

تفکر انتقادی در عصر هوش مصنوعی

نام استاد راهنما:

پژوهشگران:

چطور پاسخ‌های روان را به ادعاهای قابل بررسی تبدیل کنیم؟

تغییر پارادایم از «جستجوی اطلاعات» به «ارزیابی اطلاعات»



در گذشته چالش دانشجو یا پژوهشگر «پیدا کردن» اطلاعات بود



امروزه هوش مصنوعی اقیانوسی از پاسخ‌های حاضر و آماده تحویل می‌دهد



مهارت اصلی دیگر جمع‌آوری نیست، بلکه پالایش و داوری است



هوش مصنوعی پاسخ می‌دهد؛ اما لزوماً حقیقت را نمی‌گوید

روان بودن، مدرک نیست

- اعتمادبه‌نفس و روان بودن پاسخ، نشانهٔ درستی آن نیست.
- Fluency Heuristic
- هوش مصنوعی ذاتاً جوری آموزش دیده که آرام، متین، بدون لکنت و با اعتمادبه‌نفس بنویسد حتی وقتی کاملاً در اشتباه است.

داوری با انسان است

داوری نهایی ، اعتباربخشی و عواقب استفاده از آن پژوهش یا گزارش، روی شانهٔ انسان است.

۳. چرا این اتفاق می‌افتد؟

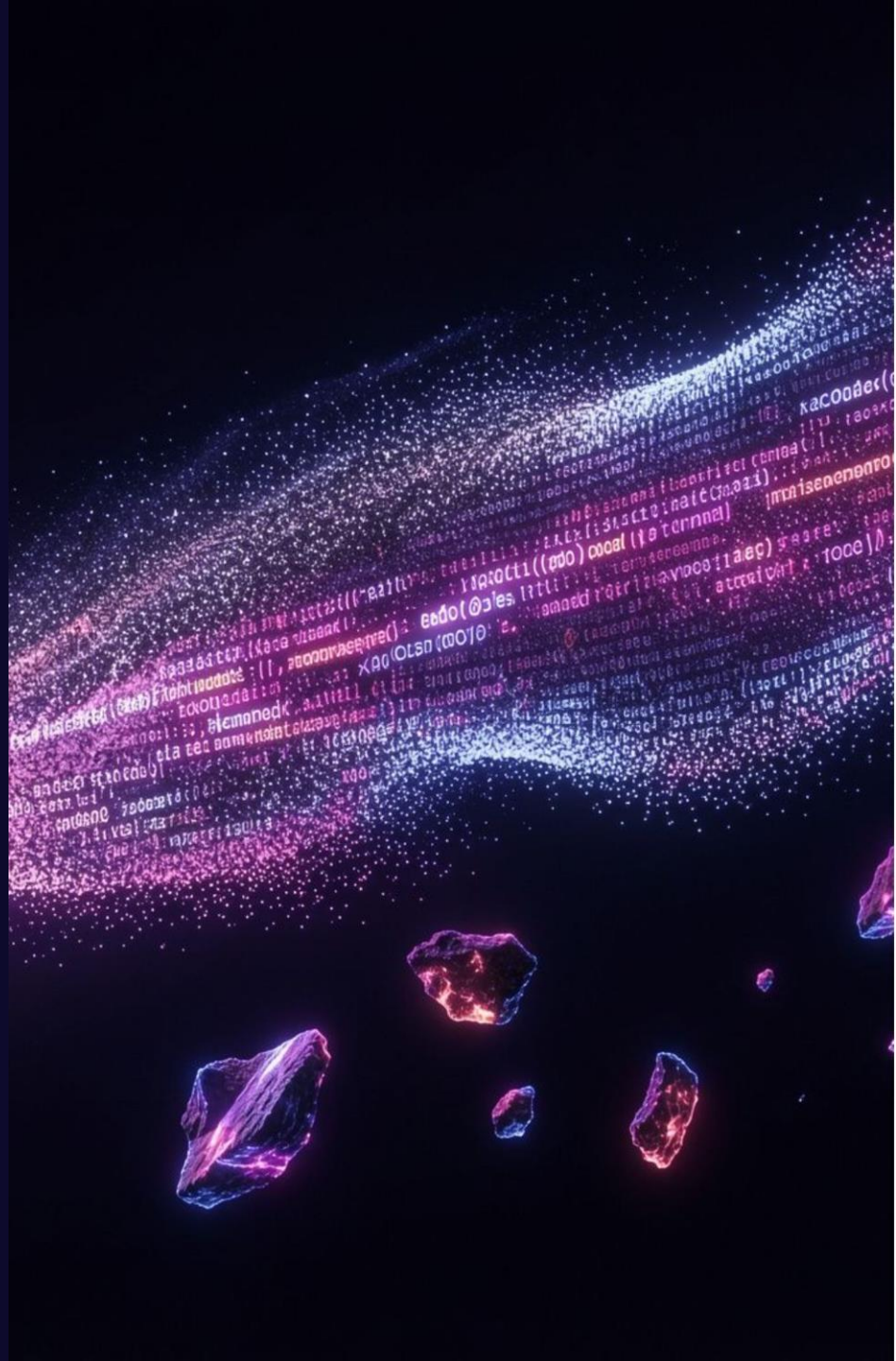
مکانیسم توهم یا Hallucination

پاسخ نادرست می‌تواند از کنار واقعیت عبور کند

مدل‌های زبانی الگوهای واژه‌ها را پیش‌بینی می‌کنند؛ این به‌معنای فهمیدن واقعیت نیست.

هدفشان ساختن جمله ای است که طبیعی و متقاعدکننده رفتار کند.

گاهی جزئیات ساختگی را کنار واقعیت‌های درست می‌گذارند. همین ترکیب، پاسخ را قانع‌کننده می‌کند.



سه خطای رایج هنگام استفاده از هوش مصنوعی

پذیرش فوری

- وقتی سوال را طوری می‌پرسیم که جهت فکری مان لو برود
- پاسخ را بدون بررسی می‌پذیریم.
- هوش مصنوعی ذاتاً تمایل به جلب رضایت کاربر (Sycophancy) دارد.

ارجاعِ بی‌بررسی

- به نام منبع یا لینک اعتماد می‌کنیم.
- **این مدل‌ها می‌توانند مقالات خیالی با فرمت DOI و اسامی اساتید واقعی بسازند که اصلاً در دنیا چاپ نشده‌اند.**

پرسش جهت‌دار

سؤالی می‌پرسیم که فقط تأیید می‌خواهد.

۵. گام اول

ادعاها را از هم جدا کن

یک پاسخ بلند را به چند سؤال کوچک و قابل بررسی تبدیل کن. (Deconstruction)

1

تقسیم کن

هر پاسخ را به ادعاهای جداگانه تقسیم کن.

2

نوع را بپرس

واقعیت است، تفسیر است یا حدس؟

3

مهم‌ها را علامت بزن

مواردی را که روی تصمیم اثر دارند، اول بررسی کن.
قانون پارتو (80 – 20)

۶. گام دوم

منبع را بررسی کن، نه فقط نام آن را

نام دانشگاه، مجله یا لینک، به‌تنهایی اعتبار نمی‌سازد.



آیا منبع واقعاً وجود دارد؟

عنوان، لینک و متن اصلی را پیدا کن.



آیا منبع از ادعای مطرح‌شده پشتیبانی می‌کند؟

منبع را باز کن و بخش مرتبط را بخوان.



تاریخ، نویسنده، اعتبار و ارتباط منبع را بسنج

منبع معتبرِ نامرتبط، پاسخ پرسش تو نیست.

۷. یک روش چهارمرحله‌ای

روش SIFT برای مکث و راستی‌آزمایی

یکی از معروف‌ترین استانداردها برای سنجش اطلاعات در فضای دیجیتال

1 مکث (Stop)

قبل از باور یا بازنشر توقف کن.

2 بررسی منبع (Investigate the source)

سازنده و سوگیری‌های احتمالی را بشناس.

3 یافتن پوشش معتبر (Find better coverage)

گزارش‌های مستقل را مقایسه کن.

4 ردیابی ادعا (Trace claims to the original context)

به منبع اصلی برگرد. (بیشتر تحریف‌ها در نقل‌قول‌ها رخ می‌دهد)



۸. آزمون مناسب را انتخاب کن

پاسخ خوب، برای هر نوع مسئله یک آزمون دارد

تفاوت روش‌شناسی ارزیابی بر اساس جنس داده‌ها (Methodology Matching)

1

عدد و محاسبه

دوباره حساب کن.

2

کد

اجرا کن و با مستندات مقایسه کن.

3

پژوهش

با منابع مستقل و معتبر تطبیق بده.

4

خبر و تصویر

منشأ و زمینه را پیدا کن.

پرسش‌های بهتر، پاسخ‌های قابل بررسی‌تر

مهندسی پرامپت نقادانه



ادعاها را جدا کن

از مدل بخواه ادعاها را جدا و فرض‌ها را آشکار کند.



منبع و اطمینان بخواه

برای هر ادعا منبع و میزان اطمینان بخواه.



نقش منتقد بده

نقش منتقد را به آن بده، اما داوری نهایی را خودت انجام بده.

جمع‌بندی

تفکر انتقادی یعنی چه چیزی را نپذیریم؟



قطعیتِ بی‌منبع

ارجاعِ ساختگی یا نامرتبط

پاسخی که فقط با پیش‌فرض ما موافق است

تصمیمی که پیامدهایش بررسی نشده است

منابع

- تدتاک - ادوایت سرکار
- ام‌ای‌تی هورایزن
- مدیوم

با تشکر از توجه شما