

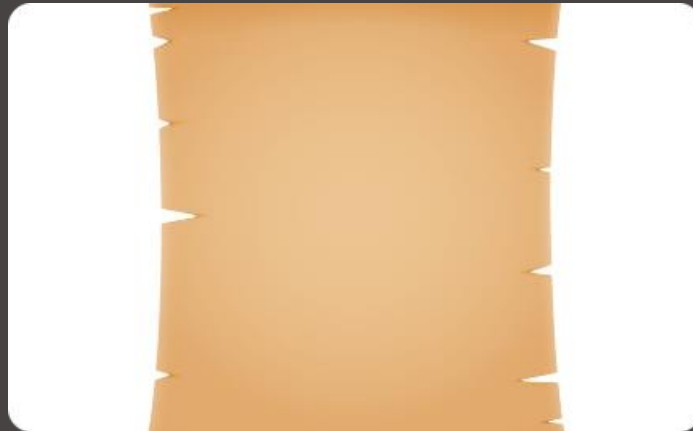
طراحی مدل پرسش و پاسخ به زبان فارسی با بازتنظیم مدل زبانی بزرگ

در این پژوهش ما به بررسی مدل‌های زبانی بزرگ، روش‌های بازتنظیم آن، مدل‌های پرسش و پاسخ، مشکلات آن برای زبان فارسی و نهایتاً به بحث در مورد کارهای انجام شده و مدل پرسش و پاسخ فارسی ساخته می‌پردازیم.



علی باقرزاده

مدل زبانی بزرگ



مدل زبانی بزرگ

مدل‌های زبانی بزرگ، شبکه‌های عصبی پیشرفته با میلیاردها پارامتر هستند و می‌توانند وظایف متنی چون ترجمه، تولید متن خلاقانه و پاسخ به سوالات را با دقت بالا انجام دهند



مدل های زبانی بزرگ چندزبانه

مدل های چندزبانه مدل‌هایی هستند که چند زبان تمرین داده شده‌اند اغلب کاربرد این مدل‌ها برای زبان‌هایی غیر از زبان انگلیسی می‌باشد.

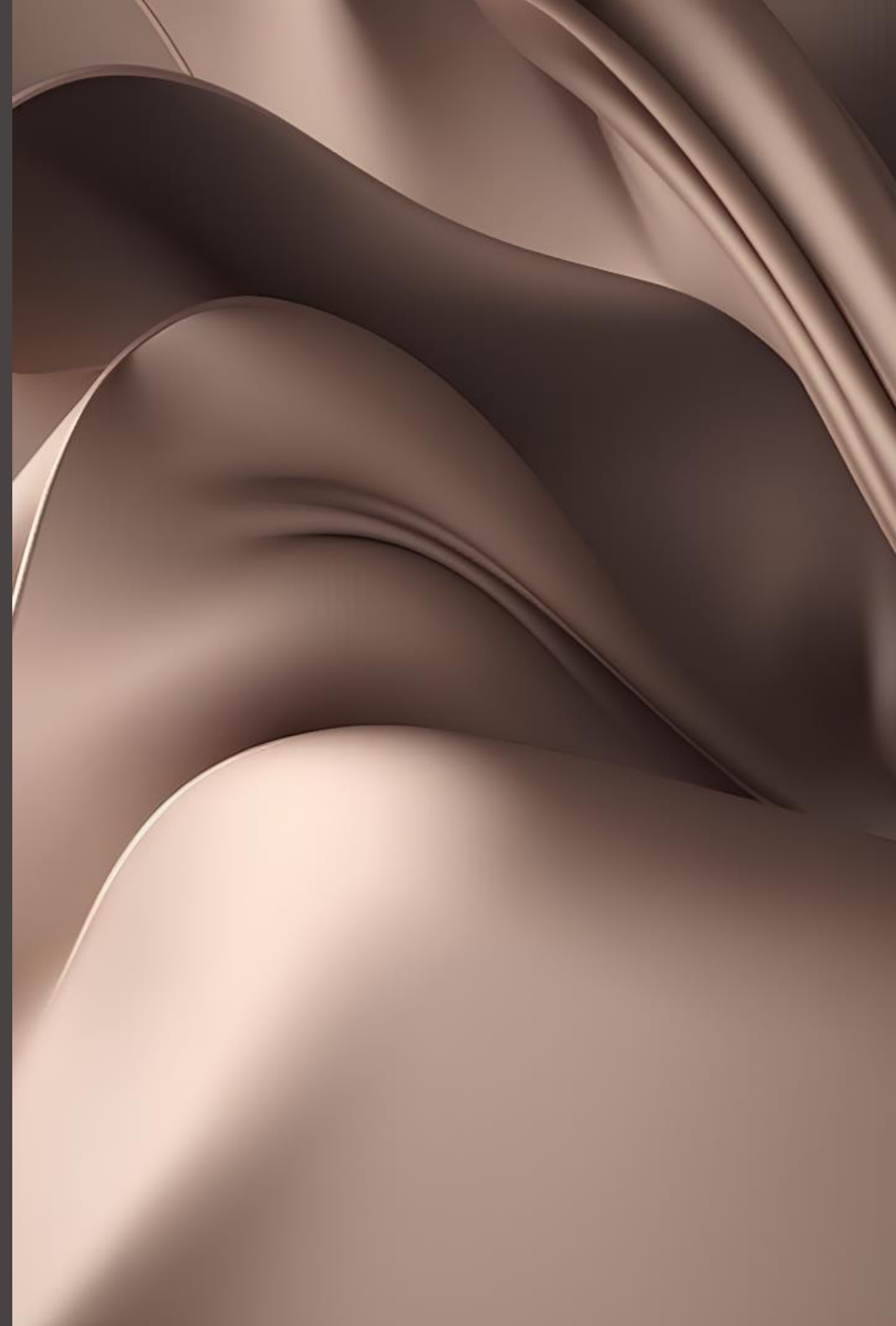


باز تنظیم مدل زبانی

برای رسیدن به مدل‌هایی با زبان غیر از انگلیسی ما به جای اختراع مجدد چرخ به بازتنظیم مدل‌های موجود برای زبان دیگر می‌پردازیم.

دلیل نیاز به بازتنظیم مدل

- محدودیت های مدل زبانی بزرگ در زبان فارسی
- عدم تطابق با ساختار زبان فارسی
- مشکلات در پردازش فعل و ترکیب کلمات
- مشکلات در تشخیص اسامی و پیوستگی افعال



مراحل باز تنظیم مدل زبانی بزرگ

1

تجزیه و تحلیل دادگان

انتخاب مناسب‌ترین روش‌های کاوش داده و انتخاب مجموعه دادگان بزرگ و مفید و متنی فارسی.

2

پیکربندی مدل

تعریف و پیکربندی دوباره مدل زبانی بزرگ برای مرورتدثر و مناسب سازی نمونه ها برای تمرین مدل.

3

آموزش مدل

آموزش درست و مناسب مدل با استفاده از الگوریتم های مناسب و داده های پراکنده و متنوع.

انواع روش‌های بازتنظیم مدل‌های زبانی بزرگ

تنظیم دستور مدل

در تنظیم دستور ما مدل را با دادن دستورات و خروجی مناسب برای آن دستور به آن تنظیم می‌کنیم مانند وظیفه ترجمه

تنظیم پرسش مدل

تنظیم پرسش به معنای دادن دستورات و prompt های جدید به یک مدل از پیش آموزش داده شده برای بهبود آن در عملکردی خاص می باشد

تنظیم نهایی مدل

تنظیم نهایی به معنای تمرین دادن لایه‌ای یا تمام مدل روی مجموعه دادگانی جدید و به منظور انجام وظیفه‌ای جدید می باشد.



مدل های پرسش و پاسخ

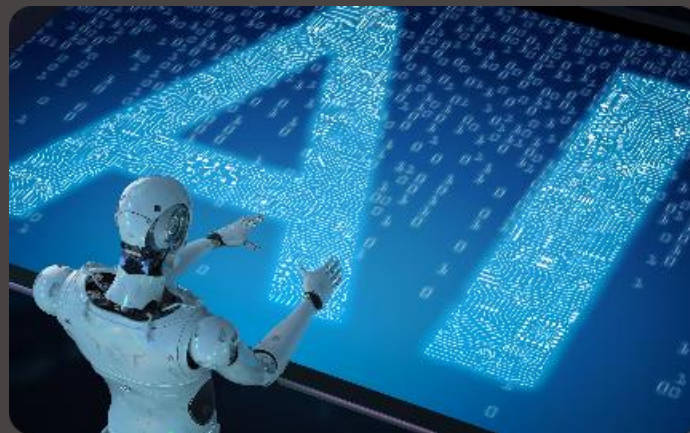
مدل های پرسش و پاسخ مدل های مبتنی بر یادگیری عمیق می باشند که می توانند به سوالات با توجه به متن داده شده و یا حتی بدون دادن متن پاسخ بدهند. برخی از این مدل ها پاسخ را از درون متن استخراج کرده و برخی دیگر پاسخ را از میان گزینه ها انتخاب می کنند.

انواع مدل های پرسش و پاسخ



تولید اجباری

این مدل ها پاسخ را به صورت کاملاً ابداعی و بدون توجه به متن انجام می دهند.



تولید آزادانه

این مدل ها پاسخ را با توجه به متن داده شده و با بیان خود تولید می کنند.



استخراجی

این نوع مدل ها پاسخ را به صورت کامل از متن داده شده استخراج می کنند

چالش های مدل پرسش و پاسخ به زبان فارسی

- عدم وجود داشتن مجموعه داده مناسب و کامل برای زبان فارسی
- پیچیدگی بالای زبان فارسی در دستور و قواعد



ساخت مجموعه دادگان به زبان فارسی

- ترجمه مجموعه های لاتین به کمک ماشین های ترجمه
- به دست آوردن جواب مناسب از متن ترجمه شده
- به دست آوردن ایندکس شروع برای جواب متن ترجمه شده

<div>{ "text": ["در شرق شهر مبارکه"], "answer_start": [114] }</div>	<div>"شرکت فولاد مبارکه در کجا واقع شده است"</div>	<div>"شرکت فولاد مبارکه اصفهان، بزرگترین واحد صنعتی خصوصی در ایران و بزرگترین مجتمع تولید فولاد در کشور ایران است، که در شرق شهر مبارکه قرار دارد. فولاد مبارکه هم اکنون محرک بسیاری از صنایع پالایشی و پتروشیمی است. فولاد مبارکه در ۱۱ دوره جایزه ملی تعالی سازمانی و ۶ دوره جایزه شرکت دانشی در کشور رتبه نخست را بدست آورده است و همچنین این شرکت در سال ۱۳۹۱ برای نخستینبار به عنوان تنها شرکت ایرانی با کسب امتیاز ۴۵۴ تندیس زرین جایزه ملی تعالی سازمانی را از آن خود کند. شرکت فولاد مبارکه اصفهان در ۲۴ دی ماه ۱۳۷۱ احداث شد و اکنون بزرگترین واحدهای صنعتی و بزرگترین مجتمع تولید فولاد در ایران است. این شرکت در زمینی به مساحت ۴۵ کیلومتر مربع در نزدیکی شهر مبارکه و در ۷۵ کیلومتری جنوب غربی شهر اصفهان واقع شده است. مصرف آب این کارخانه در کمترین میزان خود، ۱۸٪ از دبی زاینده‌رود برآورد سالانه ۲۴ میلیون متر مکعب در سال است و خود یکی از عوامل کم‌آبی زاینده‌رود شناخته میشود."</div>	<div>"شرکت فولاد مبارکه اصفهان"</div>	<div>1</div>
<div>{ "text": ["f"], "answer_start": [263] }</div>	<div>"فولاد مبارکه چند بار برنده جایزه شرکت دانشی را کسب کرده است؟"</div>	<div>"شرکت فولاد مبارکه اصفهان، بزرگترین واحد صنعتی خصوصی در ایران و بزرگترین مجتمع تولید فولاد در کشور ایران"</div>	<div>"شرکت فولاد مبارکه اصفهان"</div>	<div>2</div>
<div>{ "text": ["تندیس زرین جایزه ملی تعالی سازمانی"], "answer_start": [413] }</div>	<div>"شرکت فولاد مبارکه در سال ۱۳۹۱ چه جایزه ای برده؟"</div>	<div>"شرکت فولاد مبارکه اصفهان، بزرگترین واحد صنعتی خصوصی در ایران و بزرگترین مجتمع تولید فولاد در کشور ایران"</div>	<div>"شرکت فولاد مبارکه اصفهان"</div>	<div>3</div>

```
{
  "data": [
    {
      "paragraphs": [
        {
          "context": "The Normans (Norman: Nourmands; French: Normands; Latin: Normanni) were the people who in the 10th and 11th centuries gave their name to Normandy, a region in France. ",
          "qas": [
            {
              "answers": [
                {
                  "answer_start": 159,
                  "text": "France"
                }
              ],
              "id": "56dde6b9a695914005b9628",
              "is_impossible": false,
              "question": "In what country is Normandy located?"
            }
          ]
        }
      ],
      "title": "Normans"
    }
  ],
  "version": 2
}
```



استفاده از مدل زبانی ساخته شده

نتیجه گیری

مدل های پرسش و پاسخ اکنون به شدت نوپا هستند و در صورت وجود منابع مناسب برای زبان فارسی در این راستا می توان از این مدل ها برای یاد گرفتن مطالب جدید و دریافت پاسخ سوالات در مورد متون تازه و ناشناخته اقدام کرد و برای زبان فارسی نیز همانند زبان انگلیسی بر لبه فناوری قدم گذاشت.