metric: Various, refer to task tabs
- o Languages: English

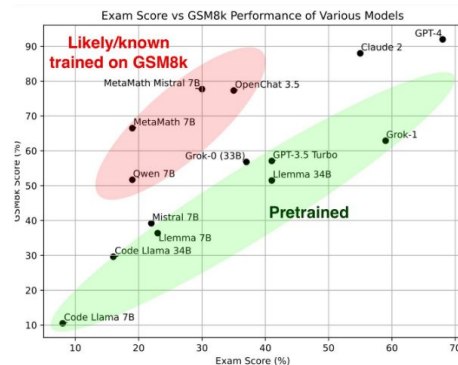
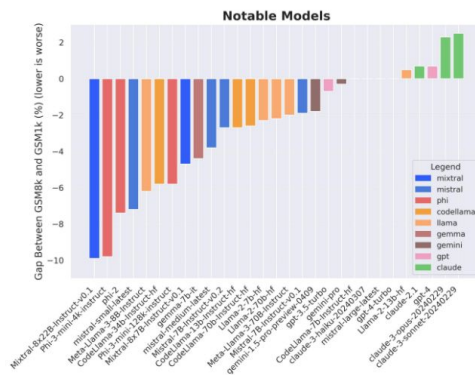
Rank ▲	Model ▲	Model Size (Million Parameters) ▲	Memory Usage (GB, fp32) ▲	Embedding Dimensions ▲	Max Tokens ▲	Average (56 datasets) ▲	Classification Average (12 datasets) ▲	Clustering Average (11 datasets)
1	NV-Embed-v2	7851	29.25	4096	32768	72.31	90.37	58.46
2	bge-en-ic1	7111	26.49	4096	32768	71.67	88.95	57.89
3	stella_en_1.5B_v5	1543	5.75	8192	131072	71.19	87.63	57.69
4	SFR-Embedding-2_R	7111	26.49	4096	32768	70.31	89.05	56.17
5	gte-Qwen2-7B-instruct	7613	28.36	3584	131072	70.24	86.58	56.92
6	stella_en_400M_v5	435	1.62	8192	8192	70.11	86.67	56.7
7	dunzhang-stella_en_400M_v5	435	1.62	1024	8192	70.11	86.67	56.7
8	bge-multilingual-gemma2	9242	34.43	3584	8192	69.88	88.08	54.65
9	NV-Embed-v1	7851	29.25	4096	32768	69.32	87.35	52.8
10	voyage-large-2-instruct			1024	16000	68.23	81.49	53.35
11	NV-Embed-Mini-1	7111	26.49	4096	32768	68.17	86.6	51.46

Leaking Benchmark Datasets

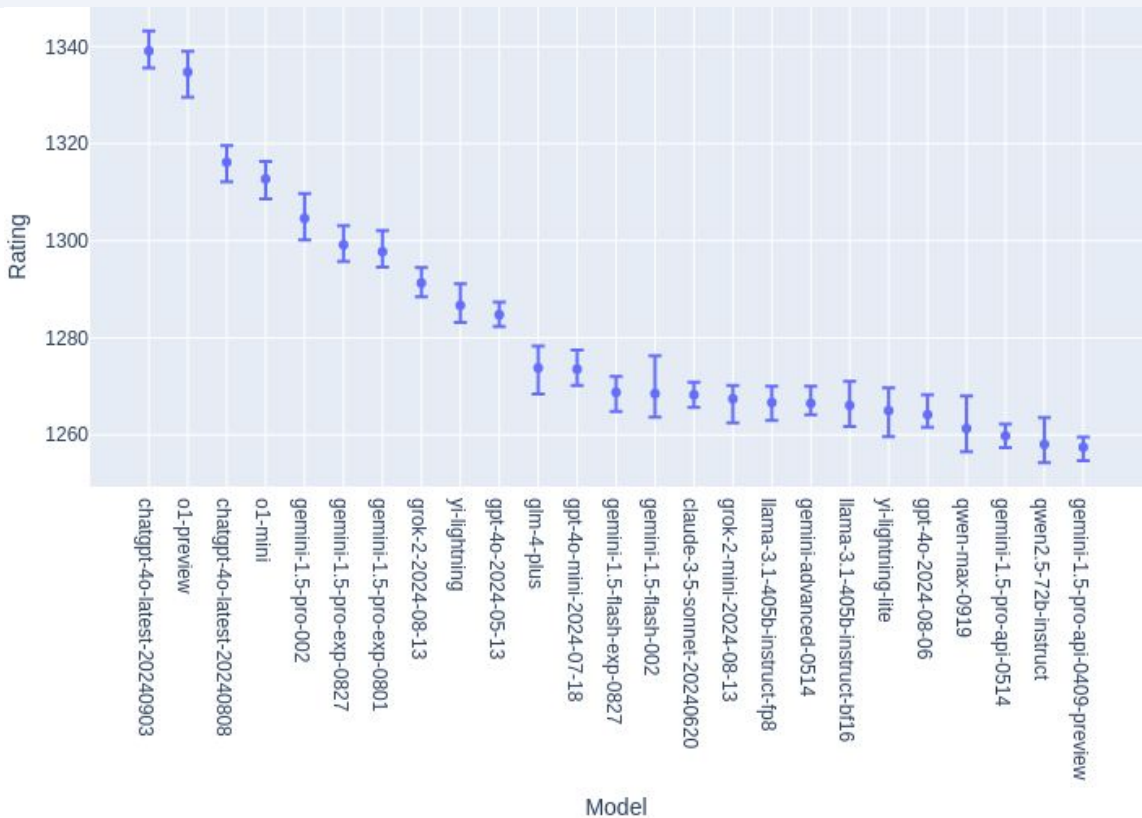


Is contamination inflating progress?

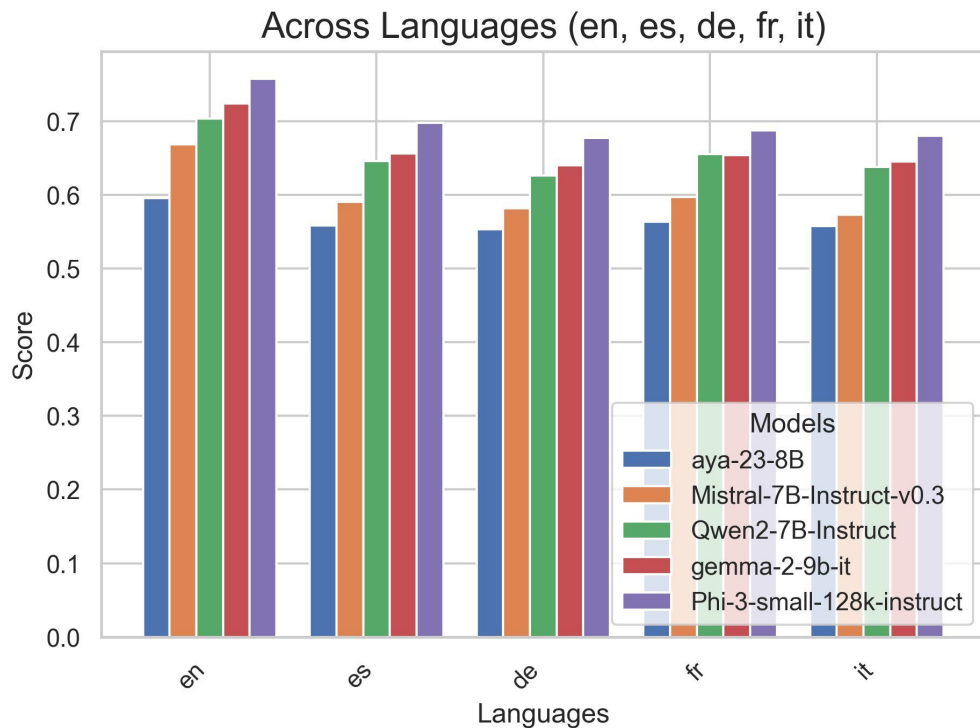
- ▶ With new model families reporting incredibly strong benchmark performance straight out-of-the-gate, researchers have increasingly been shining a light on dataset contamination: when test or validation data leaks into the training set. Researchers from Scale retested a number of models on a new Grade School Math 1000 (GSM1k) that mirrors the style and complexity of the established GSM8k benchmark, finding significant performance drops in some cases. Similarly, researchers at X.ai re-evaluated models using a dataset based on the Hungarian national finals math exam that post-dated their release, with similar results.



LMSYS = slepý test dvou modelů



ARC, Belebele, TruthfulQA, M-MMLU



CzechBench (Adam Jirkovský, CIIRC)

Model ▲	Precision ▲	Aggregate Score ▼	Grammar (Avg.) ▲	Knowledge (Avg.) ▲	Reasoning (Avg.) ▲	Math (Avg.) ▲
claude-3-opus-20240229	other	79.07	92.98	87.78	75.98	59.83
gpt-4o-2024-08-06	other	78.62	93.46	84.74	75.84	58.32
claude-3-5-sonnet-202406	other	77.69	93.94	91.78	76.83	46.47
Meta-Llama-3.1-70B-Instr	bfloat16	75.11	75.92	86.05	75.33	60.11
gemma-2-27b-it	bfloat16	73.57	78.15	81.44	75.49	54.29
gpt-4o-mini-2024-07-18	other	72.31	76.87	80.57	70.77	54.33
claude-3-haiku-20240307	other	70.32	82.14	74.36	69.13	51.82
gemma-2-9b-it	bfloat16	65.62	65.87	74.38	73.05	39.16
Meta-Llama-3.1-8B-Instru	bfloat16	55.19	53.11	59.37	66.81	20.81
gemma-2-2b-it	bfloat16	49.97	44.66	53.7	58.76	24.82

<https://gitlab.com/jirkooda/czech-bench>



BenCzechMark (Martin Fajčík, BUT)

Model	Release	Type	# 0 (B)	Average	Czech Language Understanding	Czech Math Reasoning	F
meta-llama Meta-Llama-3.1-405B-Instruct	2024-09-23	chat	406	93.959	99.444	96.667	9
meta-llama Meta-Llama-3.1-70B-Instruct	2024-09-23	chat	70.6	84.514	83.333	67.5	8
meta-llama Meta-Llama-3.1-70B	2024-09-23	pretrained	70.6	82.043	75.556	60.833	8
Qwen Qwen2-72B-Instruct	2024-09-23	chat	72.7	79.389	77.778	90.833	8
mistralai Mixtral-8x22B-Instruct-v0.1	2024-09-23	chat	141	77.478	74.444	65.833	7
meta-llama Meta-Llama-3.1-8B-Instruct	2024-09-23	chat	8.03	59.335	55	20.833	5
mistralai Mistral-Nemo-Instruct-2407	2024-09-23	chat	12.2	58.877	54.444	20	5
CohereForAI aya-23-35B	2024-09-23	chat	35	58.436	68.333	13.333	4
mistralai Mixtral-8x7B-Instruct-v0.1	2024-09-23	chat	46.7	57.497	58.333	40.833	5
spokeleash Bielik-11B-v2.3-Instruct	2024-10-05	chat	11.2	56.844	52.222	35	5
NousResearch Hermes-3-Llama-3.1-8B	2024-09-23	chat	8.03	51.481	53.333	13.333	4
meta-llama Meta-Llama-3-8B-Instruct	2024-09-23	chat	8.03	49.838	53.333	13.333	4
Qwen Qwen2-7B-Instruct	2024-09-23	chat	7.62	49.134	45	52.5	4
google gemma-2-9b-it	2024-09-23	chat	9.24	47.651	47.222	45	3
meta-llama Meta-Llama-3.1-8B	2024-09-23	pretrained	8.03	47.502	42.222	13.333	4

<https://huggingface.co/spaces/CZLC/BenCzechMark>



Rychlá evaluace

$$n = \frac{Z^2 \cdot p \cdot (1 - p)}{E^2} = \frac{(1.645)^2 \cdot 0.25 \cdot (1 - 0.25)}{0.05^2} \approx 200$$

Na rychlé otestování nám stačí **cca 200 otázek**

Na základě podobnosti výsledků **4 vybrané dataseety** z CzechBench

- AGREE (čeština)
- CzechNews (kategorizace zpráv)
- Klokánek (math)
- CTKFacts (reasoning)

<https://github.com/simecek/MiniCzechBenchmark>

Rychlá evaluace 2

Příklad modelu dotrénovaného na sentiment + zdůvodnění

Sentiment tohoto textu vzhledem ke bance je NEGATIVNÍ.

Text popisuje, jak insolvenční správkyň Sberbank musela provozovat banku pro někoho jiného a upínat síly k tomu, aby banka fungovala. Také se zmiňuje o vysoké ceně, kterou Česká spořitelna dala za banku, a o tom, že insolvenční správkyň si na sebe ušila bič. Tyto informace naznačují, že situace s bankou je obtížná a negativní.

+ elegance generované češtiny

Pomocí (velmi) velkého otevřeného modelu nebo manuálně

Níže dostaneš článek a sumarizaci. Tvým úkolem je ohodnotit, jak dobrá sumarizace je na stupnici 0-20. Postupně prosím vyhodnoť následující kritéria:

- 1) Důležitá fakta (10 bodů): jsou v sumarizaci hlavní body z článku, jsou zmíněná čísla nebo je text sumarizace balast
- 2) Gramatické chyby (5 bodů): Soustřeď se na skloňování, shodu přísudku s podnětem...
- 3) Elegance jazyka (5 bodů): Je text příjemný na čtení? Nejsou věty příliš kostrbaté a neohrabané?

ČLÁNEK

...

SUMARIZACE

...

+ elegance generovan

Pomocí (velmi) velkého otevřeného modelu

Níže dostaneš článek a sumarizaci. Ty je na stupnici 0-20. Postupně prosím

- 1) Důležitá fakta (10 bodů): jsou v s čísla nebo je text sumarizace balast
- 2) Gramatické chyby (5 bodů): Soustře podnětem...
- 3) Elegance jazyka (5 bodů): Je text kostrbaté a neohrabané?

ČLÁNEK

...

SUMARI

...

Trik: ne jedna odpověď, ale pravděpodobnosti všech

```
Perplexity Ratings Generator

def get_completion(messages, model="gpt-4o-mini", temperature=0, logprobs=True,
top_logprobs=10):
    return client.chat.completions.create(model=model, messages=messages,
temperature=temperature, logprobs=logprobs, top_logprobs=top_logprobs)

class HumorScorer:
    def __init__(self, model="gpt-4o-mini"):
        self.model = model
        self.categories = list(range(1, 11))

    def score(self, text):
        prompt = f"On a scale of 1 to 10, where 1 is not funny at all and 10 is
extremely funny, rate the humor level of the following text:\n\nText:
{text}\n\nHumor Rating:"
        response = get_completion([{"role": "user", "content": prompt}],
model=self.model)

        logprobs = {str(cat): -float('inf') for cat in self.categories}
        for token_info in response.choices[0].logprobs.content:
            for logprob in token_info.top_logprobs:
                token = logprob.token.strip()
                if token in logprobs:
                    logprobs[token] = max(logprobs[token], logprob.logprob)

        probs = {cat: np.exp(lp) for cat, lp in logprobs.items()}
        total_prob = sum(probs.values())
        normalized_probs = {cat: p / total_prob for cat, p in probs.items()}
        weighted_score = sum(int(cat) * prob for cat, prob in
normalized_probs.items())
        final_score = (weighted_score - 1) / 9

        print("Category probabilities:", normalized_probs)
        return round(final_score, 3)
```

Michael Ryaboy

Protože bez dat jsem jen další člověk s názorem.

Petr Šimeček

Research & Development, Mediaboard

LinkedIn: <https://www.linkedin.com/in/simecekpetr/>

X / Twitter: @python_tip

Web: www.mediaboard.com/cz

