

آسیب های استفاده افراطی از ابزارهای هوش مصنوعی

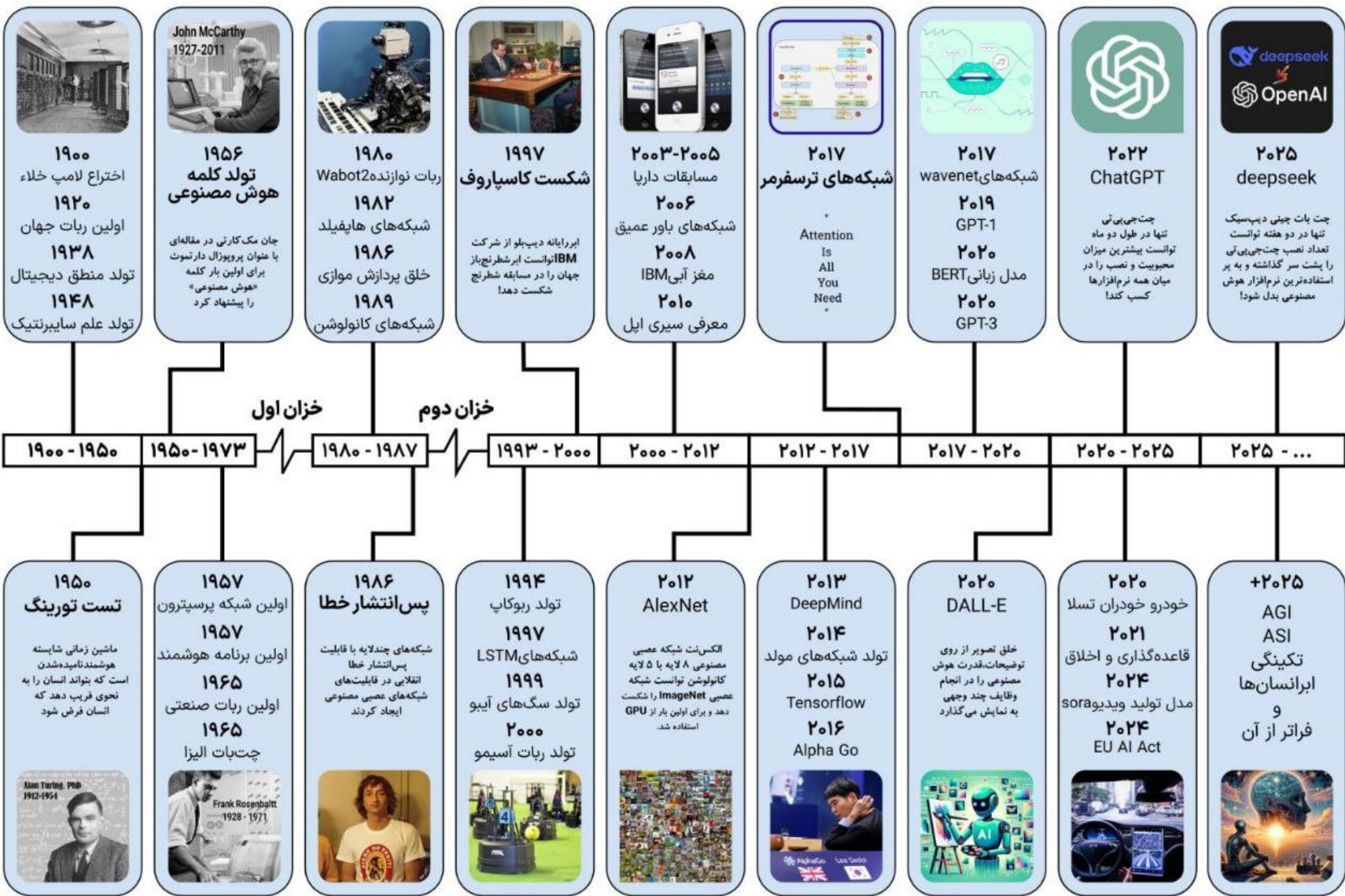
مرور آسیب های پرتکرار و ارائه روشهای برای کاهش آسیب یا هزینه آن

مرور و یادآوری نکات اصلی

هوش مصنوعی

تکامل فناوری







یان لوکون، جفری هینتون، اندرو نگ، یوشوا بنگیو، دمیس هاسابیس، فی-فی-لی،
ریچارد سوکای، پیتروویگ، دیوید سیلور، مصطفی سلیمان
ده دانشمند تراز اول هوش مصنوعی

ده دانشمند برجسته هوش مصنوعی

1. یان لوکون (Yann LeCun): یکی از پیشگامان یادگیری عمیق و شبکه‌های عصبی کانولوشنی، هم‌اکنون در Meta AI فعالیت می‌کند.
[https://en.wikipedia.org/wiki/Yann_LeCun](https://en.wikipedia.org/wiki/Yann_LeCun)
2. جفری هینتون (Geoffrey Hinton): پدر یادگیری عمیق و استاد برجسته دانشگاه تورنتو، نقش کلیدی در توسعه الگوریتم‌های شبکه عصبی دارد.
[https://en.wikipedia.org/wiki/Geoffrey_Hinton](https://en.wikipedia.org/wiki/Geoffrey_Hinton)
3. اندرو نگ (Andrew Ng): بنیان‌گذار Coursera و رهبر پروژه‌های هوش مصنوعی در Google و Baidu، متخصص در یادگیری ماشینی است.
[https://en.wikipedia.org/wiki/Andrew_Ng](https://en.wikipedia.org/wiki/Andrew_Ng)
4. یوشوا بنگیو (Yoshua Bengio): پژوهشگر برجسته در یادگیری عمیق و استاد دانشگاه مونترال، از برندگان جایزه تورینگ است.
[https://en.wikipedia.org/wiki/Yoshua_Bengio](https://en.wikipedia.org/wiki/Yoshua_Bengio)
5. دمیس هاسابیس (Demis Hassabis): بنیان‌گذار DeepMind و متخصص در هوش مصنوعی و علوم اعصاب، رهبری پروژه AlphaGo را بر عهده داشت.
[https://en.wikipedia.org/wiki/Demis_Hassabis](https://en.wikipedia.org/wiki/Demis_Hassabis)
6. فیلپ شامیت (Fei-Fei Li): پیشگام در بینایی کامپیوتری و بنیان‌گذار ImageNet، هم‌اکنون در استنفورد تدریس می‌کند.
[https://en.wikipedia.org/wiki/Fei-Fei_Li](https://en.wikipedia.org/wiki/Fei-Fei_Li)
7. ریچارد سوکای (Richard Socher): متخصص در پردازش زبان طبیعی و بنیان‌گذار MetaMind، هم‌اکنون در Salesforce فعالیت دارد.
[https://en.wikipedia.org/wiki/Richard_Socher](https://en.wikipedia.org/wiki/Richard_Socher)
8. پیتر نورویگ (Peter Norvig): مدیر تحقیقاتی Google و نویسنده کتاب‌های برجسته در هوش مصنوعی، از پیشگامان این حوزه است.
[https://en.wikipedia.org/wiki/Peter_Norvig](https://en.wikipedia.org/wiki/Peter_Norvig)
9. دیوید سیلور (David Silver): محقق ارشد DeepMind و توسعه‌دهنده الگوریتم‌های تقویت یادگیری، از جمله AlphaGo.
[[https://en.wikipedia.org/wiki/David_Silver_\(computer_scientist\)](https://en.wikipedia.org/wiki/David_Silver_(computer_scientist))]([https://en.wikipedia.org/wiki/David_Silver_\(computer_scientist\)](https://en.wikipedia.org/wiki/David_Silver_(computer_scientist)))
10. مصطفی سلیمان (Mustafa Suleyman): یکی از بنیان‌گذاران DeepMind و Inflection AI، از رهبران نوآوری در هوش مصنوعی است.
[https://en.wikipedia.org/wiki/Mustafa_Suleyman](https://en.wikipedia.org/wiki/Mustafa_Suleyman)

مفاهیم پایه ...

هوش مصنوعی AI شاخه‌ای از علوم رایانه است که به طراحی سیستم‌ها و ماشین‌هایی می‌پردازد که قادر به انجام وظایفی هستند که معمولاً به هوش انسانی نیاز دارند.

هوش مصنوعی محدود Narrow AI نوعی از هوش مصنوعی است که برای انجام یک کار خاص طراحی شده و نمی‌تواند خارج از آن حوزه عمل کند.

هوش مصنوعی مولد Generative AI نوعی از هوش مصنوعی است که قادر به ایجاد داده‌های جدید مانند متن، تصاویر، موسیقی و ویدیو می‌باشد.

هوش مصنوعی عمومی AGI نوعی فرضی از هوش مصنوعی است که قادر به انجام هر کار فکری است که یک انسان می‌تواند انجام دهد.

شبکه عصبی Neural Network مدلی محاسباتی است که از ساختار مغز انسان الهام گرفته و شامل گره‌های متصل به هم می‌باشد. این شبکه‌ها قادر به یادگیری ویژگی‌های پیچیده در داده‌ها هستند .

یادگیری عمیق DL زیرمجموعه‌ای از یادگیری ماشین است که از شبکه‌های عصبی مصنوعی با لایه‌های متعدد برای تجزیه و تحلیل داده‌ها استفاده می‌کند. این روش برای پردازش داده‌های پیچیده مانند تصاویر و صدا کاربرد دارد.

مفاهیم پایه ...

مدل زبانی Language Model مدلی است که برای درک و تولید زبان طبیعی آموزش داده شده است. این مدل‌ها به ماشین‌ها این امکان را می‌دهند که زبان انسانی را پردازش و تولید کنند.

مدل‌های زبانی بزرگ LLM مدل‌های زبانی با تعداد زیادی پارامتر که قادر به انجام وظایف پیچیده زبانی مانند ترجمه، خلاصه‌سازی و ایجاد متن هستند.

توکن Token کوچک‌ترین واحد داده‌ای است که یک مدل زبانی می‌تواند آن را پردازش کند. این واحد معمولاً کلمات، علائم نگارشی یا ترکیب‌هایی از این‌ها هستند.

داده‌های آموزشی Training Data داده‌هایی هستند که برای آموزش یک مدل یادگیری ماشین استفاده می‌شوند. این داده‌ها به مدل کمک می‌کنند تا الگوها و روابط موجود در داده‌ها را یاد بگیرد .

مفاهیم پایه ...

الگوریتم Algorithm مجموعه‌ای از دستورالعمل‌ها است که برای حل یک مشکل یا انجام یک کار باید به ترتیب دنبال شوند.

پرامپت Prompt دستور یا سوالاتی هستند که از مدل درخواست می‌شود تا به آنها پاسخ دهد.

چت‌بات Chatbot برنامه‌ای است که قادر به انجام مکالمات متنی یا صوتی با انسان می‌باشد. این ابزار از مدل‌های زبانی برای تعامل با کاربران استفاده می‌کند.

ابزارهای هوش مصنوعی AI Tools نرم‌افزارها و برنامه‌هایی هستند که از تکنولوژی هوش مصنوعی برای انجام وظایف مختلف استفاده می‌کنند. ابزارهای AI ممکن است شامل مدل‌ها، پلتفرم‌ها، یا برنامه‌های خاصی برای تحلیل داده‌ها، تولید محتوا و سایر کاربردها باشند.

مهندسی پرامپت Prompt Engineering هنر و علم طراحی پرامپت‌ها یا ورودی‌های متنی مؤثر است تا نتایج مطلوب و دقیقی از مدل‌های زبانی به دست آید

مدل، ابزار و حساب کاربری

مفهوم	تعریف ساده و کوتاه	مثال‌های معروف و قابل درک
مدل Model	مغز هوش مصنوعی؛ سیستم یادگیرنده	DeepSeek, Gemini, ChatGPT
ابزار Tool	نرم‌افزار یا برنامه‌ای برای ایجاد، مدیریت یا استفاده از محتوای هوش مصنوعی	برای ساخت ارائه‌های هوشمند و جذاب Google Classroom برای آموزش مجازی Canvas برای طراحی و ساخت محتوا
ایجنت Agent	سیستم هوشمند که مستقل تصمیم می‌گیرد و عمل می‌کند	Siri, Alexa, Google Assistant
ویجت Widget	رابط گرافیکی کوچک برای تعامل راحت کاربر با سیستم یا مدل	پنجره چت روی سایت‌های مختلف، دکمه ترجمه گوگل در مرورگر

مدل، ابزار و اکانت

مدل مدل‌ها سیستم‌هایی هستند که بر اساس داده‌ها آموزش می‌بینند تا وظایف خاصی مانند پردازش زبان طبیعی، شبیه‌سازی رفتار انسان، یا پیش‌بینی داده‌ها را انجام دهند. مدل‌ها می‌توانند هوش مصنوعی، یادگیری ماشین یا یادگیری عمیق باشند. برای مثال، مدل‌های زبانی مانند GPT-3 و BERT.

ابزار ابزارها نرم‌افزارهایی هستند که از مدل‌ها و تکنیک‌های هوش مصنوعی برای انجام وظایف مختلف استفاده می‌کنند. این ابزارها می‌توانند شامل پلتفرم‌های پردازش زبان طبیعی، تجزیه و تحلیل داده، یا حتی ایجاد محتوا باشند. به عنوان مثال، ابزارهای پردازش تصویر یا چت‌بات‌ها.

اکانت اکانت‌ها معمولاً به کاربرانی اطلاق می‌شوند که به یک سرویس یا پلتفرم دسترسی دارند. برای استفاده از مدل‌ها یا ابزارهای هوش مصنوعی، معمولاً نیاز به ایجاد اکانت است تا کاربران بتوانند به این خدمات دسترسی پیدا کنند. به عنوان مثال، اکانت کاربری در OpenAI برای دسترسی به مدل‌های GPT.



1. شرکت OpenAI

نام مدل	توانایی‌ها	وضعیت دسترسی	وضعیت در ایران
GPT-1	تولید متن، درک زبان طبیعی	-	-
GPT-2	تولید متن پیشرفته، ترجمه، خلاصه‌سازی	رایگان	نیاز به VPN
GPT-3	تولید متن شبیه به انسان، پاسخ به سؤالات، ترجمه	رایگان و پولی	نیاز به VPN
GPT-3.5	بهبود یافته GPT-3، دقت بالاتر	رایگان و پولی	نیاز به VPN
GPT-4	توانایی‌های پیشرفته‌تر در تولید و درک زبان	پولی اشتراک ماهانه	نیاز به VPN
GPT-4o	تعامل چندوجهی متن، صدا، تصویر، تشخیص احساسات	رایگان	نیاز به VPN

<https://chat.openai.com/>



2. شرکت گوگل - جیمینای

وضعیت در ایران	وضعیت دسترسی	توانایی‌ها	نام مدل
نیاز به VPN	رایگان محدود و پولی	پردازش سریع زبان طبیعی، تولید متن، ترجمه، خلاصه‌سازی	Gemini 2.0 Flash
نیاز به VPN	رایگان محدود و پولی	استدلال پیشرفته، نمایش فرآیند تفکر مدل در پاسخ‌ها	Gemini 2.0 Flash Thinking Experimental
نیاز به VPN	رایگان محدود و پولی	بهبود درک و استدلال، مناسب برای کدنویسی و پردازش درخواست‌های پیچیده	Gemini 2.0 Pro Experimental

<https://gemini.google.com/>



Claude

3. شرکت Anthropic

وضعیت دسترسی	وضعیت در ایران	توانایی‌ها	نام مدل
نسخه رایگان و پولی	نیاز به VPN	پردازش زبان طبیعی، مکالمه، پاسخ به سؤالات	Claude

<https://claude.ai/>



deepseek

4. شرکت DeepSeek

نام مدل	توانایی‌ها	وضعیت دسترسی	وضعیت در ایران
DeepSeek R1	پردازش زبان طبیعی، استدلال منطقی، جستجوی وب	رایگان	بدون نیاز به VPN؛ دسترسی آزاد برای کاربران ایرانی

<https://chat.deepseek.com>



5. شرکت xAI

نام مدل	توانایی‌ها	وضعیت دسترسی	وضعیت در ایران
Grok-1	پردازش زبان طبیعی، تولید متن، پاسخ به سوالات، کدنویسی	نیاز به عضویت در X توییتر سابق Premium+	VPN نیاز به
Grok-1.5	بهبود عملکرد در ریاضیات و کدنویسی، افزایش طول متن قابل پردازش	نیاز به عضویت در X توییتر سابق Premium+	VPN نیاز به
Grok-3	استدلال پیشرفته، تولید متن خلاقانه، پاسخ به سوالات پیچیده، درک عمیق‌تر از زبان	نیاز به عضویت در X توییتر سابق Premium+	VPN نیاز به

<https://grok.com/>

Lmarena.ai

[Chatbot Arena formerly LMSYS Free AI Chat to Compare & Test Best AI Chatbots](https://lmarena.ai)
lmarena.ai

مقایسه تکاملی مدلها

<https://public.flourish.studio/visualisation/21639967/>

Rank★ (UB)	Rank (StyleCtrl)	Model	Arena Score	95% CI	Votes	Organization	License
1	1	chocolate (Early Grok-3)	1403	+5/-6	13065	xAI	Proprietary
2	4	Gemini-2.0-Flash-Thinking-Exp-01-21	1385	+5/-5	16889	Google	Proprietary
2	3	Gemini-2.0-Pro-Exp-02-05	1380	+5/-6	14843	Google	Proprietary
2	1	ChatGPT-4o-latest (2025-01-29)	1377	+5/-5	16555	OpenAI	Proprietary
5	3	DeepSeek-R1	1363	+6/-7	8099	DeepSeek	MIT
5	9	Gemini-2.0-Flash-001	1359	+5/-6	12616	Google	Proprietary
5	3	o1-2024-12-17	1353	+4/-4	19186	OpenAI	Proprietary
8	9	Qwen2.5-Max	1335	+5/-6	11308	Alibaba	Proprietary
8	6	o1-preview	1335	+3/-4	33173	OpenAI	Proprietary

مقایسه تکاملی مدل ها

پرامپت Prompt

پیام، پیام واره یا دستور

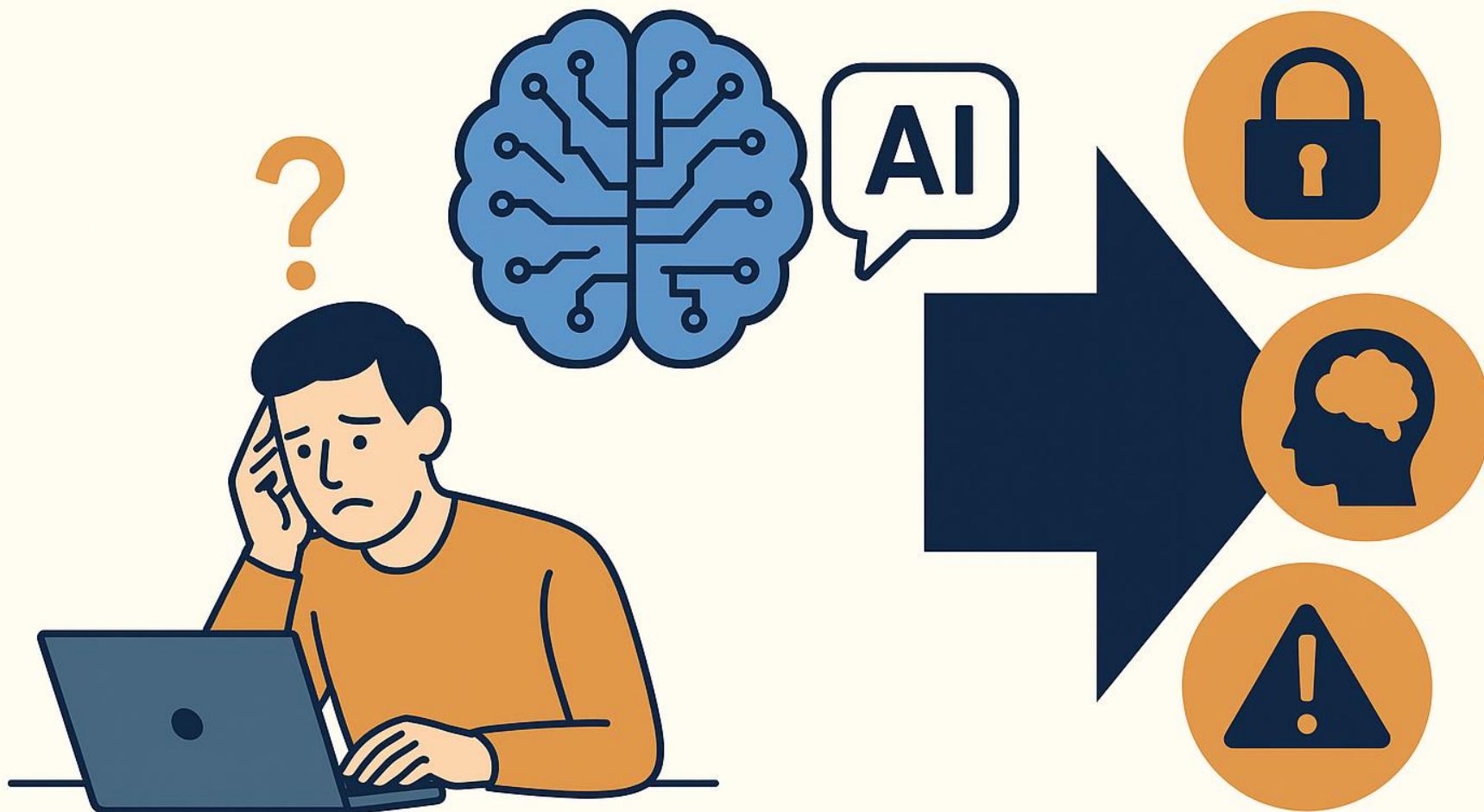
انواع پیام واره – بخش اول

نوع	تعریف	مثال
مستقیم	دستور ساده و شفاف بدون مقدمه یا چارچوب خاص.	فهرستی از کشورهای عضو ناتو بده.
زمینه‌ای	ارائه اطلاعات زمینه‌ای برای پاسخ دقیق‌تر و زمینه‌مند.	برای ارائه درسی درباره هوش مصنوعی آماده می‌شوم. چند منبع معتبر معرفی کن.
نقش‌آفرینی	قرار دادن مدل در نقش یا هویتی خاص.	فرض کن که یک روان‌شناس هستی. چه توصیه‌ای برای کاهش اضطراب امتحان داری؟
زنجیره تفکر	درخواست حل مرحله‌ای و گام‌به‌گام یک مسئله.	مرحله به مرحله توضیح بده که چطور می‌توان تابع مشتق‌پذیر را در پایتون پیاده‌سازی کرد.
مثال‌محور	استفاده از نمونه برای یادگیری الگو.	با توجه به این ایمیل نمونه، یک ایمیل مشابه برای درخواست فرصت مطالعاتی بنویس.
ضد‌الگو	مشخص کردن عباراتی که نباید استفاده شوند.	یک معرفی‌نامه برای رزومه بنویس، ولی از کلمات “شاید” و “احتمالاً” استفاده نکن.
سبک‌محور	تنظیم سبک نگارش پاسخ.	این متن را به زبان رسمی و دانشگاهی بازنویسی کن.
تعاملی	گفت‌وگوی پویا و چندمرحله‌ای با مدل.	به من کمک کن موضوع تحقیق انتخاب کنم. اول چند سوال بپرس تا علاقه‌م را بفهمی.

انواع پیام واره – بخش دوم

نوع پرامپت	تعریف	مثال
فرضی	سناریو یا شرایط خیالی.	اگر در دنیایی بدون نیروی جاذبه زندگی می‌کردیم، قوانین فیزیک چه تغییری می‌کرد؟
مقایسه‌ای	مقایسه بین چند مفهوم برای درک بهتر.	تفاوت یادگیری تقویتی و یادگیری نظارت‌شده را توضیح بده.
خلاقانه	تشویق به خلق ایده، داستان، شعر یا طرح.	یک داستان کوتاه تخیلی درباره شهری در مریخ بنویس.
تحلیلی	بررسی و تجزیه و تحلیل دقیق.	این پاراگراف را از نظر استدلال منطقی و انسجام تحلیل کن.
ساختار یافته	درخواست پاسخ در قالب خاص مثل جدول، JSON یا نمودار.	مزایا و معایب انرژی هسته‌ای را در قالب جدول دو ستونی بنویس.
هدف‌محور	تمرکز بر رسیدن به یک هدف خاص.	مخاطب من یک استاد دانشگاه است. یک متن متقاعدکننده برای درخواست فرصت پژوهشی بنویس.
چندمرحله‌ای	ترکیب چند نوع پرامپت به صورت توالی منطقی.	ابتدا مفاهیم اصلی شبکه عصبی را شرح بده، بعد آن را با درخت تصمیم مقایسه کن، و در پایان نتیجه‌گیری ارائه بده.
متا-پرامپت	طراحی یا بهینه‌سازی خود پرامپت.	چطور می‌توانم یک پرامپت بهتر برای خلاصه‌سازی مقاله بنویسم؟
بازخورد محور	ارائه پاسخ و درخواست بازخورد یا ویرایش.	این خلاصه را بررسی کن و بگو چطور می‌توانم آن را بهبود بدهم.

آسیب های استفاده افراطی



آسیب چیست؟!

آسیب: هر نوع تأثیر منفی بر افراد، جامعه یا محیط زیست

○ جسمی، روانی، اقتصادی، اجتماعی، اخلاقی

آسیب های ناشی از هوش مصنوعی

○ جسمی، فیزیکی: درگیری ربات های خودکار با انسان در شرایط ناامن

○ روانی: انتشار محتوای مضر یا گمراه کننده

○ اقتصادی: از بین رفتن مشاغل به دلیل خودکارسازی

○ اجتماعی: افزایش نابرابری یا تبعیض از طریق الگوریتم های جهت دار و اریب

○ اخلاقی: نقض حریم خصوصی یا تصمیم گیری بدون شفافیت

چرا هوش مصنوعی ذاتا آسیب زاست؟!

۱. **سوءاستفاده:** می تواند برای گسترش اطلاعات نادرست، کلاهبرداری یا جاسوسی استفاده شود.
۲. **تعصب:** اگر داده های آموزشی سوگیری داشته باشند، نتایج ناعادلانه یا تبعیض آمیز تولید می کند.
۳. **حریم خصوصی:** جمع آوری و تحلیل داده های شخصی ممکن است به نقض حریم خصوصی منجر شود.
۴. **بیکاری:** خودکارسازی مشاغل می تواند باعث از دست دادن شغل برای برخی افراد شود.
۵. **خطای انسانی:** وابستگی بیش از حد به هوش مصنوعی ممکن است خطاها را در تصمیم گیری های مهم افزایش دهد.

پیش بینی روند توسعه و تکامل هوش مصنوعی

Until 2030



توسعه علمی و فنی

تکامل هوش کامل ظرف پنج
سال آینده (حداکثر)

رقابت بین شرکتها و صنایع

2030 to 2050



فناوری به مثابه سلاح

هوش مصنوعی در دست
ابر قدرت ها

گسترش جنگ و سلطه گری

After 2050



هوش مستقل

یادگیری و تکامل مستقل از
بشر

تولید و مصرف بهینه انرژی

گروه بندی کاربران بر اساس نوع فعالیت و تخصص

○ کاربران عادی، برنامه نویسان، پژوهشگران

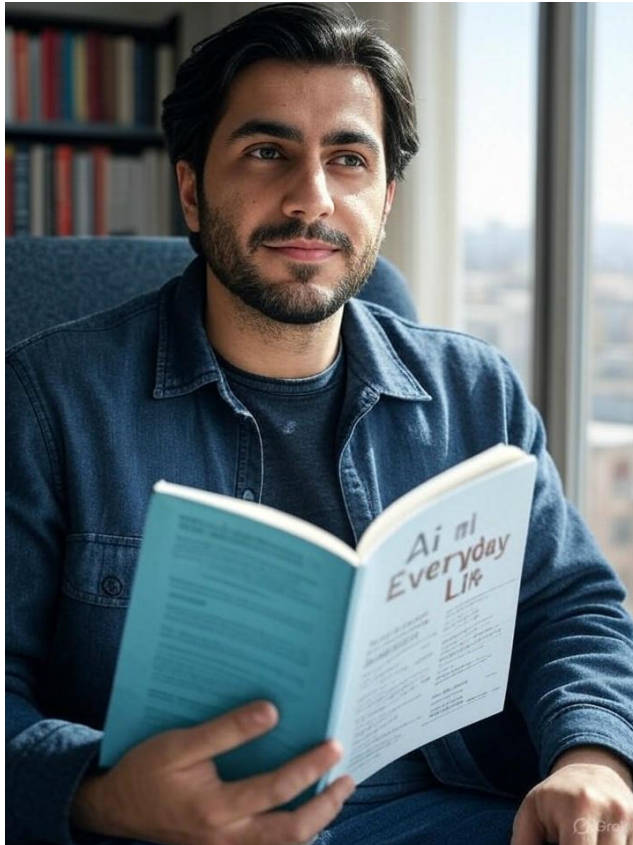
شناسایی آسیب برای هر گروه

○ نوع و منشاء آسیب

ارائه تمهیداتی برای کاستن از آسیب یا هزینه

○ مراقبت های رفتاری، تمرین های فکری و ...

کاربران عمومی



سواد دیجیتال عمومی دارند، ولی فنی نیستند.

از ابزارهای هوشی بیشتر برای کارهای روزمره، نوشتن، ترجمه، یا پرسش و پاسخ استفاده می کنند.

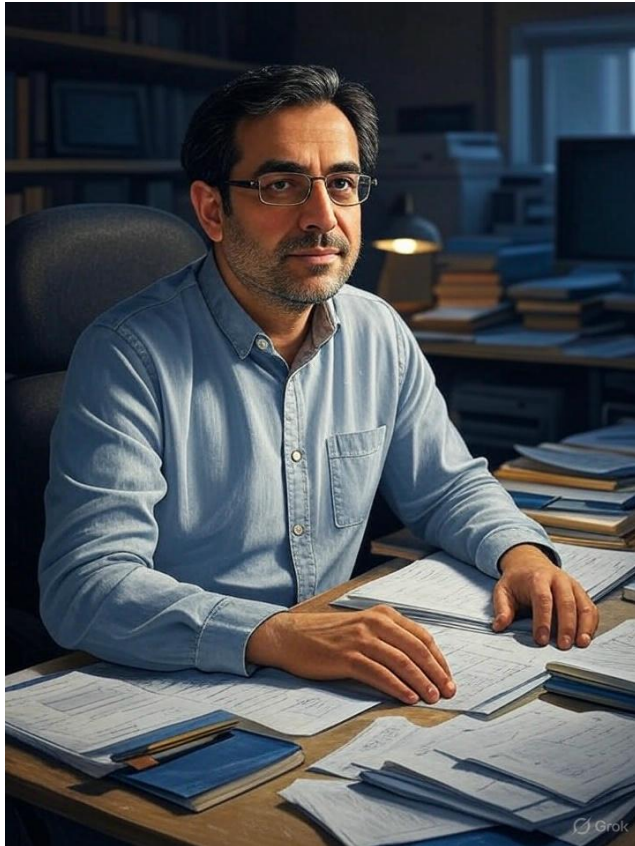
برنامه نویس، مهندس، Developer

توانایی کدنویسی دارند،
ابزار و محصولات هوشی
می سازند

به API ها، مدل های
متن باز، و کتابخانه هایی
مانند LangChain یا
HuggingFace
مسلط اند.



پژوهشگران هوش مصنوعی

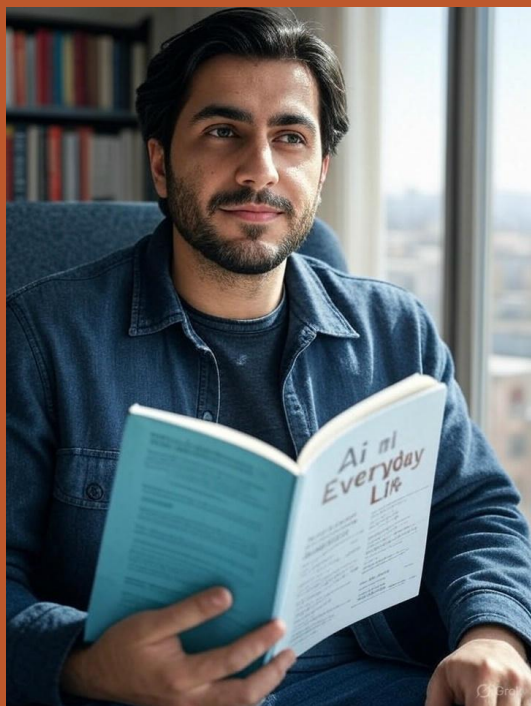


درگیر توسعه مدل‌های زبانی، ارزیابی معماری‌ها، یا طراحی الگوریتم‌های بهینه‌سازی هستند.

آشنایی عمیق با ترنسفورمرها، دیتاست‌ها، تکنیک‌های fine-tuning، evaluation و scaling دارند.

هدفشان توسعه قابلیت‌های مدل‌هاست، نه صرفاً استفاده از آن‌ها.

آسیب های متوجه کاربران عمومی



آسیب ۱: تنبلی ذهنی و فرسایش شناختی

استفاده مداوم از ابزارهای
هوش مصنوعی می‌تواند
قدرت تحلیل و تفکر
انتقادی را کاهش دهد.

مفهوم مرتبط:

Cognitive Offloading
- واگذاری تفکر به ماشین.

در بلندمدت وابستگی
شناختی ایجاد می‌کند.

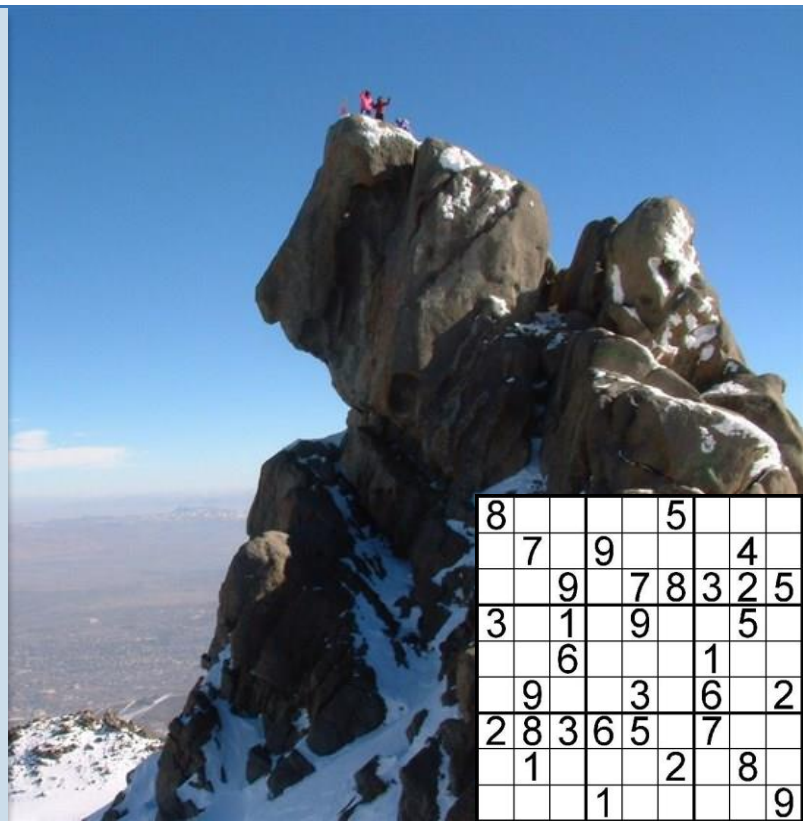


آسیب ۱: تنبلی ذهنی و فرسایش شناختی / پیشنهاد

ایده: پیروی از تحول تاریخی
کشاورزی و شهرنشینی
ورزش برای سلامتی جسم و
ذهن

پیشنهاد: بازی‌های فکری و
رایانه‌ای برای تقویت تفکر
خلاق و حل مسئله

پازل پیچیده، شبیه‌سازی
استراتژیک، چالش‌های ذهنی



آسیب ۲: نشت اطلاعات حساس

داده‌های محرمانه در
پرامپت
خطر ذخیره‌سازی یا
آموزش مجدد مدل با
داده‌ها
تهدیدی برای حریم
خصوصی شخصی و
حرفه‌ای



آسیب ۲: نشت اطلاعات حساس / پیشنهاد

استفاده از توصیف به جای داده
◦ فرض کن یک فایل اکسل با دو ستون دارم ...

گمنام سازی

- حذف ستون های دارای اطلاعات شخصی مثل نام و نام خانوادگی، شماره حساب، کد ملی
- جایگزینی این ستون ها با داده های دیگر
- کاربر ۱، کاربر ۲

استفاده از داده های جعلی

- ستون حاوی حقوق ماهیانه را با اعداد غیر واقعی پر کنید.

◦ تقریبا مشابه Differential Privacy

- استفاده از مدل های کوچک، محلی و فریز شده
- مدل نباید چیزی را به خاطر بسپارد.





آسیب ۳: هلوسینیشن مدل‌ها

مدل‌ها گاهی اطلاعات نادرست ولی با اعتماد به نفس تولید می‌کنند.

ارجاعات جعلی، استدلال نادرست، داده‌های غیرواقعی.

خطر تصمیم‌گیری بر پایه اطلاعات نادرست.

آسیب ۳: هلو سینیشن مدل‌ها/ پیشنهاد



شک و تردید، شرط لازم استفاده از ابزارهای هوش مصنوعی

استفاده از پرامپت مرتبطتر، دقیقتر و مفصل تر

- به تجربه دیده شده که برای مسائل خاص انواع معینی از پرامپت ها جواب مناسب تولید می کنند.

- پرامپت ساده: چه عواملی باعث افزایش کارایی در یادگیری ماشینی می شوند؟

- پرامپت مناسب: در زمینه یادگیری ماشین، لطفاً سه عامل اصلی را که باعث بهبود دقت مدل‌ها می‌شوند، با توضیح دقیق و مثال‌های واقعی بیان کن

- برای حل مسائل ریاضی و الگوریتم از مدل few-shot استفاده کنید و مثال مناسب ذکر کنید

- مثال: آیا عدد ۹۷ اول است؟ می دانیم که این عدد وقتی اول است که بر هیچ عددی از ۲ تا ۹۶ بخشپذیر نباشد.

برای مسائلی که شامل حقایق علمی هستند یافته ها را با منابع مختلف چک کنید

- منابع تالیفی انسانی مناسب تر هستند

- از کارشناسان می توان کمک گرفت

- یافته های فارسی و حقایقی را که محدود به محل جغرافیایی خاص، زبان، فرهنگ یا دین خاص هستند با شک و تردید بیشتری بررسی کنید

- بیشتر مدلها دانش کمتری به زبان فارسی دارند.

برای پرهیز از ارجاعات جعلی، از ابزارهای خاص پژوهش استفاده کنید.

- مثل getliner.com برای مطالعه مقالات علمی

آسیب ۴: وابستگی به زیرساخت‌های خارجی

بسیاری از ابزارهای
هوش مصنوعی در اختیار
شرکت‌های فراملی
هستند.

خطر تحریم، قطع
دسترسی، یا سوءاستفاده
سیاسی.

کاهش استقلال و
حاکمیت دیجیتال ملی.



آسیب ۴: وابستگی به زیرساخت‌های خارجی/پیشنهاد

چت جی پی تی و گراک از تحریم خارج شده اند!

تا از رقابت در جمع آوری داده ها جا نمانند؟!
از ابزارها و محصولات داخلی زیاد استفاده کنید تا بالغ شوند.

اگر مدیر و مسئول هستید از محصولات ایرانی حمایت و به توسعه آنها کمک کنید.

از مدیران و روسای سازمان بخواهید
ابزارهای محلی و ویژه خود شما تامین کنند



آسیب ۵: سوگیری الگوریتمی و بی عدالتی/پیشنهاد



دانش عمومی، تاریخی و اجتماعی خود را تقویت کنید:

◦ Black lives matter, Palestine, Persian Gulf, ...

◦ سال مرغ (ایرانی) و سال خروس (چینی)

حساسیست های محلی، قومی، زبانی و ... را بشناسید

◦ استقلال، پرسپولیس، تراکتورسازی،

◦ کنایه های توهین و تحقیر آمیز محلی و قومی را بشناسید تا بتوانید آنها را در محتوای تولیدی هوش مصنوعی بشناسید و از باز نشر وبازگویی آنها پرهیز کنید.

مدلها از ما یاد می گیرند:

◦ اگر محتوا تولید می کنید، سعی کنید در دام این تبعیض ها نیفتید

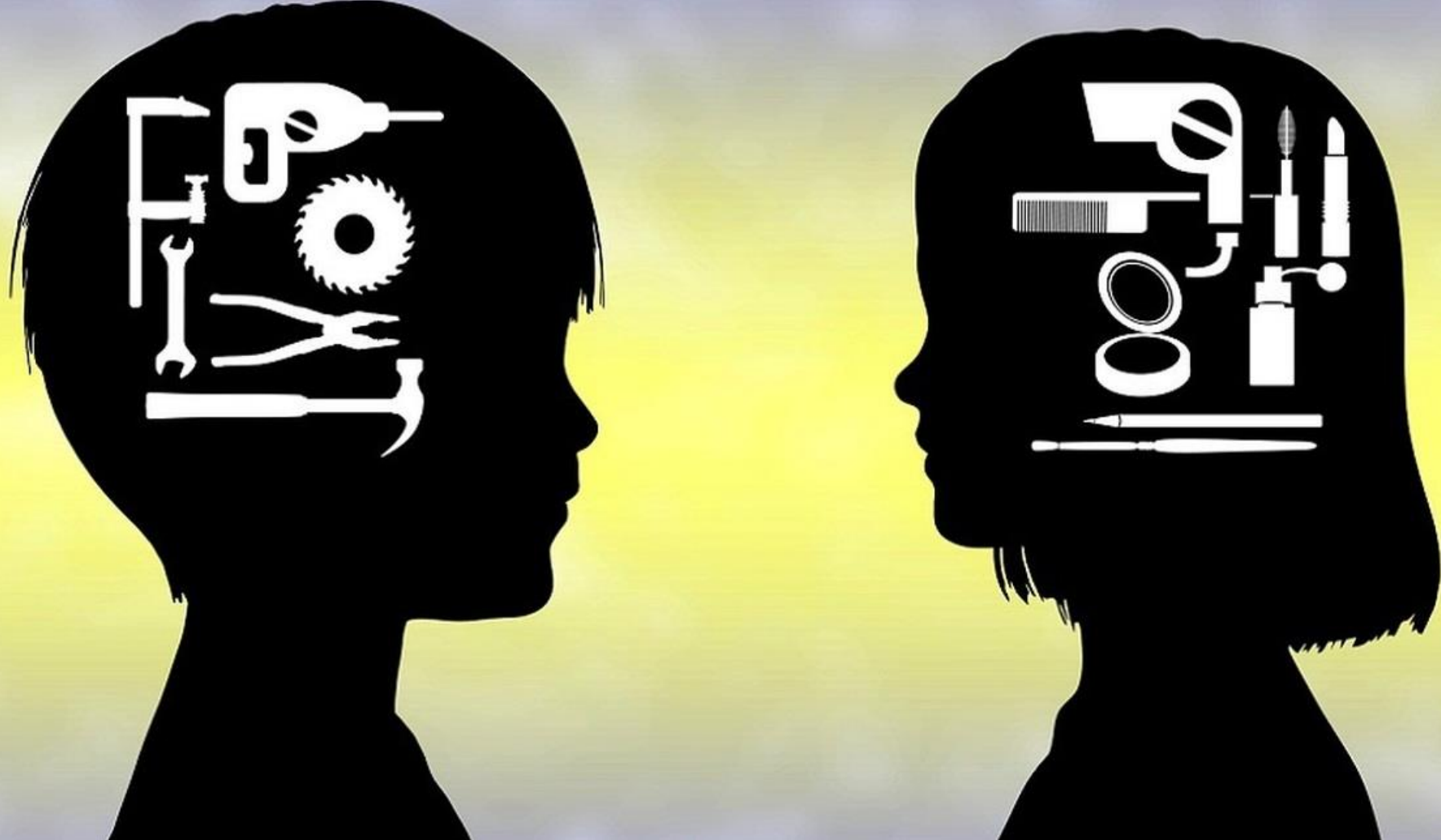
◦ به فرهنگ و زبان رسمی و معیار پایبند باشیم

◦ نقاط قوت همدیگر را برجسته کنیم

اگر مدیر هستید و استخدام می کنید:

◦ از لهجه فقط به عنوان نقطه قوت یاد کنید

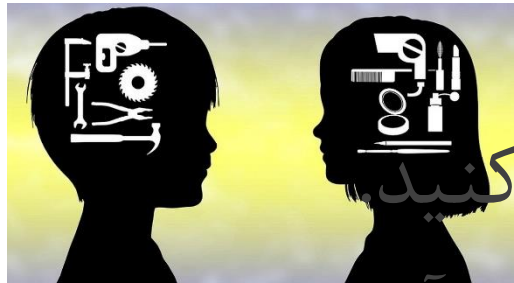
◦ توانایی های فردی را در قالب استانداردهای حرفه ای بسنجید



آسیب ۶: بازتاب تقویت شده باورهای کلیشه ای

مدل های زبانی روایت های غالب را بازتولید می کنند.
تکرار کلیشه ها باعث تقویت فرهنگ سطحی می شود.
خطر جایگزینی حقیقت با بازتاب جمعی.

آسیب ۶: بازتاب تقویت‌شده باورهای کلیشه‌ای / پیشنهاد



محتوای تولیدی هوشی را
همواره با شک و تردید ارزیابی کنید
کلیشه‌ها را بشناسید و از بازنشر آنها پرهیز
نمایید

آنچه که تکرار می‌شود، لزوماً صحیح نیست.
پُر نیست، لایف استایله!

آسیب های ویژه برنامه نویس و مهندسان نرم افزار

آسیب های ویژه برای برنامه نویسان: آسیب های عمومی

آسیب های ۱-۶ برای برنامه نویسان هم قابل
تصور هستند

برنامه نویسان و سازندگان مدل های هوشی باید
تمهیدات ویژه ای در طراحی این مدل ها داشته
باشند تا کاربران عادی دچار چالش های ۱-۶
نشوند!

آسیب های ویژه برای برنامه نویسان: آسیب های عمومی / پیشنهاد

- تولید ابزار و ایجنت با کارایی مناسب
- استفاده از ابزارهای یادگیری ماشین برای تشخیص آسیب ها
- استفاده از تکنیک های Document Classification, Sentiment Analysis, Text Clarity Score, برای تحلیل و دسته بندی محتوای متنی تولید شده توسط هوش مصنوعی
- مقایسه حقایق بیان شده در متن تولیدی هوش با حقایق بیان شده در منابع طلایی با استفاده از روشهای بازیابی اطلاعات
- فیلتر محتوای نامناسب به کمک روش های یادگیری ماشین
- نمونه های موفق: Twitter fact checking, DALLE privacy preserving , ...
- ارائه راهکار و راهنمایی مناسب به کاربران

آسیب های ویژه برای برنامه نویسان: آسیب های عمومی / پیشنهاد

تنظیمات کلیدی در مدل های هوش مصنوعی

◆ **Temperature** دمای مدل مقدار کم، پاسخ های دقیق و قابل پیش بینی تولید می کند، در حالی که مقدار بالا باعث تنوع و خلاقیت بیشتر می شود. برای پرسش های علمی مقدار کم و برای تولید متن های خلاقانه مقدار بالا توصیه می شود.

◆ **Top P** روشی برای کنترل تنوع پاسخ ها از طریق نمونه گیری. مقدار کم برای پاسخ های دقیق و مقدار بالا برای تولید متون متنوع استفاده می شود. معمولاً یا Top P یا Temperature تنظیم می شود، نه هر دو.

◆ **Max Length** حداکثر طول محدودیت تعداد توکن ها برای جلوگیری از پاسخ های طولانی یا نامربوط و کنترل هزینه ها.

◆ **Stop Sequences** توالی های توقف تعیین رشته ای که مدل پس از آن متوقف شود. به عنوان مثال، برای محدود کردن لیست به ۱۰ مورد، می توان "۱۱" را به عنوان توالی توقف تعیین کرد.

◆ **Frequency Penalty** جریمه فرکانس کاهش احتمال تکرار بیش از حد یک کلمه در پاسخ. مقدار بالاتر باعث کاهش تکرار و مقدار کمتر باعث افزایش احتمال استفاده مجدد از کلمات می شود.

◆ **RAG** از این روش برای جستجوی بهتر و متمرکز استفاده کنید.



آسیب ۷: نقض حاکمیت داده و مالکیت اطلاعات

استفاده از API ها داده ها را به بیرون از کشور منتقل می کند.

تهدیدی برای امنیت ملی و رقابت اقتصادی.

آسیب ۷: نقض حاکمیت داده و مالکیت اطلاعات/پیشنهاد

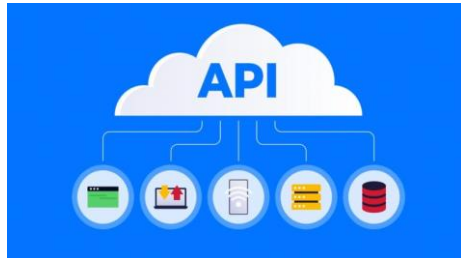
از قاعده حداقل دسترسی استفاده کنید

- 
- Order Deny, Allow,
 - Deny from ALL
 - Allow from host1, host2

تا جای ممکن از API ها استفاده نکنید

- مخصوصا دسترسی با API به ابزار و ایجنت خود را محدود کنید.
- مانیتورینگ دائمی داشته باشید و مبادلات داده ای ابزار خود با دنیای بیرون را زیر نظر بگیرید.
- با روشهای یادگیری ماشینی می توانید این مبادلات و فعالیت ها را به صورت هوشمند کنترل کنید و هشدارهای لازم را به صورت فوری صادر نمایید

آسیب ۷: نقض حاکمیت داده و مالکیت اطلاعات/پیشنهاد



استفاده از LLM تنها راه حل نیست!
○ راه حل خوبی است چون به زبان آدمیزاد با آن تعامل می کنیم
○ چالش های فنی و امنیتی زیادی دارد.

استفاده از ابزارها و تکنیک های اختصاصی جایگزین
○ بسیاری از مسائل سازمانی را می توان با ابزارها و تکنیک های کارآمد، کم هزینه و امن جایگزین کرد
○ نمونه اول: سیستم پیشنهاد دهنده برای دوره آموزشی
○ نمونه دوم: چت بات راسا

سیستم پیشنهاد دوره

جنسیت

مرد

سن

30

استان

همدان

تحصیلات

بیسواد

عنوان شغلی

بیکار قبلا شاغل(بیمه بیکاری)

در صورتی که مبتلا به مشکل پزشکی یا روحی هستید، قسمت مربوطه را علامت بزنید.

☐ عارضه قلبی

☐ مشکلات سلامتی دیگر

☐ مشکلات روحی

در صورتی که مشمول هر یک از موارد زیر هستید آن را انتخاب کنید:

☐ آزاد ☐ جانباز ☐ رزمنده ☐ بسیجی ☐ بسیجی فعال

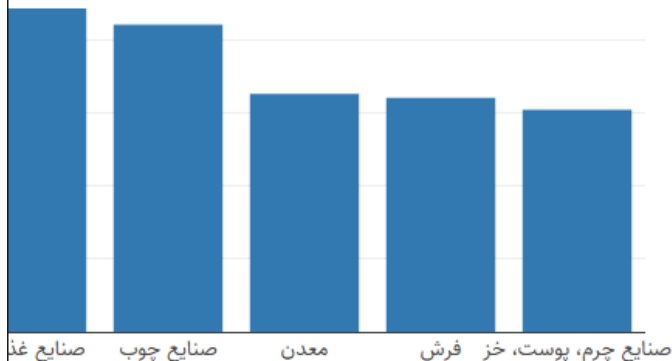
☐ خانواده شهید

پیشنهاد دوره

نتایج تحلیل

دوره های پیشنهادی برای شما به شرح زیر است:

دوره های پیشنهادی به ترتیب اولویت



صنایع غذایی

توصیه می شود موارد زیر را در مورد انتخاب این گروه آموزشی در نظر بگیرید:

- سطح تحصیلات
- سکونت در استان همدان
- جنسیت

استاندارد پیشنهادی در این گروه کارگروپیشکارانوانایی است.

راسا RASA، ابزاری برای چت بات سازمانی

Rasa یک فریم‌ورک متن‌باز برای ساخت چت‌بات‌های هوشمند و گفت‌وگو محور است که قابلیت اجرای محلی، سفارشی‌سازی کامل، و کنترل دقیق بر مکالمه را فراهم می‌کند.

مناسب برای سازمان‌هایی با داده حساس، نیاز به سفارشی‌سازی و عدم وابستگی به API‌های خارجی. شامل دو جزء اصلی:

- Rasa NLU- Language Understanding
- Rasa Core – Dialogue Management

مزایا و محدودیت‌های Rasa

مزایا: 

- کاملاً متن‌باز و رایگان
- قابل استقرار در محیط امن داخلی
- کنترل کامل بر intents، پاسخ‌ها، و منطق گفتگو
- ادغام آسان با API‌ها، پایگاه‌های داده و LLM‌ها

محدودیت‌ها: 

- نیاز به داده‌های آموزشی برجسته‌شده
- مدیریت مکالمه‌های پیچیده نیازمند طراحی دقیق است
- پاسخ‌ها معمولاً غیرمولد و تکراری مگر با اتصال به مدل‌های بزرگ زبانی

آسیب های ویژه پژوهشگران و توسعه دهندگان مدل‌های هوشمند

آسیب های ویژه پژوهشگران و توسعه دهندگان مدل های هوشمند

به عنوان کاربر، همه آسیب های ۱-۶

به عنوان به برنامه نویس، آسیب ۷

به عنوان دانشمند: هلو سینی شن علم، ریسک باز تولید
نتایج، محدودیت منابع، تب یادگیری عمیق

آسیب های خاص پژوهشگران

چالش

عدم شفافیت مدل های بسته

محدودیت در دسترسی به مدل های مقیاس پذیر

عدم تکرارپذیری Reproducibility

سوگیری های آموزش یافته نامشخص

پیچیدگی در ارزیابی مدل ها Benchmarking

توضیح

مثل GPT-4 یا Claude، معماری و دیناست آموزشی شان مشخص نیست، مانعی جدی برای پژوهش باز.

منابع GPU سنگین، نیاز به داده های بزرگ، و دسترسی به مدل های با لایسنس محدود.

مدل های تجاری اغلب اجازه تست مستقل با شرایط کنترل شده نمی دهند.

نمی دانند که کدام منابع دقیقاً مدل را هدایت کرده اند.

تست در task های مختلف، گاهی منجر به نتایج گمراه کننده می شود.

در نتیجه این چالش ها بخصوص کمبود منابع و دادگان، عمده پژوهش های توسعه مدل های زبانی دچار سوگیری داده ای زیاد، بیش برآزش، عملکرد نامطمئن و نادقیق در شرایط متنوع و در نهایت عدم اقبال عمومی می شوند.

آسیب های خاص پژوهشگران/پیشنهاده

چالش

عدم شفافیت مدل های بسته

محدودیت در دسترسی به مدل های مقیاس پذیر

عدم تکرارپذیری Reproducibility

سوگیری های آموزش یافته نامشخص

پیچیدگی در ارزیابی مدل ها Benchmarking

پیشنهاد

از این مدلها در مقایسه و ارزیابی پژوهش استفاده نکنید. مدل های متن باز و در دسترس مثل لاما گزینه مناسبتری هستند.

سعی کنید پژوهش خود را به جنبه خاصی از مدل محدود کنید تا منابع کمتری نیاز داشته باشد. با سازمانها و گروه های برخوردار همکار داشته باشید.

سعی کنید از روشهایی که به توضیح پذیری کمک می کنند استفاده کنید.

از روشهای یادگیری تقویتی و مشارکت انسانی استفاده کنید.

پژوهش در زمینه یادگیری عمیق به صورت عام و مدل زبانی به صورت خاص نیازمند منابع محاسباتی زیاد است. این چالش صرفا با مشارکت و هم افزایی قابل حل است. بنابراین ضروری است برای ایجاد زیرساخت های قوی و مشترک تلاش کنیم.

ویژگی \ گروه	کاربران معمولی	برنامه‌نویسان	پژوهشگران
سطح فنی	پایین	متوسط تا بالا	بالا
استفاده	روزمره و عمومی	ادغام در پروژه‌ها	تحلیل و تولید علم
مخاطرات کلیدی	پذیرش کورکورانه، نشت داده	API ضعف امنیتی، اعتماد به	هلوسینیشن علمی، عدم بازتولید
راهکار کلیدی	آموزش، سادگی، هشدار	local فیلتر امنیتی، مدل	مدل‌های مستند، راستی‌آزمایی

جمع بندی

پژوهشگران	برنامه نویسان	کاربران معمولی	
بالا	متوسط تا بالا	پایین	دانش فنی
تحلیل و تولید علم	ادغام در پروژه ها	روزمره و عمومی	استفاده
هلو سینی شین علمی، عدم باز تولید	ضعف امنیتی، اعتماد به API	پذیرش کورکورانه، نشت داده	آسیب
مدل های مستند، راستی آزمایی	فیلتر امنیتی، مدل local	آموزش، سادگی، هشدار	راهکار

منابع و مراجع



بخش قابل توجهی از مطالب و اسلایدها برگرفته از ارائه دوست عزیزم دکتر عادل فاطمی است

سایر منابع و مراجع علمی :

۱. راسل، استوارت. و نوروینگ، پیت. (۲۰۲۱). هوش مصنوعی: رویکردی نوین (ویرایش چهارم). انتشارات پیرسون.
۲. میچل، تام. (۱۹۹۷). یادگیری ماشین. نیویورک: مک‌گرو-هیل.
۳. گودفلو، ایان، بنجیو، یوشوا، و کورویل، آرون. (۲۰۱۶). یادگیری عمیق. انتشارات MIT.
۴. واسوانی، آشوین، شازیر، نوآ، یارمار، نوکی، و دیگران. (۲۰۱۷). توجه همه چیز است: معرفی معماری ترنسفورمر. کنفرانس Advances in Neural Information Processing Systems. <https://arxiv.org/abs/1706.03762>
۵. براون، تام، مان، بن، رایدر، نلیا، و دیگران. (۲۰۲۰). مدل‌های زبانی یادگیرنده با نمونه‌های کم (arXiv:2005.14165). <https://arxiv.org/abs/2005.14165>
۶. OpenAI. (۲۰۲۳). گزارش فنی GPT-4. <https://openai.com/research/gpt-4>
۷. لیو، پینگ، یوان، ونی، فو، جینگ، و دیگران. (۲۰۲۳). پیش‌آموزش، پرامپت‌دهی و پیش‌بینی: مرور سیستماتیک روش‌های پرامپت‌نویسی در پردازش زبان طبیعی. arXiv:2107.13586. <https://arxiv.org/abs/2107.13586>
۸. OpenAI. (بدون تاریخ). دفترچه راهنمای OpenAI Cookbook. <https://github.com/openai/openai-cookbook>
۹. دیپ‌لرنینگ‌ای‌آی. (بدون تاریخ). مهندسی پرامپت ChatGPT برای توسعه‌دهندگان. <https://learn.deeplearning.ai/chatgpt-prompt-eng/>
۱۰. هاجینگ‌فیس. (بدون تاریخ). مستندات Transformers. <https://huggingface.co/docs/transformers>
۱۱. LMSYS. (بدون تاریخ). چت‌بات آرنا: مقایسه و آزمون مدل‌های زبانی بزرگ. <https://chat.lmsys.org/>

منابع این دوره،
اسلایدها و

AI Workshop 4032



AI Workshop 4032

Address: <https://b2n.ir/basuai>

Password(all small): basuaiworkshop