

# درباره Chat GPT | فرصت‌ها، تهدیدها و تاثیرات چت جی پی تی بر دنیای محتوا (فلسفه تکنولوژی دیجیتال - ۲ مرداد، ۱۴۰۲)

آن چه در این نوشته می‌خوانید (فعلاً):

[پیش‌نوشت‌ها](#)

[زبان به عنوان زنجیره‌ای از کلمات](#)

[مدل‌سازی زبان به عنوان زنجیره‌ای از کلمات چه کاربردهایی دارد؟](#)

- [پیش‌بینی کلمات بعدی هنگام تایپ](#)
- [حدس زدن پرسش‌های کاربران بر اساس نخستین کلمات‌شان](#)
- [افزایش دقت OCR](#)
- [کمک به نرم‌افزارهای تبدیل صدا به متن](#)

[دو رویکرد متفاوت به مدل‌سازی زبان](#)

- [رویکرد بر پایه احتمال \(ایده زنجیره مارکوف\)](#)
- [رویکرد مبتنی شبکه‌های عصبی \(مدل‌سازی حافظه‌دار\)](#)

[یک مثال از شبکه‌های عصبی \(بسیار ساده‌شده\)](#)

پیش‌نوشت یک: مدت زیادی است که به علل مختلف، از نوشتن درباره حوزه تکنولوژی فاصله گرفته‌ام. یکی از علت‌ها این است که معمولاً رویکرد من به تکنولوژی، انتقادی است یا لاقلاً به جریان اصلی نزدیک نیست. انتقادی نه به این معنا که مخالف تکنولوژی هستم که اتفاقاً عاشق تکنولوژی‌ام. بلکه از آن رو که فکر می‌کنم نگاه انتقادی می‌تواند به ما کمک کند تا از ظرفیت‌های تکنولوژی، با پرداخت کم‌ترین هزینه (پیدا و پنهان) بیشترین بهره ببریم.

اصرار بر نگاه انتقادی به تکنولوژی در ایران امروز ما، می‌تواند مخرب باشد. چون اظهارنظر در خلاء انجام نمی‌شود و باید بستر و زمینه را هم دید. وقتی زیرساخت‌های اصلی ارتباط با جهان به درستی در اختیار ما نیست، فرض کنید کسی بیاید از اثرات منفی شبکه‌های اجتماعی بگوید. این شکل از حرف زدن، بیشتر

از آن‌که به کار مردم بیاید، خوراک کسانی می‌شود که معتقدند حق اتصال به دنیای آزاد را فقط باید با سیمکارت‌های ویژه به «خارجی‌ها» داد.

در چنین شرایطی، این که کارشناسی چشم خود را ببندد و بگوید «من فقط می‌خواهم تحلیل‌های علمی و کارشناسی‌ام را بگویم» نه مصداق حماقت، که از جنس خیانت است (و متأسفانه در این مدت هر چند روز یک بار مجبور بوده‌ام این نکته واضح را به برخی دوستان و عزیزان یادآوری کنم). امروز باید از مزایای اینترنت و دسترسی آزاد به اطلاعات گفت و روزی که این امکانات برای مردم فراهم شد، آن روز درباره دروسها و سختی‌هایش گفتگو کنیم و بکوشیم اثرات منفی استفاده از آن‌ها را مدیریت کنیم.

پی‌نوشت دو: به همین علت هم در این چند سال، به رغم میل درونی درباره رمزارزها هم چیزی ننوشتم. در حالی که می‌توانید تصور کنید برای کسی که مردم جهان را نه بر پایه کفر و ایمان که بر پایه نگاه توزیع‌شده و نگاه متمرکز در شکل‌گیری پدیده‌های عالم در یک طیف می‌گنجاند، چقدر سخت است که درباره یکی از بزرگ‌ترین دستاوردهای نگاه توزیع‌شده به پول چیزی ننویسد و حرف نزند. نمی‌گویم حرف کارشناسی عجیب‌وغریبی داشته‌ام. اما حرفم این است که چه دانشی در دنیا بهتر از سیستم‌های پیچیده برای بررسی و تحلیل و ارزیابی نقاط قوت و ضعف و ظرفیت‌ها و خطرات رمزارزها وجود دارد؟ و وقتی افراد بسیاری به اشتباه، اقتصاد کلان را بهترین بستر تحلیل این محصول جدید می‌بینند، مقاومت در برابر وسوسه نوشتن سخت بوده است. اما می‌دانسته‌ام که در این مواقع، از ده جمله حرف، نه جمله رها می‌شود و آن تک‌جمله‌هایی که از رگولاتوری و تمرکز می‌گویند، پررنگ می‌شود و شلاق تنظیم‌گران غیرمتخصص بر پیکر این صنعت تازه نواخته خواهد شد. باز هم منظورم این نیست که کسی حرف امثال من را جدی می‌گیرد یا اساساً کسی مثل من حرف خاصی برای گفتن دارد، اما ماهیت این کار - توجیه ساختن برای کسانی که تشنه توجیه رفتارهای توجیه‌ناپذیر خود هستند - چیزی نیست که با روحیات من جور درآید.

پیش‌نوشت سه - یک: با وجود توضیحاتی که گفتم، تصمیم دارم اندکی درباره چت جی پی تی بنویسم. علت نخست این است که چند بار در گفتگوهای دوستانه در این باره حرف شده و حس کردم شاید به جای تکرار یک مجموعه حرف ثابت در چند مکالمه مختلف، اگر آن‌ها را کمی بهتر و هدف‌مندتر در یک ساختار منظم‌تر بگنجانم، چیز بهتری از آب در می‌آید.

پیش‌نوشت سه - دو: علت دوم این که پردازش زبان طبیعی برای من همیشه موضوعی جذاب بوده و از سال‌ها پیش، گاه‌وبی‌گاه برایش وقت می‌گذاشته‌ام (البته بیشتر کارهایی از جنس رویکرد DCG یا Definite Clause Grammar؛ شبیه چیزی که در واتسون آی‌بی‌ام توسعه پیدا کرد و با مدل‌های زبانی به معنای رایج در یادگیری ماشینی فرق دارند). طبیعتاً به خاطر این علاقه شخصی نگاهی هم به پژوهش‌های

این حوزه و رویکردهای دیگر - که سهم انسانی در آن‌ها کمتر و سهم Computational بیشتر بوده - می‌انداخته‌ام. البته حاصل این گشت‌وگذارها، من را چندان از سطح مخاطب عام این ابزارها فراتر نبرده و حرف‌ها و نظراتم، کارشناسی محسوب نمی‌شوند.

پیش‌نوشت سه - سه: بعضی اظهارنظرهایی که در این مدت درباره Chat GPT شد، آن قدر عجیب بود که نتوانستم در مقابل نوشتن مقاومت کنم. به سه موردشان اشاره می‌کنم.

یکی آن آقای که در حضور آقای رئیسی گفت که «با چت جی پی تی مکاتبه کرده» و درباره شب قدر پرسیده و گفته «بالاخره شب قدر چیز خوبی یا نه؟ باور داشته باشیم یا نه؟» و البته چت جی پی تی هم - به روایت ایشان - پاسخ داده که «یه اعتقاد شخصیه و ...» (+). یادم هست در دوران مدرسه راهنمایی که هنوز کامپیوترها به شکل عمومی برای کاربردهای روزمره رایج نبودند و ما با غرور از این که کمودور ۶۴ حدوداً ۶۴۰۰۰ «تا» حافظه دارد حرف می‌زدیم، یکی از بستگان - که آدم معتقدی بود و همه چیز را بر آن معیار می‌سنجید - پرسید: از خدا در موردش پرسیده‌ای؟ چه می‌گوید؟ (جهان به پیش می‌رود. اما عده‌ای مدام خود را تکرار می‌کنند).

دوم این که دیدم یک استاد ارتباطات دانشگاه لایه‌لای صحبت‌هایش درباره تهدید شغل خبرنگاری توسط این نوع سرویس‌ها گفته که (+) «آنچه هم که شما در هوش مصنوعی می‌بینید در زبان انگلیسی است و به زبان فارسی هم به زودی اتفاق نخواهد افتاد. اتفاق عجیبی هم نیست و دیتا خاصی در آن پیدا نمی‌شود.» البته ایشان به نکات دیگری هم اشاره کردند که بعضاً درست هستند. اما به گمانم در بررسی فرصت و تهدید یک تکنولوژی جدید، این شکل از اطمینان دادن که بالاخره زبان ما آن‌قدرها هم در جهان زنده نیست و مطالب کمی در آن هست و چون زبان‌مان در دنیا رواج ندارد، هنوز تهدید جدی برای ما نشده، روش چندان مناسبی نیست. در ارزیابی‌ها باید ظرفیت‌ها را سنجید و نه این که از ضعف‌هایمان به عنوان نقاط امیدبخش نام ببریم.

سومین مورد هم استوری یکی از دوستان خوبم بود که اتفاقاً در حوزه آی تی هم فعالند درباره Chat GPT چیزی با این مضمون (دقیق در ذهنم نیست) گفته بودند که بالاخره از چیزی که همزمان آشپزی را به اندازه آشپز می‌داند و فلسفه را به اندازه فیلسوف و تاریخ را به اندازه مورخ و پزشکی را به اندازه پزشک و ... باید ترسید (به معنای مثبت. یعنی می‌گفتند این غول چراغ جادو تقریباً هر غلطی می‌تواند بکند).

این نوع نگاه‌ها هم نشان می‌دهد که ما ظرفیت‌های تکنولوژی‌های مختلف و متعددی را که در سیستم‌های Generative هوش مصنوعی به کار می‌روند، به درستی نمی‌شناسیم یا نقاط تشابه و تمایز آن‌ها را نمی‌دانیم. مثلاً از تفاوت Inference Engine و Expert System و Language

Model غافل می‌شویم و همین باعث می‌شود در برآورد فرصت‌ها و تهدیدها و ضعف‌ها و ظرفیت‌ها خطا کنیم.

پیش‌نوشت سه - چهار: من خودم را یکی از علاقه‌مندان حوزه محتوا می‌دانم (البته در کار تولید محتوا تقریباً هیچ سابقه‌ای ندارم و در این فضا خودم را بیشتر به نویسندگی می‌شناسم). با توجه به این که یکی از کارکردهای اصلی سیستم‌هایی Pre-trained شبیه چت جی پی تی تولید محتواست، علاقه چند ساله باعث می‌شود ترغیب شوم درباره‌اش بنویسم.

پیش‌نوشت چهارم: قاعدتاً این مطلب به زودی جمع نمی‌شود و به پایان نمی‌رسد. من هم که خداوندگار نوشتن نیمه‌کاره هستم. اولویت فعلی‌ام هم بحث [فرهنگ](#) است. با این حال امید دارم به اندازه‌ای که دست و وقتم اجازه داد، این بحث را پیش ببرم و به جنبه‌های مختلفش بپردازم. این مطلب از جنس آموزش چت جی پی تی نخواهد بود. بلکه صرفاً در همین حد است که بینیم چت جی پی تی چیست و استفاده از چت جی پی تی چه تأثیری بر آینده کسب و کارها و به طور خاص، صنعت محتوا دارد. اگر فرصت کافی دست دهد، بحثم را با خطرات چت جی پی تی به پایان می‌برم و این که دسترسی به چت جی پی تی چه تأثیری بر زندگی روزمره ما و نیز فضای رقابتی میان کسب و کارها خواهد داشت.

چون کار چت جی پی تی از جنس حرف زدن است، طبیعی است که باید بحث را با «زبان» آغاز کنم. به این شکل، هم بهتر درک می‌کنیم که چی جی پی تی چگونه کار می‌کند و هم در ادامه بحث متوجه می‌شویم که چه سوالاتی از چت جی پی تی بپرسیم تا پاسخ‌های بهتر و مفیدتری بگیریم.

## زبان به عنوان زنجیره‌ای از کلمات

بیا یک بازی بسیار ساده انجام دهیم. فرض کنید با قبیله جدیدی آشنا شده‌اید و زبان‌شان را نمی‌شناسید. آن‌ها از الفبای فارسی استفاده می‌کنند و نوشتن هم بلدند. آن‌چه در ادامه می‌بینید، بخشی از گفتگوهای آن‌هاست که در یک کتیبه ثبت شده.

لچفعغ مملا؟

فقی. مالابغ سیق بلیل.

لفعغ منل محح؟

فقی. قغیغ حخت منل.

قفب مالابغا قیصی قغیغ.

لفعغ غالع داغق عهب مملا؟

فقی. ضصسق عهچغ سیص عهب منل مح؟ صق.

قاعدتاً چیزی از این گفتگو متوجه نمی‌شوید. حالا به یک مسئله فکر کنید:

یک نفر از این قبیله چنین سوالی از شما می‌پرسد: «قغیغ حخت داغق مملا؟» به نظر شما پاسخ‌تان باید چه باشد؟ هیچ‌وقت نمی‌شود با اطمینان کامل گفت. اما بر اساس داده‌های موجود اگر جهان دانش‌تان به همین تک‌مکالمه روی کتیبه محدود باشد، پاسخ خواهید داد: «فقی».

چرا؟ چون در دو نمونه‌ای که سوالی با «مملا» به پایان رسیده، با «فقی» پاسخ داده شده است. حالا فرض کنید سوال دیگری از شما می‌پرسند: «مالابغا قیصی قغیغ غالق مح؟»

در این جا بین دو گزینه گیر می‌کنید (باز هم بر اساس داده‌های موجود). جواب احتمالاً باید یا «صق» باشد یا «فقی»

چون در متن دو سوال با مح به پایان رسیده و یک بار صق جواب گرفته و یک بار فقی.

البته ممکن است کمی پیچیده‌تر هم فکر کنید که فعلاً به سراغش نمی‌رویم. ممکن است بگویید جواب «پرسش ۵ کلمه‌ای که با «مح» تمام شده، صق بوده. بنابراین با توجه به این که پرسش جدید هم پنج کلمه‌ای است، من بین «فقی» و «صق» ترجیح می‌دهم از صق استفاده کنم.

این بازی یک نکته مهم در خود دارد. شما در تلاش هستید بدون این که بفهمید درباره چه حرف می‌زنید، با طرف مقابل حرف بزنید. این روش عادی ما در گفتگو نیست. اما به هر حال، شکلی از تلاش برای گفتگو است.

شما با این سبک تلاش برای حرف زدن، یک نمونه بسیار ابتدایی از هوش مصنوعی مکالمه‌محور (Conversational AI) را شبیه‌سازی کرده‌اید.

این شکل از نگاه به زبان را می‌توان «مدل کردن زبان به عنوان زنجیره‌ای از نمادها» یا «Modelling language as a sequence of symbols» نامید. به این معنا که شما می‌گویید از نظر من زبان یعنی کلماتی که پشت هم به زنجیر کشیده شده‌اند و من معتقدم هر انتخاب حلقه زنجیر به نوعی به این ربط دارد که حلقه یا حلقه‌های قبل چه بوده‌اند. هر چقدر بتوانم ویژگی‌های این زنجیر را بهتر بفهمم و به شکلی تشخیص دهم که ارتباط کلمات متوالی چیست، راحت‌تر می‌توانم به این زبان صحبت کنم.

از نظر ریاضی، مسئله بالا تفاوت جدی با این مسئله ندارد که به شما بگویند آخرین رقم از رشته اعداد زیر

حذف شده و از شما خواسته شود بگویید در رشته زیر به جای X چه رقمی قرار می گیرد:

۳۴۳۵۴۳۴۵۳۲۴۳۲۵۳۴۵۳۴۲۵۳۴۲۳

۹۸۷۰۵۶۸۷۹۶۲۷۹۴۶۷۳۹۶۴۵۹۰۲۳۹

۶۹۷۲۴۶۲۳۰۴۶۷۹۲۳۷۶۴۹۲۶۳۴

۶۷۸۴۳۶۳۴۵۳X

و شما هم بخواهید جدول زیر را پر کنید:

احتمال این که ۰ باشد ...٪ است

احتمال این که ۱ باشد ...٪ است

احتمال این که ۲ باشد ...٪ است

و ...

## مدلسازی زبان به عنوان زنجیره کلمات چه کاربردهایی دارد؟

فعلاً فرض کنید الگوریتمی طراحی کرده ایم که فقط می تواند یک کلمه بعد را در زنجیره ای از کلمات حدس بزند. خروجی اش هم شبیه همین چیزی است که کمی بالاتر دیدیم. یعنی چند کلمه پیشنهاد می دهد و می گوید که هر کلمه با چه احتمالی در ادامه خواهد آمد.

آیا چنین الگوریتمی به کار می آید؟

پاسخ مثبت است. با وجودی که این الگوریتم فعلاً در این مرحله بسیار ساده است، هم چنان کاربردهای فراوانی دارد. به این چند نمونه توجه کنید:

پیش بینی کلمات هنگام تایپ

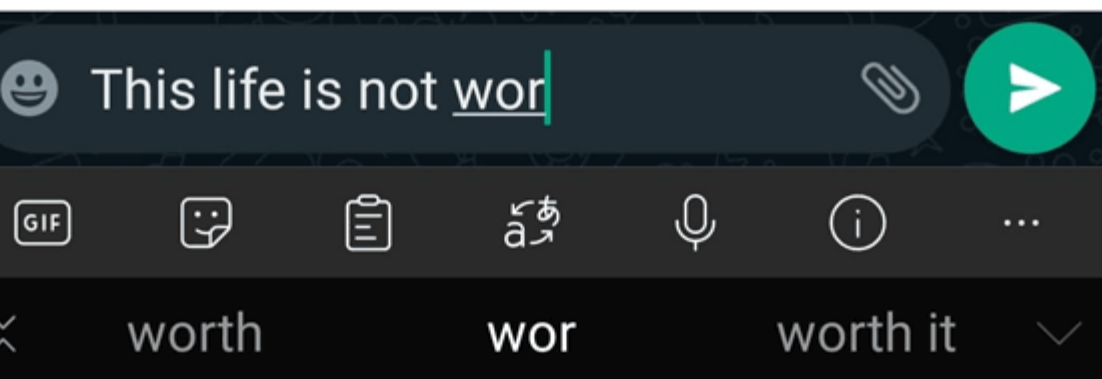
حتماً تا کنون کیبوردهای پیش بینی کننده یا Predictive Keyboards را دیده اید. تقریباً تمام

گوشی‌های هوشمند موجود در بازار، می‌توانند این قابلیت را در اختیار کاربران خود قرار دهند که با تایپ حروف اول یک کلمه (مثلاً هنگام تایپ مسیج) کلمهٔ بعدی را حدس بزنند.

تصویر زیر نمونه‌ای از پیش‌بینی کیبورد سامسونگ است:



کیبورد مایکروسافت - نصب شده روی سیستم عامل اندروید - هم پیش‌بینی‌های خودش را ارائه می‌دهد:



حدس زدن پرسش کاربران بر اساس نخستین کلمات آن‌ها

اگر در گوگل، بینگ یا هر موتور جستجوی دیگری بخواهید سوال خود را در قالب یک عبارت یا جمله بپرسید، می‌بینید که آن‌ها پرسش‌های شما را حدس می‌زنند و بر اساس احتمال و میزان اطمینانی که به هر کدام دارند فهرست می‌کنند (Sorted by Confidence).

شبیه همین کار را در بخش FAQ و بخش ثبت تیکت برخی از سایت‌ها هم می‌بینید. اگرچه این کارها بسیار شبیه همان Predictive Text Typing است، اما به خاطر برخی تفاوت‌های ظریف - که خارج از بحث ماست - معمولاً آن‌ها را زیر چتر Text Autocompletion قرار می‌دهند.

## افزایش دقت OCR

تکنولوژی OCR یا Optical Character Recognition یکی از تکنولوژی‌هایی است که زندگی همه ما را - حتی اگر ندانیم و متوجه نباشیم - تغییر داده است. هر وقت از یک متن عکس می‌گیرید و موبایل‌تان کلمات داخل متن را استخراج می‌کند، یا وقتی یک PDF قدیمی حجیم با صفحات اسکن‌شده دارید و می‌بینید که می‌توانید کلمات مد نظرتان را داخل آن جستجو کنید، مشغول استفاده از دستاوردهای OCR هستید.



OCR تکنولوژی جدیدی نیست. دغدغه‌اش حدود یک قرن وجود داشته و نمونه‌های کاربردی‌اش نزدیک به نیم قرن است که وجود دارند. خود OCR در ساده‌ترین شکل خود هم از هوش مصنوعی بهره می‌برد و زیرمجموعه الگوریتم‌های تشخیص الگو (Pattern Recognition) محسوب می‌شود. همان الگوریتم‌هایی که اثر انگشت یا چهره ما را تشخیص می‌دهند. اما این تکنولوژی در ذات خود از جنس ترتیبی یا Sequential نیست. یعنی وقتی عکسی از یک متن را در اختیار OCR قرار می‌دهید، هر کلمه می‌تواند یک «مسئله جداگانه» باشد. به عنوان مثال در تصویر زیر، من با موبایلم از دست‌نوشته‌ام عکس گرفته‌ام و اپلیکیشن کلمات را تشخیص داده و جدا کرده است:

با این حال، مدل‌های زنجیره‌ای از کلمات می‌توانند به کیفیت کار OCR کمک کرده و خطای آن را کاهش دهند. مثلاً فرض کنید OCR توانسته عبارت **feature of my Android** را تشخیص دهد. حتی اگر کلمه **phone** را بدخط و ناخوانا نوشته باشم، حدس زدن این کلمه دشوار نیست.

برای این که بهتر بتوانید کاربرد تحلیل زنجیره‌ای در OCR را درک کنید، در کیبورد پیش‌بینی‌کننده سامسونگ چند کلمه تایپ کردم:

همان‌طور که می‌بینید، یک تحلیل ساده زنجیره نشان می‌دهد که کلمه بعدی من به احتمال زیاد یکی از سه کلمه phone و app و device است. پس همین که OCR بتواند تشخیص دهد که دستخط من به کدامیک از سه کلمه نزدیک است، برای تشخیص کلمه کافی است. یا اگر بخواهم دقیق‌تر بگویم: بعد از این که OCR حدس زد کلمه من phone است، می‌تواند با تحلیل مدل زنجیره‌ای، اطمینان خود را از نتیجه‌ی محاسباتش افزایش دهد.

این کار را می‌توان Sequence Modeling Assisted OCR نامید. این نوع استفاده از الگوریتم‌ها به عنوان دستیار یکدیگر در دنیای هوش مصنوعی بسیار رایج است.

کمک به نرم‌افزارهای تبدیل صدا به متن

نرم‌افزارهای تشخیص صدا (Speech Recognition) و تبدیل صدا به متن هم از جمله نرم‌افزارهای پرکاربرد در سال‌های اخیر هستند. بسیاری از ما از سرویس‌هایی مثل Google Voice Typing استفاده کرده‌ایم. قدیمی‌ترها هم بیش از دو دهه است که با نرم‌افزارهایی مانند Dragon Naturally

**Speaking** که هم‌اکنون شرکت نوانس آن را عرضه می‌کند کار کرده‌اند (**Nuance** اکنون در مالکیت مایکروسافت است). **Dragon** آن‌قدر تخصصی شده که اگر وکیل یا مذاکره‌کننده و تنظیم‌کننده قراردادها بین‌المللی باشید، به جای نسخه معمولی نرم‌افزار تشخیص صدای خود، نرم‌افزار **Dragon Legal** **Anywhere** را به شما عرضه می‌کند تا در تشخیص صدای شما و تبدیل آن به کلمات و متن قرارداد، کمترین خطا به وجود بیاید.

همه نرم‌افزارهای تشخیص صدا از مدل‌سازی زنجیره کلمات استفاده نمی‌کنند. بسیاری از آن‌ها صرفاً یک شبکه عصبی متعارف هستند که با صدا و متن آموزش دیده‌اند. مثلاً فرض کنید شما هزاران کتاب صوتی دارید که خوانندگان حرفه‌ای آن‌ها را خوانده‌اند. از سوی دیگر متن همان کتاب‌ها را هم در اختیار دارید. اگر شبکه عصبی با این داده‌ها آموزش ببیند (**Train** شود) می‌تواند با دقت قابل‌قبولی صدا را تشخیص دهد. منظورم از **Train** شدن، به ساده‌ترین زبان، این است که کلمه به کلمه و جمله به جمله، صدا را به شبکه عصبی بدهند و به شبکه بگویند که هر صدا با چه کلمه‌ای متناظر است. بعد از مدتی که شبکه تلفظ‌های متعددی از یک کلمه را دریافت کرد و آموزش دید، می‌تواند تلفظ‌هایی را هم که اندکی تفاوت دارند تشخیص دهد.

اما همین نرم‌افزارهای تشخیص صدا، اگر به مدل‌سازی زنجیره کلمات مجهز شوند، خطای آن‌ها بسیار کمتر خواهد شد. مثلاً اگر عبارت **The Universal Laws of Life and Death** را برای یک برنامه تشخیص صدا بخوانید و **Death** را نامفهوم تلفظ کنید، یک برنامه شبکه عصبی که تحلیل زنجیره‌ای انجام نمی‌دهد، ممکن است آن را **there** تشخیص دهد. اما مدل‌سازی زنجیره کلمات هم به کمکش بیاید، می‌داند که بعد از **of Life and** بعید است کلمه **there** وجود داشته باشد و احتمالاً کلمه **death** تلفظ شده است.

در واقع اگر در سال‌های اخیر می‌بینیم اغلب ما می‌توانیم به سادگی با نرم‌افزارهای **Voice Typing** حرف بزنیم و جملات مان هم با دقت خوبی تشخیص داده می‌شود، بیشتر از این که نشان‌دهنده بهبود تلفظ ما باشد، ناشی از تقویت موتورهای تشخیص صدا با پشتیبانی مدل‌های زنجیره کلمات است.

## دو رویکرد متفاوت به مدل‌سازی زبان

هر چه تا این‌جا گفتیم، فقط درباره یک فرض بود: «فرض زبان به عنوان یک زنجیره از نمادها.» حالا سوال این است که این زنجیره از نمادها را با چه رویکردهایی می‌توان تحلیل کرد.

برای مدل‌سازی هر زنجیره‌ای از نمادها دو رویکرد وجود دارند که بسیار شناخته‌شده هستند و به کار ما می‌آیند. برای کسی که قرار است صرفاً کاربر چت جی پی تی باشد، همین که کلیت این دو رویکرد را بشناسد کافی

است. حتی با یک شناخت اندک هم می‌توان نقاط قوت و ضعف چت جی پی تی (و سایر مدل‌های هوش مصنوعی پردازش زبان طبیعی) را تا حد خوبی درک کرد.

یک رویکرد مدل‌سازی زبان، مدل‌سازی بر مبنای احتمال است. چنین مدل‌هایی را Probabilistic Language Models می‌نامند. مثالی که من اوایل همین مطلب آوردم، با رویکرد احتمالی بود. چون تقریباً همه حرف‌هایم از این جنس بود که «معمولاً وقتی کلمه ... در جمله می‌آید، احتمال این که کلمه ... بعد از آن بیاید زیاد است.»

رویکرد دومی که در مدل‌سازی به کار می‌رود، استفاده از شبکه‌های عصبی و یادگیری عمیق است. در ادامه هر یک از این دو رویکرد را در چند جمله توضیح می‌دهم (با قربانی کردن کامل دقت علمی، البته بدون این که به هدف بحث خدشه‌ای وارد شود).

### مدل‌سازی مبتنی بر احتمال

در این روش، شما هر چقدر می‌توانید متن جمع می‌کنید. متن اخبار، روزنامه‌ها، کتاب‌ها، گفتگوها و ...

سپس تصمیم می‌گیرید که زنجیره چند کلمه‌ای برای شما مهم است. مثلاً ممکن است بگویید من فقط می‌خواهم دو کلمه را معیار قرار دهم. یعنی اگر دیدم «سیب قرمز» و «سیب سرخ» و «سیب رسیده» و «سیب کال» و «سیب درشت» در متن‌ها زیاد به کار رفته، دیگر هر جا سیب دیدم، فرض می‌کنم بعد از سیب باید یکی از کلمه‌ها باشد. خودم هم جایی سیب را به کار بردم، اگر خواستم کلمه‌ای به آن اضافه کنم، از همین‌ها استفاده می‌کنم.

حالا یک نفر می‌پرسد: کلمه قبل از سیب برایت مهم نیست: «مثل سیب ...» و «این سیب ...» و «برایم سیب ...» را یک‌جور فرض می‌کنی؟

شما می‌بینید بهتر است سه کلمه را در نظر بگیرید. چون «مثل سیب ...» با کلماتی مثل سرخ و رسیده کامل می‌شود. اما «برایم سیب ...» علاوه بر سرخ و درشت و ... می‌تواند با «بیاور» و «پوست» و ... هم کامل شود (برایم سیب بیاور. برایم سیب پوست بکن و ...).

به این نوع مدل‌سازی احتمالی n-gram می‌گویند. n تعداد کلماتی است که در آمارگیری و محاسبه احتمال لحاظ می‌کنید. روش دو کلمه‌ای را Bi-gram می‌گویند. روش سه کلمه را Tri-gram می‌گویند و ...

اگر گفتند  $n$ -gram منظور این است که همه ترکیب‌های  $n$  کلمه‌ای را استخراج کرده‌اند و هر وقت  $n-1$  کلمه را به آن‌ها بدهید، از روی گزارش‌های آماری خود حدس می‌زنند که برای آخرین حلقه این زنجیر (کلمه  $n$  ام) چه گزینه‌هایی و با چه احتمالی وجود دارد.

این روش ظاهراً جذاب به نظر می‌رسد. اما ضعف بزرگی دارد و آن این است که این نوع تحلیل زنجیره، بی‌حافظه است (Memorylessness).

بگذارید این مفهوم را با یک مثال ساده نشان دهیم. فرض کنید تحلیل five-gram انجام می‌دهیم (پنج کلمه را با هم در نظر می‌گیریم). من چهار کلمه اول یک ترکیب پنج‌کلمه‌ای را به شما می‌دهم و از شما می‌خواهم با توجه به داده‌های آماری خود بگویید کلمه پنجم چیست:

«نظام جمهوری اسلامی در ...»

چه ایده‌هایی در مورد کلمه پنجم دارید؟ چهل، این، عالم، جهان، ناتوانی، ناکارآمدی، جریان‌سازی، قلب، ایران، همه‌پرسی، روزهای، همه، عرصه، طول، بخش، مواجهه، مقابل، شکل، مدار، مسیر، برابر، مراحل، اندیشه، صدر و ...

طبیعتاً بر اساس داده‌های خود می‌توانید برای هر یک از این کلمات (و ده‌ها کلمه دیگر که می‌توانند به عنوان کلمه پنجم بیانند) یک احتمال حدس بزنید. اما همه می‌دانیم که این کلمات از زمین تا آسمان با هم فرق دارند. این جمله در زبان یکی از مسئولان نظام، یک روزنامه‌نگار، تربیون نماز جمعه، یکی از براندازها، یک اصلاح‌طلب، یک کاندیدای نمایندگی مجلس و ... می‌تواند به شکل‌های متفاوتی تکمیل شود. اگر تحلیل خود را از five-gram به six-gram ارتقاء دهیم، کمی وضع بهتر می‌شود. همین‌طور اگر سیستم هوشمند خود را به seven-gram و eight-gram و ... برسانیم، نتایج رضایت‌بخش‌تری خواهیم داشت.

وقتی می‌گوییم Memorylessness یا بی‌حافظگی، منظورمان این است که در  $n$ -gram برای سیستم ما اصلاً فرق نمی‌کند که قبل از این  $n$  کلمه چه چیزی گفته شده است. و همین بی‌توجهی به گذشته، باعث می‌شود خطای الگوریتم افزایش پیدا کند.

متخصصان ریاضی معمولاً این نوع نگاه به متن را رویکرد مارکوفی می‌نامند. چون مدل مارکوف در احتمال به سراغ مدل‌سازی پدیده‌هایی می‌رود که «فرض می‌کند» گذشته آن‌ها برای دانستن آینده‌شان مورد نیاز نیست.

ساده‌ترین مثال زنجیره مارکوف، پرتاب سکه یا همان شیر یا خط کردن است. به شما می‌گویند که سکه

سالمی روی زمین افتاده است. سکه را برمی داریم و پرت می کنیم تا دوباره روی زمین بیفتد، چقدر احتمال دارد که وضعیت سکه تغییر کند؟ (اگر شیر بوده روی خط بیفتد و اگر خط بوده روی شیر بیفتد). شما به سادگی می گوئید ۵۰٪. آیا برایتان مهم است که قبلاً که با این سکه بازی کرده اند شیر آمده یا خط یا با چه ترتیبی شیر و خط آمده؟ قطعاً نه.

حتی اگر سکه شما سالم نباشد و اصطلاحاً Unfair باشد، یعنی مثلاً احتمال شیر آمدن ۴۰٪ و خط آمدنش ۶۰٪ باشد، باز هم پرتاب سکه را می توان با زنجیره مارکوف مدل کرد. چون همین که احتمال ۴۰٪ و ۶۰٪ را می دانید (که یک ویژگی ثابت سکه است) و وضعیت الان سکه را می دانید، می توانید احتمال تغییر وضعیت را حساب کنید. اتفاق های گذشته و این که قبلاً به چه ترتیبی شیر یا خط آمده، هیچ تأثیری رو پیش بینی شما ندارد. سکه ۴۰ / ۶۰ را در نظر بگیرید:

شیر - خط - شیر - شیر - خط - شیر - خط - بعدی؟

شیر - شیر - شیر - خط - خط - خط - خط - بعدی؟

در هر دو حالت، شما باید با احتمال ۴۰ - ۶۰ درباره وضعیت بعدی حرف بزنید. این که چند بار آخر چه اتفاقی افتاده، مهم نیست.

زنجیره مارکوف Memoryless یا بی حافظه است (دقیق تر بگویم: برای مدلسازی رویدادهای بی حافظه و ناوابسته به گذشته به کار می آید). زنجیره مارکوف در پیچیده ترین حالت خود می گوید: «آینده فقط تابع حال است و نه گذشته.» در n-gram هم دقیقاً همین کار را می کنیم. مثلاً در five-gram می گوئیم: من برای تشخیص کلمه پنجم، فقط به چهار کلمه قبل نگاه می کنم. اصلاً مهم نیست که کلمات قبل از آن چه بوده اند.

با این حال، نباید قدرت زنجیره مارکوف را دست کم بگیریم. همین شیوه n-gram بسیاری از کارهای تحلیل زبان طبیعی را به خوبی انجام می دهد. تقریباً تمام مثال هایی که در بالا گفتم، یعنی تشخیص صوت و تشخیص دستنوشته و مانند این ها می تواند با دقت قابل قبولی با همین الگوریتم های آماری انجام شود.

توضیحی که خیلی مهم نیست | آیا اگر ما ظرفیت محاسباتی نامحدود داشتیم و مثلاً می توانستیم به جای five-gram سراغ five-hundred-gram برویم و همه الگوهای قبلی را هم با هم ترکیب کنیم، این سیستم بی نقص می شد؟ یعنی مثلاً شما داده های دو-گرام و سه-گرام و چهار-گرام و پنج-گرام تا پانصد-گرام را گردآوری کنید و با ترکیب آن ها (به شکلی که فعلاً برایمان مهم نیست) یک الگوریتم پردازش زبان طبیعی

احتمال محور بسازید.

پاسخ این است که: نه! مشکل دیگر این روش، **Overfitting** است. داده‌های بیشتر می‌توانند کار را خراب کنند. اگر **Overfitting** را نمی‌شناسید، این چند خط را نادیده بگیرید | پایان توضیحی که خیلی مهم نبود.

## مدل زبانی با تکیه بر شبکه های عصبی

دومین روشی که برای مدلسازی زبان (یا هر زنجیره‌ای از نمادها) به کار می‌رود، استفاده از شبکه های عصبی (Neural Networks) است. ایده شبکه های عصبی ایده‌ای نسبتاً قدیمی است. البته معمولاً هر چندسال یک‌بار با یک نام‌گذاری جدید دوباره شنیده و شناخته می‌شود. مثلاً شاید برایتان جالب باشد که کسانی که امروز از یادگیری عمیق و **Deep Learning** حرف می‌زنند، تقریباً مبتنی بر روشی کار می‌کنند که حوالی ۱۹۵۰ ابداع شد و و البته حدود ۱۹۸۰ تحولی بنیادین را تجربه کرد. بعد از آن، اگر چه پیشرفت‌های مهمی در یادگیری با شبکه های عصبی انجام شده، اما آن قدر که بعضی تازه‌متخصصان هوش مصنوعی درباره‌اش به ما می‌گویند، بنیادین نبوده است (این توضیح از آن جهت مهم است که از اسم‌ها و اصطلاحات عجیبی که این روزها می‌شنوید نترسید).

بگذارید با یک مقایسه استعاری توضیح دهم.

اگر کارهای دهه پنجاه دونالد هب (Donald Hebb) را - در مدل‌سازی عملکرد نورو و یادگیری در آن - شبیه کارهای کوپرنیک و کپلر در نظر بگیریم، نسل جدیدی از شبکه های عصبی را که در دهه هشتاد شکل گرفت، می‌توان گامی بزرگ به اندازه کار نیوتن تلقی کرد. حالا به یک طراح سیستم‌های دینامیکی خودرو فکر کنید که سه قرن بعد از نیوتن، خودروهای امروز ما را طراحی می‌کند. به یک معنا، می‌توان گفت او خیلی جلوتر از تفکر نیوتنی است. به معنایی دیگر، او هنوز هم شکل توسعه‌یافته همان مدل را به کار می‌گیرد.

ماجرای شبکه های عصبی هم همین است. گام بزرگ اول را دونالد هب برداشت. دوران هب دوران جدیدی بود که ما نورو را یک المان نسبتاً ساده دیدیم که سینگال‌های الکتریکی را پردازش می‌کند (قبل از آن از جن و موجودات غیرارگانیک تا روح، بسته به باورهای شما، در سوراخ‌های بین همین نوروها لانه کرده بودند). در دهه هشتاد، این ایده مطرح شد که اگر ما تعدادی نورو را شبیه‌سازی می‌کنیم، چرا از خروجی خود نوروها به عنوان ورودی‌شان استفاده نکنیم؟ (فرض کنید یک نفر امروز بگوید: چرا چت جی پی تی فقط باید متن‌های انسانی را بخواند؟ می‌توانیم خروجی‌های خود چت جی پی تی را هم به خودش بدهیم تا آن‌ها را هم بررسی کند و الگوی یادگیری خودش قرار دهد). ایده ورودی گرفتن از خروجی، ظاهراً ساده است، اما

ویژگی‌های دینامیکی سیستم را به کلی عوض می‌کند و ظرفیت‌های جدیدی برای آن می‌سازد. این شبکه‌ها که آن‌ها را با نام RNN (مخفف Recurrent Neural Networks) سالها وجود داشتند و از آن‌ها استفاده می‌شد، تا ایده‌های جدیدی برای بهبود ظرفیت پردازش آن‌ها شکل گرفت (هم برای عمق پردازش و هم برای پردازش موازی). اصطلاحاتی مثل Attention Mechanisms (مکانیزم‌های توجه) و Self-attention (توجه به خود) که بعداً با نام شیک‌تر و جذاب‌تر Transformer ها همراه شدند، به این نوع تغییرات اشاره دارند.

اما با همه این تغییرات و تحولاتی که طی نزدیک به نیم قرن اخیر در شبکه‌های عصبی به وجود آمده، دو ویژگی که نقطه قوت و ماهیت شبکه عصبی را تشکیل می‌داده تغییر نکرده است: ذخیره سازی توزیع شده، خاطره داشتن از گذشته (ویژگی‌ها بیشتر است. دو مورد به کار ما می‌آید).

این دو ویژگی بسیار مهم هستند و هر چقدر آن‌ها را بهتر درک کنیم، ظرفیت‌ها و محدودیت‌های شبکه‌های عصبی و مدل‌های زبانی مبتنی بر شبکه‌های عصبی (مانند Chat GPT) را بهتر درک خواهیم کرد.

## یک مثال بسیار ساده‌شده از یادگیری توزیعی (شبیه شبکه عصبی)

دوستانی که با شبکه‌های عصبی و یادگیری ماشینی آشنایی دارند، اگر مثال زیر را با الگوهایی که از شبکه‌های عصبی در ذهن دارند مقایسه کنند، احتمالاً از ساده‌سازی بیش‌ازحد من خوشحال نمی‌شوند. اما اگر بپذیرند که صرفاً می‌خواهیم با مفهوم توزیع‌شدگی (مستقل از پیچیدگی‌های شبکه‌های عصبی و سیستم‌های یادگیری عمیق) آشنا شویم، به گمانم این مثال را خواهند پسندید. این مثال صرفاً از این جهت به ذهنم رسید که الان یک فروشگاه عطر روبه‌رویم قرار دارد و فروشنده مدتی است که با جدیت من را نگاه می‌کند. اگر جای دیگری نشسته بودم و چیز دیگری جلوی من بود، مثال دیگری می‌زدم.

فرض کنیم به یک کارگاه ساده تولید عطر رفته‌ایم. این کارگاه، پنج ماده مختلف دارد که آن‌ها را به نسبت‌های متفاوت با هم ترکیب می‌کند و انواع عطرها را می‌سازد.

چهار نفر هم داریم که سلیقه عطر نسبتاً متفاوتی دارند. هر عطری که از این کارگاه بیرون می‌آید، به هر چهار نفر عرضه می‌شود و هر کدام صرفاً با یک کلمه پاسخ می‌دهند: پسندیدم. نپسندیدم.

فرض کنید من یک دستگاه به شکل زیر درست کرده‌ام:





روی این دستگاه ۳۵ ولوم (شبه ولوم رادیو - پتانسیومترهایی که برای تنظیم صدا و ... استفاده می‌کنیم) قرار دارد و بالای آن چهار عدد لامپ نصب شده است. این پتانسیومترها با مدار پیچیده‌ای که ما اطلاعی از آن نداریم به هم وصل هستند. و احتمالاً خروجی بعضی از آن‌ها ورودی چند پتانسیومتر دیگر را تأمین می‌کند. به هر حال، هیچ نوع دسترسی به داخل جعبه نداریم و برایمان مهم هم نیست.

اما می‌دانیم که وقتی پتانسیومترها را می‌چرخانیم، در بعضی محدوده‌ها که قرار می‌گیرند، بعضی از این لامپ‌ها خاموش یا روشن می‌شوند.

حالا می‌توانیم یک بازی جالب انجام دهیم:

من به شما ترکیب مواد یک عطر را می‌گویم: ۲۳٪ از ماده یک و ۲۷٪ از ماده دو و ۳۰٪ از ماده سه و ۲۰٪ از ماده چهار و ۰٪ از ماده پنج.

هم‌چنین به شما می‌گویم که این عطر را نفر اول و دوم و چهارم پسندیدند و نفر سوم آن را نپسندید.

شما باید پنج پتانسیومتر سمت چپ را بر اساس ترکیب مواد عطر تنظیم کنید. سپس با سی پتانسیومتر دیگر آن‌قدر بازی کنید تا چراغ‌های بالا دقیقاً مشابه نظر آن چهار آدم روشن شوند. یعنی چراغ اول و دوم و چهارم روشن شوند و سومی خاموش بماند.

تا این‌جا کار بازی سخت نیست. اما حالا می‌توانیم یک گام جلوتر برویم.

مشخصات عطر بعدی را به شما می‌دهم: ۵٪ - ۱۰٪ - ۱۰٪ - ۲۰٪ - ۵۵٪ و باز هم نظر آن چهار نفر: اولی و چهارمی پسندیدند.

حالا شما باید باز با پتانسیومترها بازی کنید. و دنبال تنظیمی بگردید که چراغ اول و چهارم را روشن کند. اما باید سعی کنید این تنظیم را طوری انجام دهید که اگر دوباره پتانسیومترهای چپ را روی همان تنظیم قدیمی (۲۳-۲۷-۳۰-۲۰-۰) بردیم، چراغ‌ها مشابه حالت اول روشن شوند (اول، دوم و چهارم).

این بازی را می‌توانیم ده‌ها بار با عطرهاى مختلف انجام دهیم. کم‌کم دستگاه شما به نقطه‌ای می‌رسد که اگر ترکیب عطری را به شما بگویم که قبلاً نگفته بودم، دستگاه می‌تواند حدس بزند که کدامیک از این چهار نفر آن را می‌پسندند و کدام نمی‌پسندند.

خوش‌بختانه اوضاع در دنیای هوش مصنوعی بسیار بهتر است. چون رابطه بین این پتانسیومترها را می‌دانیم.

همچنین محققان هوش مصنوعی سال‌ها جستجو کرده‌اند و توانسته‌اند پتانسیومترهایی با ویژگی‌های مناسب بسازند که ورودی‌ها را به شکل کارآمدی ترکیب کند و خروجی بدهد. ضمناً الگوریتم‌ها و روش‌های متنوعی هم ایجاد شده که کمک می‌کند شما پتانسیومترها را به شکل بهینه‌ای تنظیم کنید. یعنی به شکلی که تا حد امکان، نتایج قبلی از بین نرود و ظرفیت تحلیل نمونه‌های جدید هم به سیستم اضافه شود.

اما فعلاً بحث ما این نیست. ما با همان جعبهٔ خودمان کار داریم. چون فقط می‌خواهیم مفهوم Distributedness یا توزیع‌شدگی را بفهمیم.

بعد از این که چند روز با دستگاه بازی کردید و پتانسیومترها را تنظیم کردید، نهایتاً تمام تلاش شما در یک ماتریس ۵ در ۶ خلاصه می‌شود. مثلاً چیزی شبیه این:

هر عددی که در این جدول می‌بینید، متناظر با یکی از پتانسیومترها در اسباب‌بازی بالاست. عملاً تمام تلاش شما و یادگیری شما و آنچه از سلیقهٔ آن چهار نفر می‌دانید، در این جدول خلاصه شده است.

چند اتفاق جالب که در این جدول افتاده را با هم مرور کنیم:

یکی این که داده‌های این ماتریس، کاملاً توزیع‌شده هستند. شما نمی‌دانید و نمی‌توانید بگویید کدام عدد به

سلیقه‌ی چه کسی مربوط است. فقط می‌توانید بگویید تمام چیزی که از سلیقه‌ی این آدم‌ها دارم در این جدول ذخیره شده است.

دیگر این‌که این ماتریس، حافظه دارد. یعنی تمام نمونه‌هایی که به آن نشان داده‌اید و تجربیاتی که داشته، به نوعی جایی در آن ذخیره شده است. شما طبیعتاً هر بار سعی می‌کنید با کمترین تغییر در تنظیمات پتانسیومترها، کاری کنید که اسباب‌بازی‌تان با نمونه‌ی جدید هم جور در بیاید و بتواند آن را هم تحلیل کند.

سوم این‌که در n-gram سیستم شما عملاً می‌تواند در مورد زنجیره‌هایی که قبلاً دیده و در پیکره یا دیتابیس آن وجود داشته اظهارنظر (قابل‌اتکا) کند. اما در این‌جا سیستم این ظرفیت را دارد که در مورد ترکیب عطری که پیش از این هرگز ندیده هم (طبیعتاً با یک محدوده‌ی خطا) نظر بدهد. یعنی یک ترکیب عطر را که اصلاً تا حالا وجود نداشته به آن بدهید و حدس بزنند که چه کسانی آن را می‌پسندند.

## این مطلب - با اولویت پایین - تکمیل می‌شود.

چند مطلب پیشنهادی:

- [دنیای دیجیتال و شکل‌گیری حافظه بیرونی - قسمت اول](#)
- [آیا تسلط بر زبان انگلیسی، توانایی درک و تحلیل ما را افزایش می‌دهد؟](#)
- [دگرگونی نقش روایت‌گران در عصر تکثیر دیجیتالی](#)
- [نظریه اطلاعات - کلود شانون و درک کلان سیستم‌های پیچیده \(+ کمی طعم ریاضی\)](#)
- [خودرو بدون راننده گوگل: راننده‌هایی که آدم نیستند!](#)
- [شبکه‌های اجتماعی: شکل جدید یک ساختار قدیمی](#)

با متمعن:

- [کتاب ادیان مرتبط با یوفوها \(اشیاء ناشناس پرنده\)](#)
- [مالی رفتاری در زندگی روزمره | جریمه داده‌ام اما ناراحت نیستم!](#)
- [آیا برندگان جایزه نوبل بعد از دریافت جایزه افت می‌کنند؟](#)
- [روش اصلاح دولت | همه را بیاورید یک پله پایین‌تر!](#)
- [اولین تماس تلفنی تاریخ | قدرت و محدودیت یک ابزار در یک جمله!](#)
- [خواندن اخبار کافی نیست](#)
- [اثر تماشاگر | خلاقیت در طراحی مینیمال جلد کتاب](#)

فایل‌های صوتی مذاکره آموزش زبان انگلیسی آموزش ارتباطات و مذاکره خودشناسی

آموزش مدیریت کسب و کار (MBA) کارآفرینی کسب و کار دیجیتال

ویژگی‌های انسان تحصیل کرده آموزش حرفه‌ای‌گری در محیط کار