چت جی‌پی‌تی، ساختار هوش مصنوعی مولد

به‌همراه مقایسه فناوری‌های چت باتی ۲۰۲۴ جهان

استفان ولفرام

ترجمه: عباس عجمی، احسان نجاری، سید حسن موسوی نژاد

انتشارات پندار پارس

سرشناسه : ولفرم، استیون، ۱۹۵۹ - م. Wolfram, Stephen
عنوان و نام پدیدآور : چت جی پی تی، ساختار هوش مصنوعی مولد به همراه مقایسه فناوری‌های چت باتی ۲۰۲۴ جهان/ استفان ولفرام؛ ترجمه عباس عجمی، احسان نجاری، سیدحسن موسوی‌نژاد.
مشخصات نشر : تهران: پندار پارس، ۱۴۰۳.
مشخصات ظاهری : ۱۷۰ ص.
شابک : - - - - -
وضعیت فهرست نویسی : فیپا
یادداشت : عنوان اصلی: What is ChatGPT doing ... and why does it work?, [2023].
موضوع : چت جی پی تی - ChatGPT
موضوع : پردازش زبان طبیعی -- برنامه‌های کامپیوتری
Natural language processing (Computer science) -- Computer programs
شبکه‌های عصبی)کامپیوتر(Neural networks (Computer science) ولفرم)زبان برنامه‌نویسی کامپیوتر(

شناسه افزوده : عجمی، عباس، ۱۳۶۳-، مترجم
شناسه افزوده : نجاری، احسان، ۱۳۷۱-، مترجم
شناسه افزوده : موسوی‌نژاد، سید حسن، ۱۳۶۹-، مترجم
رده بندی کنگره : QA۷۶/۹
رده بندی دیویی : ۰۰۶/۳۵
شماره کتابشناسی ملی : ۹۶۶۲۴۶۲
اطلاعات رکورد کتابشناسی : فیپا

انتشارات پندارپارس

دفتر فروش: انقلاب، ابتدای کارگرجنوبی، کوی رشتچی، شماره ۱۴، واحد ۱۶ **www.pendarepars.com**
تلفن: ۶۶۵۷۲۳۳۵ت ۶۶۹۲۶۵۷۸ همراه: ۰۹۱۲۲۴۵۲۳۴۸ **info@pendarepars.com**

نام کتاب : چت جی‌پی‌تی، ساختار هوش مصنوعی مولد (به‌همراه مقایسه فناوری‌های چت باتی ۲۰۲۴ جهان)
ناشر : انتشارات پندار پارس
تالیف : استفان ولفرام
ترجمه : عباس عجمی، احسان نجاری، سید حسن موسوی نژاد
چاپ نخست : مرداد ۱۴۰۳
شمارگان : ۱۰۰
چاپ، صحافی : روز
قیمت : ۲۰۰.۰۰۰ تومان شابک: ۳-۳۳-۷۷۸۵-۶۲۲-۹۷۸

هرگونه کپی برداری، تکثیر و چاپ کاغذی یا الکترونیکی از این کتاب بدون اجازه ناشر تخلف بوده و پیگرد قانونی دارد

فهرست

مقدمه ۸
پیشگفتار ۹
درباره نویسنده ۱۰

فصل اول: چرایی و چگونگی چت جی‌پی‌تی ۱۱
مثل اضافه کردن یک کلمه در زمان است ۱۱
احتمالات از کجا می آیند؟ ۱۷
مدل چیست؟ ۲۲
مدل هایی برای کارهای شِبهِ انسانی ۲۵
شبکه های عصبی ۲۷
یادگیری ماشینی و آموزش شبکه های عصبی ۳۸
تمرین و دانش آموزش شبکه عصبی ۴۴
مطمئناً یک شبکه به اندازه کافی بزرگ می تواند هر کاری را انجام دهد! ۵۲
مفهوم جانشانی ۵۵
درون چت جی‌پی‌تی ۶۱
آموزش چت جی‌پی‌تی ۷۰
فراتر از آموزش پایه ۷۳
واقعاً چه چیزی به چت جی‌پی‌تی اجازه می دهد کار کند؟ ۷۵
معنی فضا و قوانین معنایی حرکت ۸۰
گرامر معنایی و قدرت زبان محاسباتی ۸۶
بنابراین ... چرایی و چگونگی چت جی‌پی‌تی ۹۰
تقدیر و تشکر نویسنده ۹۳

فهرست

فصل دوم: وافرام آلفا به عنوان راهی برای آوردن ابرقدرت های دانش محاسباتی به چت جی‌پی‌تی ۹۴
چت جی‌پی‌تی و ولفرام آلفا ۹۴
یک مثال ساده ۹۶
چند مثال دیگر ۹۹
راه پیش رو ۱۱۱

فصل آخر: مقایسه معروف ترین چت بات های دنیا در سال ۲۰۲۴ ۱۱۵
چت جی‌پی‌تی۳٫۵ در مقابل چت جی‌پی‌تی۴ ۱۱۵
گوگل بارد در مقابل چت جی‌پی‌تی ۱۲۳
کوپایلوت در مقابل چت جی‌پی‌تی ۱۴۷
گرامرلی در مقابل چت جی‌پی‌تی ۱۶۱

مقدمه

این روزها یکی از کلیدی‌ترین و چالش برانگیزترین موضوعات فناوری و هوش مصنوعی، «هوش مصنوعی مولّد[1]» است. کتاب هایی نیز در این حوزه در دنیا به چاپ رسیده است. یکی از اثرگذارترین و کامل‌ترین کتاب های هوش مصنوعی در این موضوع مهم، کتابی با عنوان

«What Is ChatGPT Doing … and Why Does It Work» است. این کتاب مرجع از پرفروش‌ترین کتاب‌های هوش مصنوعی سایت آمازون در سال ۲۰۲۳ شد. سَم آلتمَن، رئیس شرکت اوپن‌ای‌آی که مبدع چت جی‌پی‌تی است، نظری نسبت به کتاب مطرح شده دارد. وی گفته است « این کتاب بهترین توصیفی است که تاکنون در حوزه چت جی‌پی‌تی دیده‌ام »

حال کتابی که پیش روی شماست، حاصل ترجمه و تألیف است. در قسمت ترجمه، ما آن کتاب مرجع تخصصی در حوزه هوش مصنوعی را به طور کامل و به همراه پیوست آن، ترجمه و در دو فصل اول این کتاب، ارائه کرده ایم. در فصل سوم تالیف نویسندگان نیز چت جی‌پی‌تی با فناوری های رقیب خود در سال میلادی ۲۰۲۴ به روش مقایسه ای بررسی شده است تا مخاطب کتاب، کاملا بینش عمیق و عملیاتی نسبت به مطالب علمی و پژوهشی کتاب بدست آورد. امید است این اثر مورد قبول حضرت حق قرار گیرد.

[1] Generative artificial intelligence

پیشگفتار

این کتاب تلاشی برای توضیح اصول اولیه و همچنین چرایی و چگونگی فناوری چت جی‌پی‌تی است. مطمئناً این داستان در مورد تلاقی فناوری دستیاران هوشمند با علم و فلسفه است. برای گفتن داستان، باید طیف قابل‌توجهی از ایده‌ها و اکتشافات را که در طول قرن‌ها انجام شده است، گرد هم بیاوریم. برای من هیجان‌انگیز است که ببینم مسائل زیادی که مدت‌هاست به آنها علاقه‌مند بوده‌ام در یک پیشرفت ناگهانی گرد هم آمده‌اند. از رفتار پیچیده برنامه‌های ساده گرفته با محوریت زبان و معنا، همچنین کاربردهای آن در سیستم‌های رایانه‌ای بزرگ، همه اینها بخشی از داستان چت جی‌پی‌تی هستند.

چت جی‌پی‌تی مبتنی بر مفهوم شبکه‌های عصبی است که در ابتدا در دهه ۱۹۴۰ به‌عنوان شیوه ای ایده‌آل برای بیان عملکرد مغز اختراع شد. من خودم اولین‌بار در سال ۱۹۸۳ یک شبکه عصبی را برنامه ریزی کردم و متاسفانه کار جالبی را انجام نداد. اما ۴۰ سال بعد، با رایانه‌هایی که به طور مؤثر یک میلیون بار سریع‌تر هستند، با پردازش میلیاردها صفحه وب و پس از یک سری نوآوری‌های مهندسی، وضعیت کاملاً متفاوت گردید و در کمال تعجب همگان، یک شبکه عصبی با مقیاس یک میلیارد برابر بزرگ‌تر از شبکه‌ای که من در سال ۱۹۸۳ داشتم، کارهایی را انجام می‌دهد که تصورش به دلیل منحصر به فرد بودن در حیطه تولید زبان معنادار انسانی، دشوار می‌باشد.

این کتاب از دو بخش تشکیل شده است. بخش اول توضیح چت جی‌پی‌تی و توانایی آن در انجام کارهای شبه‌انسانی تولید زبان است. دومین بخش، بیان ابزارهای محاسباتی برای فراتر رفتن از حد محاسباتی انسان است. در این بخش، موضوع استفاده از دانش ابرقدرتِ محاسباتیِ سیستم ولفرام آلفا است. تنها سه ماه از راه‌اندازی چت جی‌پی‌تی می‌گذرد، و ما تازه در حال درک پیامدهای آن چه از نظر عملی و چه فکری هستیم. اما در حال حاضر ورود این فناوری به ما یادآوری می‌کند که حتی پس از اختراعات و نوآوری‌های جدید، همچنان شگفتی‌ها ادامه دارد.

استیون ولفرام

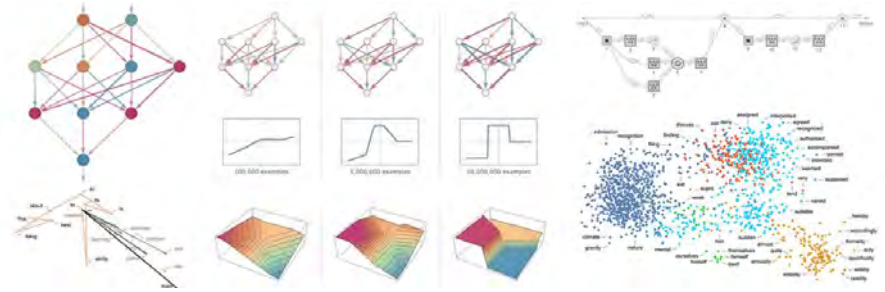
۲۸ فوریه ۲۰۲۳

درباره نویسنده

استیون ولفرام، در ۲۹ اوت ۱۹۵۹ در لندن به دنیا آمد. نویسنده کتاب «یک نوع جدید علم» و ۸ عنوان کتاب دیگر که در ۱۵ سالگی اولین مقاله علمی[۲] خود را نوشت و در سن ۲۰ سالگی مدرک دکتری فیزیک نظری خود را گرفت. او دانشمند کامپیوتر، فیزیک‌دان و تاجر بریتانیایی آمریکایی است. او به‌خاطر کارش در علوم کامپیوتر، ریاضیات و فیزیک نظری شناخته شده است. در سال ۲۰۱۲، او به عضویت انجمن ریاضی آمریکا انتخاب شد. همچنین وی به‌عنوان یک تاجر، بنیان‌گذار و مدیرعامل شرکت نرم‌افزاری ولفرام[۳] است که در آنجا به‌عنوان طراح ارشد سامانه محاسبات فنی مدرن مَتِمَتیکا[۴] و موتور پاسخگو ولفرام آلفا و مبدع زبان ولفرام[۵] مشغول به فعالیت است.

[۲] https://www.stephenwolfram.com/publications/academic/all/
[۳] Wolfram Research
[٤] https://www.wolfram.com/mathematica/
[٥] https://www.wolfram.com/language/

فصل اول: چرایی و چگونگی چت جی‌پی‌تی



مثل اضافه کردن تنها یک کلمه در زمان است

اینکه چت جی‌پی‌تی می‌تواند به طور خودکار مطلبی از جنس متن نوشته شده توسط انسان را تولید کند، قابل‌توجه و غیرمنتظره است. اما سؤال مهم این است که چگونه این کار را انجام می‌دهد؟ و چرا کار می‌کند؟ هدف من در اینجا ارائه یک طرح کلی از آنچه در داخل چت جی‌پی‌تی می‌گذرد است و سپس به بررسی چرایی توان تولید متنی معنادار می‌رسد. در ابتدا باید بگویم که من بر روی تصویر بزرگ آنچه در حال وقوع است تمرکز خواهم کرد، درحالی‌که برخی جزئیات مهندسی را ذکر می‌کنم، عمیقاً وارد آنها نمی‌شوم و ماهیت چیزی که من می‌گویم به همان اندازه در مورد دیگر «مدل‌های بزرگ زبانی[۶]» فعلی نیز صدق می‌کند.

اولین موضوع موردبحث این است که آنچه چت جی‌پی‌تی انجام می‌دهد، یک «دنباله منطقی[۷]» از هر متنی است که تاکنون به دست آورده. منظور ما از واژه «منطقی» آنچه که ما از پاسخ به نوشته‌های افراد بر روی میلیاردهای صفحه وب انتظار داریم، است.

پس بیایید فرض بگیریم که متن «بهترین مطلب در مورد هوش مصنوعی توانایی آن نسبت به» را در اختیار داریم. با خود تصور کنید میلیاردها صفحه متنی وب را که توسط انسان (در اینترنت و در کتاب های الکترونیکی) نوشته شده است را اسکن می‌کنید و همه نمونه‌های متنی را ذخیره و پردازش می‌کنید، حالا ببینید چه کلمه‌ای را بلافاصله بعد این عبارت در کسری از زمان دارید. چت جی‌پی‌تی به طور مؤثر چیزی شبیه به این را انجام می‌دهد، با این تفاوت که (همان‌طور که توضیح خواهم داد) به متن تحت‌اللفظی نگاه نمی‌کند، بلکه به دنبال نکاتی می‌گردد که با معنای

---

[۶] مدل زبانی بزرگ large language model یا به اختصار LLM، سیستم های هوش مصنوعی هستند که برای درک، تولید و پاسخگویی به زبان انسان طراحی شده اند. آنها «بزرگ» هستند زیرا حاوی میلیاردها پارامتر هستند که به آنها امکان می دهد الگوهای پیچیده در داده های زبان را پردازش کنند.

[۷] Reasonable continuation

جمله مطابقت دارد و نتیجه نهایی این است که فهرست رتبه‌بندی شده‌ای از کلمات به همراه «احتمالات» آنها را تولید می‌کند. بهترین مزیت هوش مصنوعی توانایی آن در **یادگیری، پیش‌بینی، ساختن، درک‌کردن** و **انجام‌دادن** است، در جدول زیر درصدهای این توانایی‌ها ذکر شده است:

۴.۵٪	یادگیری	**Learn**
۳.۵٪	پیش‌بینی	Predict
۳.۲٪	ساختن	Make
۳.۱٪	درک‌کردن	Understand
۲.۹٪	انجام‌دادن	Do

نکته قابل‌توجه این است که وقتی چت جی‌پی‌تی کاری مانند نوشتن یک مطلب را انجام می‌دهد، اساساً این سؤال را بارهاوبارها می‌پرسد: «باتوجه‌به متن، کلمه بعدی چه باید باشد؟» و هر بار کلمه‌ای را اضافه می‌کند (به طور دقیق‌تر، همان‌طور که توضیح خواهم داد، اضافه‌کردن ممکن است فقط یک «توکن[۸]» باشد که می‌تواند فقط بخشی از یک کلمه را در برگیرد، به همین دلیل است که گاهی اوقات چت جی‌پی‌تی کلمات جدیدی را می‌سازد).

در هر مرحله فهرستی از کلمات با احتمالات آن‌ها را دریافت می‌کند. اما در واقع [تصمیم مهم آن است که] کدام یک را باید برای افزودن به مطلب در حال نوشتن انتخاب کند؟ ممکن است کسی فکر کند که این کلمه باید دارای «بالاترین رتبه[۹]» باشد (یعنی کلمه‌ای که بیشترین احتمال به آن اختصاص‌داده‌شده است). اما اینجاست که گویا جادوگری مخفیانه وارد می‌شود. به دلایلی - که شاید روزی از پس درک وجه علمی آن بر بیاییم - اگر همیشه کلمه‌ای که بالاترین رتبه را دارد انتخاب کنیم، معمولاً به یک مطلب «یکنواخت» می‌رسیم و در نوشتهٔ ما خلاقیتی دیده نمی شود (و حتی گاهی اوقات کلمه‌به‌کلمه، لغات تکرار می‌شوند). اما اگر گاهی (به طور تصادفی) کلماتی با رتبه پایین‌تر انتخاب کنیم، یک مطلب «جذاب‌تر» خواهیم داشت.

این واقعیت که در اینجا [ویژگی] تصادفی بودن وجود دارد به این معنی است که اگر یک درخواست را چند بار ارسال کنیم، احتمالاً هر بار مطالب مختلفی دریافت می‌کنیم. در همراهی با ایده جادو، یک پارامتر مشخص که «دما[۱۰]» نامیده می شود وجود دارد که تعیین می‌کند چند بار از کلمات با رتبه پایین‌تر استفاده شود. برای تولید مقاله [از طریق هوش مصنوعی] به نظر می‌رسد که دمای ۰.۸ بهترین باشد. (شایان‌ذکر است که هیچ نظریه‌ای در اینجا استفاده نمی‌شود،

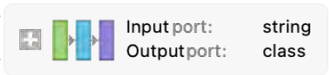
[۸] Token
[۹] Highest-ranked
[۱۰] Temperature

بلکه فقط موضوع این است که چه چیزی در عمل کارایی بهتری دارد، به‌عنوان‌مثال از مفهوم «دما» در توزیع نمایی فیزیک آماری استفاده می‌شود، اما هیچ ارتباط «فیزیکی» در اینجا وجود ندارد، حداقل تا آنجا که ما می‌دانیم).

قبل از اینکه ادامه دهیم، باید توضیح دهم که برای اهداف نمایش، من عمدتاً از سیستم کامل چت جی‌پی‌تی استفاده نمی‌کنم. در عوض، من معمولاً با سیستم ساده تر جی‌پی‌تی ۲ کار می‌کنم که این ویژگی خوب را دارد که به‌اندازه کافی کوچک است که می‌تواند روی یک رایانه رومیزی استاندارد اجرا شود؛ بنابراین، اساساً برای همه چیزهایی که نشان می‌دهم، می‌توانم کد زبان آشکار ولفرام را اضافه کنم که می‌توانید فوراً روی رایانه خود اجرا کنید.

برای مثال، در اینجا نحوه به‌دست‌آوردن جدول احتمالات بیان شده است. ابتدا، ما باید شبکه عصبی و «مدل زبان[۱۱]» را بازیابی کنیم:

```
In[ ]:= model =
          NetModel[{"GPT2 Transformer Trained on WebText Data",
            "Task" → "LanguageModeling"}]

Out[ ]= NetChain[ 
                  Input port:  string
                  Output port: class ]
```

بعداً، به درون این شبکه عصبی نگاه خواهیم کرد و در مورد نحوه عملکرد آن صحبت خواهیم کرد. اما در حال حاضر فقط می‌توانیم این «مدل شبکه ای» را به‌عنوان یک جعبه سیاه روی متن خود اعمال کنیم و ۵ کلمه برتر را بر اساس احتمالی که مدل می‌گوید می بایست دنبال کنیم:

```
In[ ]:= model["The best thing about AI is its ability to", {"TopProbabilities", 5}]

Out[ ]= {do → 0.0288508, understand → 0.0307805,
          make → 0.0319072, predict → 0.0349748, learn → 0.0445305}
```

[۱۱] مدل زبانی نوعی مدل یادگیری ماشین (Machine Learning) است که برای ایجاد یک توزیع احتمال بر روی کلمات، آموزش (Train) داده می‌شود. به بیان ساده، این مدل سعی می‌کند با توجه به متن داده شده، کلمه مناسب بعدی را برای پر کردن یک فضای خالی در جمله یا عبارت، پیش‌بینی کند.

مدل این نتیجه را می‌گیرد و آن را به یک «مجموعه‌داده» با فرمت آشکار تبدیل می‌کند:

```
In[•]:= Dataset[ReverseSort[Association[%]],
          ItemDisplayFunction → (PercentForm[#, 2] &)]
```

Out[•]=

learn	4.5%
predict	3.5%
make	3.2%
understand	3.1%
do	2.9%

اینکه اگر کسی به طور مکرر «مدل را اعمال کند» چه اتفاقی می‌افتد در پایین نمایش داده شده است. در هر مرحله کلمه‌ای را که بیشترین احتمال را دارد اضافه می‌کند (که در این کد به‌عنوان «تصمیم[۱۲]» در مدل مشخص شده است):

```
In[•]:= NestList[StringJoin[#, model[#, "Decision"]] &,
          "The best thing about AI is its ability to", 7]
```

```
Out[•]= {The best thing about AI is its ability to,
  The best thing about AI is its ability to learn,
  The best thing about AI is its ability to learn from,
  The best thing about AI is its ability to learn from experience,
  The best thing about AI is its ability to learn from experience.,
  The best thing about AI is its ability to learn from experience. It,
  The best thing about AI is its ability to learn from experience. It's,
  The best thing about AI is its ability to learn from experience. It's not}
```

چه اتفاقی می‌افتد اگر یکی بیشتر ادامه دهد؟ در این مثال با دمای صفر درجه، آنچه بعد از آن می آید نسبتاً گیج کننده و تکرارشونده است.

[۱۲] Decision

The best thing about AI is its ability to learn from experience. It's not just a matter of learning from experience, it's learning from the world around you. The AI is a very good example of this. It's a very good example of how to use AI to improve your life. It's a very good example of how to use AI to improve your life. The AI is a very good example of how to use AI to improve your life. It's a very good example of how to use AI to

اما اگر به جای انتخاب همیشه کلمه «دارای رتبه بالا» گاهی اوقات به طور تصادفی کلمات «غیر رتبه دار» را انتخاب کنیم (با «تصادفی بودن» برابر با «دما» ۰.۸)؟ دوباره می توان متن را ساخت:

{The best thing about AI is its ability to,
The best thing about AI is its ability to create,
The best thing about AI is its ability to create worlds,
The best thing about AI is its ability to create worlds that,
The best thing about AI is its ability to create worlds that are,
The best thing about AI is its ability to create worlds that are both,
The best thing about AI is its ability to create worlds that are both exciting,
The best thing about AI is its ability to create worlds that are both exciting,}

و هر بار که کسی این کار را انجام می‌دهد، انتخاب‌های تصادفی متفاوتی ایجاد می‌شود و متن متفاوت خواهد بود. مانند این ۵ مثال:

The best thing about AI is its ability to learn. I've always liked the

The best thing about AI is its ability to really come into your world and just

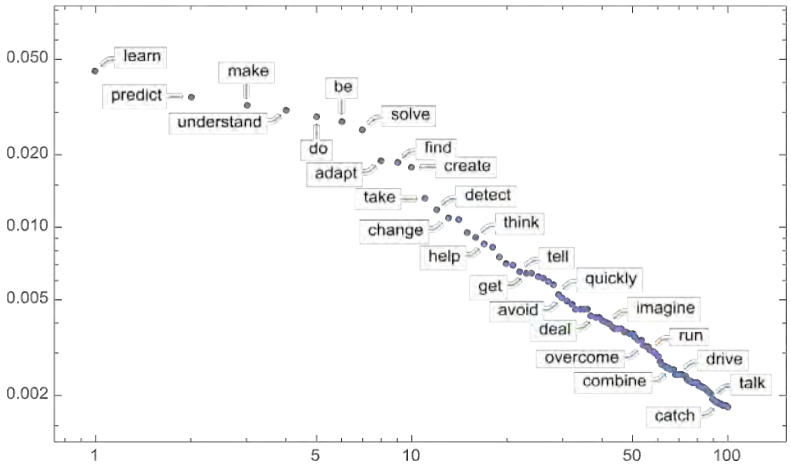
The best thing about AI is its ability to examine human behavior and the way it

The best thing about AI is its ability to do a great job of teaching us

The best thing about AI is its ability to create real tasks, but you can

شایان‌ذکر است که حتی در مرحله اول تعداد زیادی «کلمات بعدی» برای انتخاب وجود دارد (در دمای ۰.۸)، اگرچه احتمالات آنها به سرعت کاهش می یابد (بله خط مستقیم در این نمودار

log-log مطابق با «قانون توان[۱۳]» n^{-1} و نزولی است که از شاخصه های عمومی آماری زبان است:



پس چه اتفاقی می‌افتد اگر یکی بیشتر ادامه دهد؟ در اینجا یک مثال به‌صورت تصادفی ذکر می‌شود. این حالت بهتر از حالت استفاده از کلمات دارای رتبه بالا(دمای صفر) است، اما هنوز جای کار دارد:

The best thing about AI is its ability to see through, and make sense of, the world around us, rather than panicking and ignoring. This is known as AI "doing its job" or AI "run-of-the-mill." Indeed, taking an infinite number of steps, developing a machine that can be integrated with other systems, or controlling one system that's truly a machine, is one of the most fundamental processes of AI. Aside from the human-machine interaction, AI was also a big part of creativity

این کار با ساده‌ترین مدل جی‌پی‌تی۲ (از سال ۲۰۱۹) انجام پذیرفته است. با استفاده از مدل‌های جدیدتر و بزرگ‌تر جی‌پی‌تی۳ نتایج بهتری بدست می‌آید. در اینجا متن اصلی (دمای صفر) تولید شده به همان صورت قبل اما با بزرگترین مدل جی‌پی‌تی۳ است:

[۱۳] قانون توان رابطه ای است که در آن یک تغییر نسبی در یک کمیت، مستقل از اندازه اولیه آن کمیت ها، تغییر نسبی متناسبی را در کمیت دیگر ایجاد می کند.

The best thing about AI is its ability to automate processes and make decisions quickly and accurately. AI can be used to automate mundane tasks, such as data entry, and can also be used to make complex decisions, such as predicting customer behavior or analyzing large datasets. AI can also be used to improve customer service, as it can quickly and accurately respond to customer inquiries. AI can also be used to improve the accuracy of medical diagnoses and to automate the process of drug discovery.

و در اینجا یک مثال تصادفی دیگر در «دمای ۰.۸» آمده است:

The best thing about AI is its ability to learn and develop over time, allowing it to continually improve its performance and be more efficient at tasks. AI can also be used to automate mundane tasks, allowing humans to focus on more important tasks. AI can also be used to make decisions and provide insights that would otherwise be impossible for humans to figure out.

احتمالات از کجا می‌آیند؟

چت جی‌پی‌تی همیشه کلمه بعدی خود را بر اساس احتمالات انتخاب می‌کند. اما این احتمالات از کجا می‌آیند؟ بیایید با یک مشکل ساده‌تر شروع کنیم. بیایید تولید متن انگلیسی را هر بار با یک حرف (به‌جای کلمه) در نظر بگیریم. چگونه می‌توانیم بفهمیم که احتمال هر حرف چقدر باید باشد؟

حداقل کاری که می‌توانیم انجام دهیم این است که فقط یک نمونه از متن انگلیسی را برداریم و محاسبه کنیم که چند بار حروف مختلف در آن تکرار شده‌اند. برای مثال، این حروف در مقاله‌های ویکی‌پدیا برای «گربه‌ها» را حساب می‌کنیم:

```
In[ ]:= LetterCounts[WikipediaData["cats"]]

Out[ ]= ⟨| e → 4279, a → 3442, t → 3397, i → 2739, s → 2615, n → 2464, o → 2426,
 r → 2147, h → 1613, l → 1552, c → 1405, d → 1331, m → 989, u → 916,
 f → 760, g → 745, p → 651, y → 591, b → 511, w → 509, v → 395, k → 212,
 T → 114, x → 85, A → 81, C → 81, I → 68, S → 55, F → 42, z → 38, E → 36
```

و مشابه کار بالا را نیز برای «سگ‌ها» انجام می‌دهیم:

```
In[ ]:= LetterCounts[WikipediaData["dogs"]]

Out[ ]= ⟨| e → 3911, a → 2741, o → 2608, i → 2562, t → 2528, s → 2406,
  n → 2340, r → 1866, d → 1584, h → 1463, l → 1355, c → 1083, g → 929,
  m → 859, u → 782, f → 662, p → 636, y → 500, b → 462, w → 409,
  v → 406, k → 151, T → 90, C → 85, I → 80, A → 74, x → 71, S → 65,
```

نتایج مشابه هستند، اما یکسان نیستند («o» بدون شک در مقاله «سگ» بیشتر رایج است، زیرا در خود کلمه «سگ» نیز آمده است). بااین‌حال، اگر نمونه بزرگی از متن انگلیسی را در نظر بگیریم، می‌توان انتظار داشت که در نهایت حداقل نتایج نسبتاً ثابتی به دست آوریم:

```
In[ ]:= English LANGUAGE [ character frequencies ]

Out[ ]= {e → 12.7%, t → 9.06%, a → 8.17%, o → 7.51%, i → 6.97%, n → 6.75%,
  s → 6.33%, h → 6.09%, r → 5.99%, d → 4.25%, l → 4.03%, c → 2.78%, u → 2.76%,
  m → 2.41%, w → 2.36%, f → 2.23%, g → 2.02%, y → 1.97%, p → 1.93%, b → 1.49%,
  v → 0.978%, k → 0.772%, j → 0.153%, x → 0.150%, q → 0.0950%, z → 0.0740%}
```

اگر فقط دنباله‌ای از حروف را با این احتمالات تولید کنیم، نمونه‌ای از آنچه به دست می‌آوریم در اینجا آمده است:

```
rronoitadatcaeaesaotdoysaroiyiinnbantoioestlhddeocneooewceseciselnodrtrdgriscsatsepesdcnio
    uhoetsedeyhedslernevstothindtbmnaohngotannbthrdthtonsipieldn
```

می‌توانیم با اضافه‌کردن فاصله‌هایی که گویی حروفی با احتمال مشخص هستند، جمله را به «کلمات» زیر تقسیم کنیم:

```
sd n oeiaim satnwhoo eer rtr ofiianordrenapwokom del oaas ill e h f
    rellptohltvoettseodtrncilntehtotrkthrslo hdaol n sriaefr hthehtn ld gpod a  h y oi
```

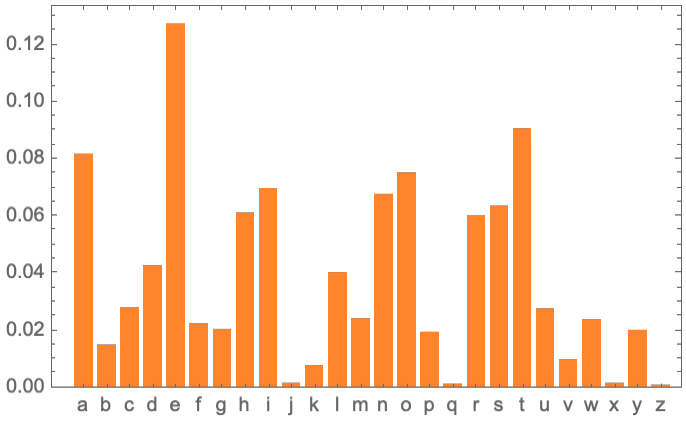
با پیاده‌کردن توزیع بر اساس «طول کلمات» در انگلیسی می‌توانیم ساختن «کلمات» را کمی بهتر نماییم:

ni hilwhuei kjtn isjd erogofnr n rwhwfao rcuw lis fahte uss cpnc
nlu oe nusaetat llfo oeme rrhrtn xdses ohm oa tne ebedcon oarvthv ist

ما در اینجا به «کلمات واقعی» نرسیدیم، اما نتایج کمی بهتر به نظر می‌رسند. بااین‌حال، برای ادامه‌دادن، باید کارهای بیشتری از انتخاب هر حرف جداگانه به‌صورت تصادفی انجام دهیم.

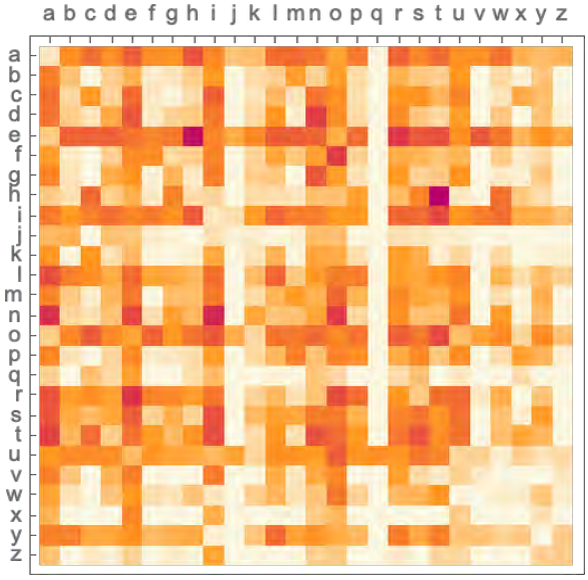
به‌عنوان‌مثال، می‌دانیم که اگر «q» داشته باشیم، حرف بعدی اساساً باید «u» باشد.

در اینجا نموداری از احتمالات حروف آورده شده است:



و در اینجا نموداری است که احتمال جفت حروف «۲ گرم»[۱۴] را در متن معمولی انگلیسی نشان می‌دهد. حروف اولِ ممکن به صورت افقی و حروف دوم به صورت عمودی در صفحه نشان داده شده اند:

[۱۴] ۲-grams



و در اینجا می‌بینیم که برای مثال، ستون «q» خالی است (احتمال صفر) به جز در ردیف «u». پس اکنون به‌جای اینکه کلمات خود را هر بار با یک حرف واحد تولید کنیم، با استفاده از این احتمالات «۲ گرمی» آنها را با نگاهی به دو حرف در یک زمان تولید میکنیم. در اینجا نمونه‌ای از نتیجه که شامل چند کلمه واقعی است، نمایش داده می‌شود:

on inguman men ise forerenoft weat iofobato buc ous corew ousesetiv
falle tinouco ryefo ra the ecederi pasuthrgr cuconom tra tesla wil tat pere thi

با استفاده از متن انگلیسی به‌اندازه کافی، می‌توانیم تخمین‌های بسیار خوبی را نه‌تنها برای احتمال تک حرف یا جفت حرف (۲ گرم)، بلکه برای حروف طولانی‌تر نیز به دست آوریم. و اگر «کلمات تصادفی» را با احتمالات n-gram[۱۵] به‌تدریج طولانی تر تولید کنیم، می‌بینیم که آنها به‌تدریج «واقع بینانه‌تر» می‌شوند:

[۱۵] n-gram مجموعه ای از n مورد متوالی در یک سند متنی است که ممکن است شامل کلمات، اعداد، نمادها و علائم نگارشی باشد. مدل‌های n-gram در بسیاری از برنامه‌های تحلیل متنی که توالی کلمات مرتبط هستند، مانند تحلیل احساسات، طبقه‌بندی متن و تولید متن، مفید هستند.

0	on gxeeetowmt tsifhy ah aufnsoc ior oia itlt bnc tu ih uls
1	ri io os ot timumumoi gymyestit ate bshe abol viowr wotybeat mecho
2	wore hi usinallistin hia ale warou pothe of premetra bect upo pr
3	qual musin was witherins wil por vie surgedygua was suchinguary outheydays theresist
4	stud made yello adenced through theirs from cent intous wherefo proteined screa
5	special average vocab consumer market prepara injury trade consa usually speci utility

بیایید فرض کنیم که شبیه چت جی‌پی‌تی این کار را انجام دهیم که با کلمات کامل سروکار داشته باشیم نه با حروف. در زبان انگلیسی حدود ۴۰۰۰۰ کلمه پرکاربرد وجود دارد و با نگاه‌کردن به مجموعه بزرگی از متن انگلیسی (مثلاً چند میلیون کتاب، در مجموع با چند صد میلیارد کلمه) می‌توانیم تخمینی از رایج بودن هر کلمه به دست آوریم و با استفاده از آن می‌توانیم شروع به تولید «جملات» کنیم که در آن هر کلمه به طور مستقل و به طور تصادفی انتخاب می‌شود، با همان احتمالی که در مجموعه ظاهر می‌شود. در اینجا نمونه‌ای از آنچه به دست می‌آوریم، نمایش داده شده است:

of program excessive been by was research rate not here of of other is men
were against are show they the different the half the the in any were leaved

جای تعجب ندارد که این کار هنوز کامل نیست. پس چگونه می‌توانیم بهتر عمل کنیم؟ درست مانند حروف، برای این کار می‌توانیم نه‌تنها احتمالات را برای تک‌تک کلمه‌ها، بلکه برای جفت‌ها (n گرم) کلمات نیز در نظر بگیریم. با انجام این کار برای جفت کلمه‌ها، ۵ نمونه از آنچه به دست می‌آوریم، را نمایش می‌دهیم.

در همه موارد از کلمه «گربه» شروع می‌شود:

cat through shipping variety is made the aid emergency can the

cat for the book flip was generally decided to design of

cat at safety to contain the vicinity coupled between electric public

cat throughout in a confirmation procedure and two were difficult music

cat on the theory an already from a representation before a

کمی جملات «معقولانه‌تر» می‌شوند. در اینجا امکان دارد تصور کنیم که اگر از n-gram های به‌اندازه کافی طولانی استفاده کنیم، اساساً یک چت جی‌پی‌تی خواهیم داشت. این بدین معناست که خروجی ما به طول مقاله، توالی‌هایی از کلمات را با «احتمالات کلی مقاله صحیح» ایجاد می‌کند؛ اما مشکل اینجا است: به‌اندازه کافی متن انگلیسی وجود ندارد که بتوان آن احتمالات را استنتاج کرد.

در برداشت داده از محیط وب ممکن است چند صد میلیارد کلمه یافت شود. در کتاب‌هایی که دیجیتالی شده‌اند ممکن است صد میلیارد کلمه دیگر وجود داشته باشد. اما با ۴۰۰۰۰ کلمه رایج، حتی با جمع تعداد ۲ گرم ها نیز ۱.۶ میلیارد کلمه وجود دارد و با جمع تعداد ۳ گرم ها به چیزی در حدود ۶۰ تریلیون کلمه خواهیم رسید. بنابراین هیچ راهی وجود ندارد که بتوانیم احتمالات را حتی برای همه اینها از روی متون موجود تخمین بزنیم. زمانی که به «قطعات مقاله» ۲۰ کلمه‌ای می رسیم، تعداد احتمالات از تعداد ذرات جهان بیشتر است، بنابراین به یک معنا هرگز نمی توان همه آنها را نوشت.

پس چه کاری می‌توانیم انجام دهیم؟ ایده بزرگ این است که مدلی بسازیم که به ما امکان دهد احتمالات توالی‌ها را تخمین بزنیم، حتی اگر هرگز آن توالی‌ها را به‌صراحت در مجموعه متنی که بررسی کرده‌ایم نباشد. در هسته چت جی‌پی‌تی دقیقاً یک مدل به‌اصطلاح زبان بزرگ[۱۶]وجود دارد که برای تخمین این احتمالات ساخته شده است.

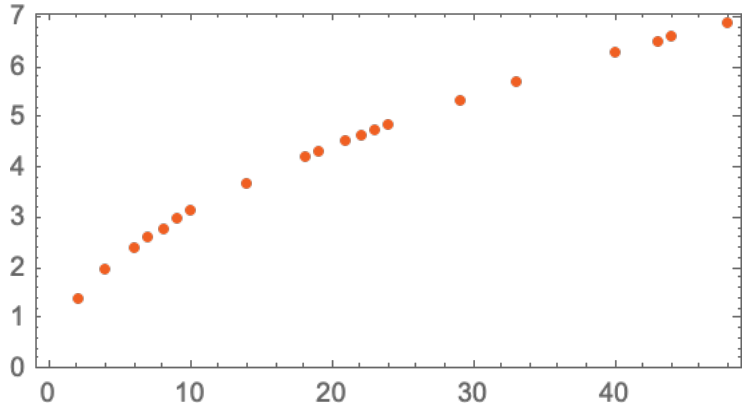
مدل چیست؟

می‌خواهید بدانید (همان‌طور که گالیله در اواخر دهه ۱۵۰۰ انجام داد) چه مدت طول می‌کشد تا یک گلوله توپی که از هر طبقه از برج پیتزا پرتاب می‌شود به زمین بخورد؟

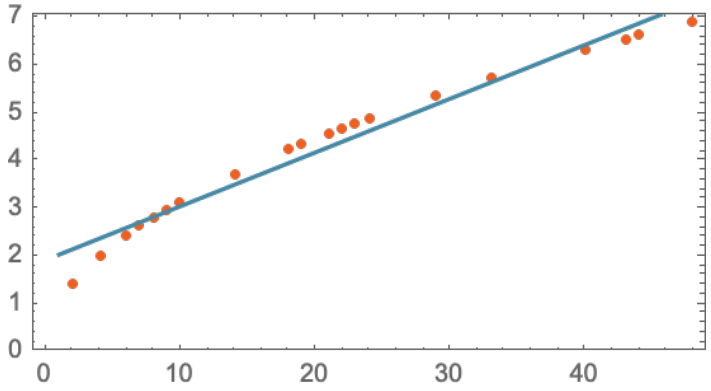
خوب، شما فقط می‌توانید آن را در مواردی اندازه‌گیری و یک جدول از نتایج آن تهیه نمایید یا می‌توانید آنچه را که جوهره علم نظری است انجام دهید: مدلی بسازید که به‌جای اندازه‌گیری و به‌خاطر سپردن هر مورد، نوعی روش برای محاسبه پاسخ ارائه دهد.

[۱۶] LLM

بیایید تصور کنیم داده‌هایی (که تا حدودی ایده‌آل شده‌اند) برای مدت زمانی که توپ از طبقات مختلف سقوط می‌کند را دارا هستیم:



چگونه بفهمیم چقدر طول می‌کشد تا از طبقه‌ای بیفتیم که صراحتاً اطلاعاتی درباره آن نداریم؟ در این مورد خاص، ما می‌توانیم از قوانین شناخته شده فیزیک برای حل آن استفاده کنیم. اما اگر بگوییم فقط داده‌ها را داریم و نمی‌دانیم چه قوانینی بر آن حاکم است، احتمال دارد به سراغ یک حدس ریاضی برویم، مانند این که شاید باید از یک خط مستقیم به‌عنوان مدل استفاده کنیم:

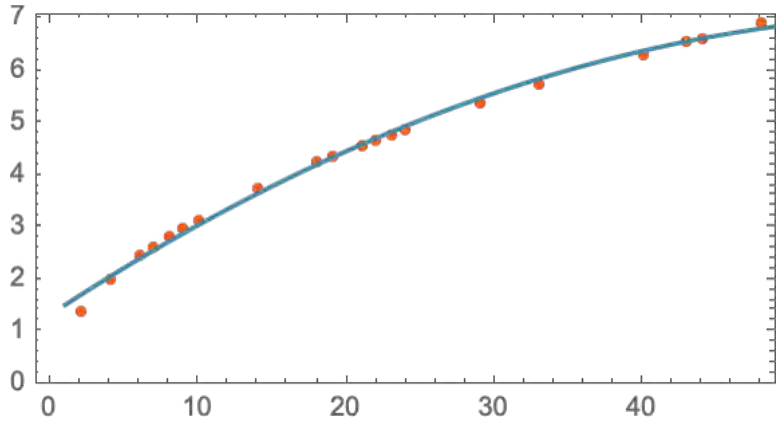


ما می‌توانستیم خطوط مستقیم مختلفی را انتخاب کنیم. اما این نمونه به داده‌هایی که ما داریم نزدیک تر است. از این خط مستقیم می‌توانیم زمان سقوط را برای هر طبقه تخمین بزنیم.

از کجا فهمیدیم در اینجا باید از خط مستقیم استفاده کنیم؟ در واقع نمی‌دانستیم. بخاطر این بوده که از نظر ریاضی [خط مستقیم] بسیار ساده است و ما به این واقعیت عادت کرده‌ایم که بسیاری از داده‌هایی که اندازه‌گیری می‌کنیم به خوبی با موارد ساده ریاضی مطابقت دارند.

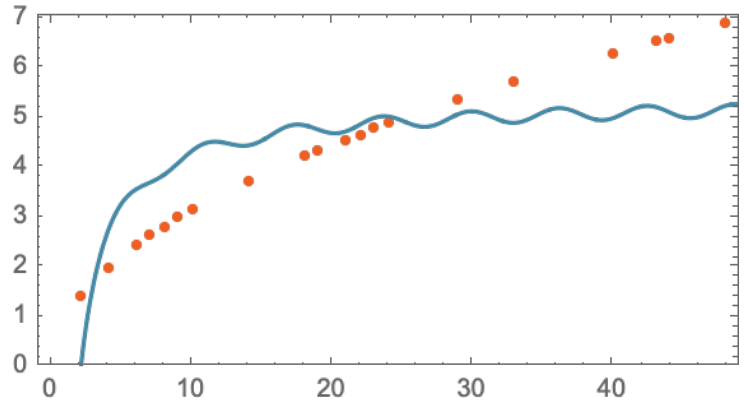
ما می‌توانیم موارد ریاضیاتی پیچیده‌تری را نیز امتحان کنیم تا به نتیجه بهتری برسیم:

$a + b\,x + c\,x^2$



بااین‌حال، همه چیز ممکن است کاملاً اشتباه باشد. مانند بهترین کاری که می‌توانیم با

a + b / x + c sin(x) انجام دهیم:



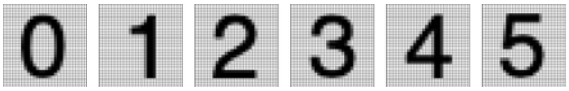
قابل‌درک است که هرگز «مدلی بدون شکل» وجود ندارد. هر مدلی که شما استفاده کنید ساختار زیربنایی خاصی دارد. سپس مجموعه خاصی از پارامترهایی که می‌توانید تنظیم کنید در اختیار قرار می‌دهد تا با داده‌های شما مطابقت داشته باشد. در مورد چت جی‌پی‌تی، تعداد زیادی از این پارامترها یعنی ۱۷۵ میلیارد از آنها استفاده می‌شود

اما نکته قابل‌توجه این است که ساختار زیربنایی چت جی‌پی‌تی با «فقط» انبوهی از این پارامترها برای ایجاد مدلی که احتمالات کلمه بعدی را محاسبه کند و قطعات متنی با طول مقاله معقول را به ما ارائه دهد، کافی است.

مدل‌هایی برای کارهای شِبهِ انسانی

مثالی که در بالا آوردیم شامل ساختن مدلی برای داده‌های عددی است که اساساً از فیزیک ساده نشات می‌گیرد. جایی که برای چندین قرن می‌دانستیم ریاضیات ساده کاربرد دارد. اما برای چت جی‌پی‌تی باید از مدل متنی به زبان انسان استفاده نماییم و برای چنین چیزی هنوز مانند «ریاضیات ساده» نداریم؛ بنابراین، یک مدل از آن چگونه باید باشد؟

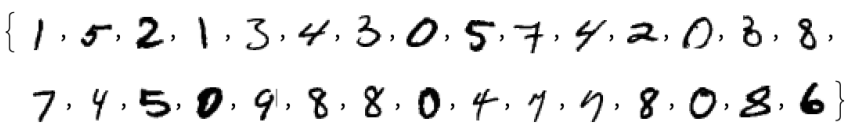
قبل از اینکه در مورد زبان صحبت کنیم، اجازه دهید در مورد یکی دیگر از کارهای شبه انسانی صحبت کنیم: تشخیص تصاویر. به‌عنوان یک مثال ساده، بیایید تصاویر ارقام را در نظر بگیریم (این یک مثال کلاسیک یادگیری ماشین است)



یکی از کارهایی که می‌توانیم انجام دهیم این است که برای هر رقم یک دسته عکس نمونه بگیریم:

سپس برای اینکه بفهمیم آیا تصویری که به‌عنوان ورودی داده می‌شود با یک رقم خاص مطابقت دارد یا نه می‌توانیم یک مقایسه واضح پیکسل به پیکسل با نمونه‌هایی که داریم انجام دهیم. اما به‌عنوان انسان مطمئناً به نظر می‌رسد که این کار را به نحو بهتری انجام می‌دهیم، زیرا هنوز

می‌توانیم ارقام را تشخیص دهیم، حتی زمانی که مثلاً دست‌نویس هستند، و انواع تغییرات و تحریف‌ها را دارند:



هنگامی که ما یک مدل را برای داده‌های عددی در بالا ساختیم، توانستیم یک مقدار عددی x را که به‌عنوان ورودی به ما داده شده بود را بگیریم و فقط a + bx را برای a و b مشخص شده محاسبه کنیم؛ بنابراین، اگر مقدار سطح خاکستری هر پیکسل را در اینجا به‌عنوان یک متغیر x_i در نظر بگیریم، آیا تابعی از همه آن متغیرها وجود دارد که هنگام ارزیابی به ما بگوید تصویر چه رقمی را نشان می‌دهد؟ به نظر می‌رسد که امکان ساخت چنین تابعی وجود دارد. بااین‌حال جای تعجب نیست که این کاری دشوار است. یک مثال معمولی ممکن است شامل نیم‌میلیون عملیات ریاضی باشد.

اما نتیجه نهایی این است که اگر مجموعه مقادیر پیکسل برای یک تصویر را به این تابع وارد کنیم، عددی را که خروجی می دهد مشخص می‌کند تصویر کدام رقم ها را داریم. بعداً در مورد چگونگی ساخت چنین تابعی و ایده شبکه‌های عصبی صحبت خواهیم کرد؛ اما در حال حاضر اجازه دهید این تابع را به‌عنوان جعبه سیاه در نظر بگیریم که در آن تصاویر مثلاً ارقام دست‌نویس (به‌عنوان آرایه‌هایی از مقادیر پیکسل) را به‌عنوان منبع در نظر بگیریم و اعداد مطابق آنها را دریافت کنیم:

```
In[•]:= NetModel[ ][{ }]
Out[•]= {7, 0, 9, 7, 8, 2, 4, 1, 1, 1}
```

اما واقعاً اینجا چه خبر است؟ فرض کنید یک رقم را به‌تدریج محو می‌کنیم. برای مدت کوتاهی تابع ما هنوز آن را به‌عنوان عدد ۲ «تشخیص می‌دهد» اما زود [توانایی شناسایی] آن را از دست می‌دهد و شروع به دادن خروجی «اشتباه» می‌کند:

```
In[•]:= NetModel[ ][{ }]
Out[•]= {2, 2, 2, 1, 1, 1, 1, 1, 1}
```

اما چرا می‌گوییم که نتیجه «اشتباه» است؟ در این مورد، ما می‌دانیم که همه تصاویر را با محوکردن عدد «۲» دریافت کرده‌ایم. اما اگر هدف ما تولید مدلی شبیه تشخیص تصاویر توسط انسان باشد، درحقیقت این سؤال مطرح می‌گردد که اگر انسان با یکی از تصاویر تار مواجه شود، بدون اینکه بداند از کجا آمده است چه می‌کند؟

اگر نتایجی که از عملکردمان به دست می‌آوریم معمولاً با آنچه یک انسان می‌گوید مطابقت داشته باشد، یک «مدل خوب» داریم. واقعیت علمی قابل ملاحظه این است که برای یک کار تشخیص تصویر مانند این، اساساً می‌دانیم که چگونه توابعی را بسازیم که این کار را انجام دهند.

آیا می‌توانیم از نظر ریاضی ثابت کنیم که آن توابع کار می‌کنند؟ خب نه. زیرا برای انجام این کار باید یک نظریه ریاضی در مورد آنچه انسان‌ها انجام می‌دهند داشته باشیم. تصویر «۲» را بگیرید و چند پیکسل را تغییر دهید. ممکن است تصور کنیم تنها با وجود چند پیکسل «در غیرجای خود» همچنان باید تصویر را عدد «۲» در نظر بگیریم؛ اما این روند تا کجا پیش خواهد رفت؟ مسئله اصلی ادراک بصری انسان است. بله، بدون شک این پاسخ برای زنبورها یا اختاپوس‌ها متفاوت خواهد بود و همچنین به طور بالقوه برای موجودات فضایی فرضی نیز کاملاً متفاوت خواهد بود.

شبکه‌های عصبی

با این تفاسیر، مدل‌های معمولی ما برای کارهایی مانند تشخیص تصاویر چگونه کار می‌کنند؟ محبوب‌ترین و موفق‌ترین رویکرد فعلی، از شبکه‌های عصبی استفاده می‌کند. شبکه عصبی که می‌توان آن را تصور ساده‌ای از نحوه عملکرد مغزها در نظر گرفت، در دهه ۱۹۴۰، به شکلی بسیارنزدیک به کاربرد امروزی اختراع شد.

در مغز انسان حدود ۱۰۰ میلیارد نورون (سلول عصبی) وجود دارد که هر کدام قادر به تولید یک پالس الکتریکی است (شاید هزار بار در ثانیه). نورون‌ها در یک شبکه پیچیده به هم متصل هستند و هر نورون دارای شاخه های درخت‌مانند است که به آن اجازه می‌دهد سیگنال‌های الکتریکی را به هزاران نورون دیگر منتقل کند. در یک محاسبه تقریبی، اینکه آیا هر نورون معینی در یک لحظه مشخص یک پالس الکتریکی تولید می‌کند، بستگی به این دارد که چه پالس‌هایی از نورون‌های دیگر دریافت می‌کند. (اتصالات بین نورون‌های مختلف دارای وزن‌های مختلفی می باشند)

هنگامی که ما «یک تصویر» را می‌بینیم، آنچه اتفاق می‌افتد این است که وقتی فوتون‌های نور از تصویر بر روی سلول‌های «گیرنده نوری» در پشت چشم ما می‌افتند، سیگنال‌های الکتریکی در سلول‌های عصبی تولید می‌کنند. این سلول‌های عصبی به سلول‌های عصبی دیگر متصل هستند و در نهایت سیگنال‌ها از طریق یک توالی کامل از لایه‌های عصبی عبور می‌کنند. در این فرایند است که ما تصویر را تشخیص می‌دهیم و در نهایت «فکری را شکل می‌دهیم» که عدد «۲» را می‌بینیم (و شاید در پایان کاری مانند گفتن کلمه «دو» با صدای بلند انجام دهیم).

تابع «جعبه سیاه» از بخش قبل، یک نسخه «ریاضی شده» از چنین شبکه عصبی است. اتفاقاً ۱۱ لایه دارد البته فقط ۴ «لایه اصلی»:

NetChain

		image
	Input	array (size: 1×28×28)
1	ConvolutionLayer	array (size: 20×24×24)
2	Ramp	array (size: 20×24×24)
3	PoolingLayer	array (size: 20×12×12)
4	ConvolutionLayer	array (size: 50×8×8)
5	Ramp	array (size: 50×8×8)
6	PoolingLayer	array (size: 50×4×4)
7	FlattenLayer	vector (size: 800)
8	LinearLayer	vector (size: 500)
9	Ramp	vector (size: 500)
10	LinearLayer	vector (size: 10)
11	SoftmaxLayer	vector (size: 10)
	Output	class

هیچ‌چیز خاصی در مورد این شبکه عصبی «از لحاظ نظری مشتق شده» وجود ندارد. این فقط چیزی است که در سال ۱۹۹۸ به‌عنوان یک فن مهندسی ساخته شده و بکار گرفته شده است. (البته، تفاوت چندانی با نحوه توصیف مغز ما که از طریق فرایند تکامل بیولوژیکی تولید شده است، ندارد.)