

گزارش نشست علمی

«دستاوردهای جدید آزمایشگاه هوش مصنوعی نور»



اشاره

در بخش جنبی اولین همایش علوم اسلامی و انسانی دیجیتال، نشست علمی با ارائه جناب مهندس علی دانش، مدیر اداره پردازش هوشمند نور، به منظور معرفی دستاوردهای آزمایشگاه هوش مصنوعی نور برگزار شد. در پایان این جلسه، به سؤالات حاضران پاسخ داده شد.

آنچه در این نوشتار می‌آید، گزارشی است اجمالی از این نشست علمی که امید است برای خوانندگان عزیز مفید باشد. این گزارش در چهار بخش کلی: «معرفی مرکز نور»، «پیشینه محصولات فناورانه نور»، «تازه‌ترین دستاوردها» و «چشم‌اندازها و فعالیت‌های آتی نور در حوزه هوش مصنوعی» تهیه و تنظیم شده است.



معرفی مرکز نور

«همان‌طور که می‌دانید، این مرکز در سال ۱۳۶۸ش با فرمان مقام معظم رهبری تأسیس شد و در واقع، هدف ما این بود که از همان ابتدا در حوزه علوم اسلامی فعالیت داشته باشیم و متون این عرصه را پردازش کنیم تا در مسیر افزایش سرعت در کارهای پژوهشی و دسترسی به منابع علوم اسلامی هموار باشد. در گذشته، بابت دسترسی به منابع فیزیکی، زمان و هزینه زیادی باید صرف می‌شد که با استفاده از فناوری اطلاعات، این کار خیلی تسهیل شد و هزینه‌ها کاهش یافت.

در همان ایامی که هنوز کامپیوتر خیلی رواج پیدا نکرده بود، یک‌سری سمینارهای ملی به همت این مرکز برگزار شد که در راستای آشنایی بیشتر با کامپیوتر، فرهنگ استفاده از آن در حیطه علوم اسلامی و نیز فرهنگ‌سازی در این زمینه بود. بدیهی است که مخاطبان باید با تکنولوژی‌ها و محصولات نور آشنا می‌شدند. مستحضر هستید که نخستین نرم‌افزارهای دستک‌تاپی که تولید شد، با سیستم‌عامل داس کار می‌کرد و بعداً سیستم‌عامل ویندوز آمد و ما هم متناسب با فناوری‌های جدید، محصولات نور را به‌روزرسانی می‌کردیم و بحمدالله، برنامه‌های خوب و معروفی تولید شدند که بیشتر کاربران و محققان با آنها آشنا هستند و اساساً گاهی مرکز را به‌واسطه همین تولیدات می‌شناسند؛ مثل نرم‌افزار جامع التفاسیر و جامع الأحادیث نور.

با فعال شدن بحث اینترنت، مرکز نور به تولید محصولات وبی هم ورود پیدا کرد؛ مانند پایگاه‌های مختلفی که همواره به‌روز شده است؛ از جمله می‌توانم به نورمگز و نور لایب اشاره کنم و همین‌طور پایگاه قرآن، احادیث و قاموس که مخاطبان بسیاری دارد.

نکته‌ای که می‌خواستم تأکید کنم، این است که ما قابلیت‌های مربوط به هوش مصنوعی را در اکثر محصولات نور استفاده کرده‌ایم؛ البته شاید کاربر خیلی به هنگام استفاده از آن، متوجه این موضوع نشود. اساساً بحث ورود اطلاعات و فناوری اطلاعات و پردازش‌های مربوط به آنها، با کمک هوش مصنوعی انجام شده است. در همین راستا، مرکز نور در زمینه تولید و توسعه نرم‌افزار تلفن همراه نیز همسو با فناوری‌های جدید گام برداشته است.»



ما قابلیت‌های مربوط به هوش مصنوعی را در اکثر محصولات نور استفاده کرده‌ایم؛ البته شاید کاربر خیلی به هنگام استفاده از آن، متوجه این موضوع نشود. اساساً بحث ورود اطلاعات و فناوری اطلاعات و پردازش‌های مربوط به آنها، با کمک هوش مصنوعی انجام شده است. در همین راستا، مرکز نور در زمینه تولید و توسعه نرم‌افزار تلفن همراه نیز همسو با فناوری‌های جدید گام برداشته است



پیشینه دستاوردهای نور در حوزه هوش مصنوعی

«در این چند سال، با آمدن مدل‌های بزرگ زبانی، ابزارهای گفت‌وگو یا چت‌بات‌ها در فرهنگ عمومی هوش مصنوعی خیلی پُررنگ شده است. مرکز نور حدود پانزده سال پیش به عرصه هوش مصنوعی ورود پیدا کرد و اولین سمینار هوش مصنوعی علوم اسلامی را در سال ۱۳۹۰ برگزار نمود. طی این سال‌ها، تکنولوژی‌های مرتبط با هوش مصنوعی را در محصولات خودمان استفاده کردیم که در سه سطح قابل توضیح است.

سطح اول، مربوط به بحث ورود اطلاعات است؛ یعنی دیجیتالی کردن متون و تایپ اطلاعات. در این بخش، مباحثی مثل ا.و.سی.آر مطرح است و یا اشکالاتی که در زمینه تایپ وجود دارد و نحوه برطرف کردن آنها، در سطح ورود اطلاعات بحث و بررسی می‌شود.

سطح بعدی، بیانگر بحث فرآوری اطلاعات است. در این مرحله، به‌طور هوشمند روی متون برچسب‌گذاری‌ها یا تگ‌گذاری‌هایی انجام می‌شود. سطح سوم هم به مقوله شیوه ارائه اطلاعات می‌پردازد. قابلیت‌هایی همچون گروه‌بندی و مشابهت‌یابی متون، در این مرحله وجود دارد؛ مانند مشابهت‌یابی احادیث که در نرم‌افزار جامع الأحادیث تعبیه شده است.

ما در حوزه فناوری ا.و.سی.آر در این چند سال اخیر، با آمدن مدل‌های زبانی بزرگ، تحول خیلی خوبی داشتیم و کیفیت آن هم ارتقای قابل‌قبولی پیدا کرده است. سیستم‌های قبلی ما، مبتنی بر شبکه عصبی عمیق نبودند و بعد از آنکه شبکه‌های عصبی عمیق آمدند، یک بهبودی چشمگیری در این حوزه حاصل شد.

یکی از مباحث در روند تولیدات نور، بحث تبدیل اطلاعات است. در گذشته، زبان فارسی در کامپیوتر خیلی راحت قابل استفاده نبود. برای همین منظور، مرکز نور فرمت‌های خاص خودش را تعریف کرد. همین امر، نیازمند یک‌سری تغییرات بود و اساساً برای تبدیل اطلاعات، باید مقدماً کارهایی را انجام می‌دادیم تا به بهترین شکل، به این هدف برسیم. فرایند تصحیح متون هم به‌صورت هوشمند صورت گرفته است.

همان‌طور که عرض کردم، در سطح دوم که بیانگر فرآوری اطلاعات است، برچسب‌گذاری‌هایی روی متون، مانند آیات و احادیث، به‌صورت ماشینی انجام شد و به همین شکل، برچسب‌گذاری‌های موضوعی، کلیدواژه‌ها و نمایه‌هایی نیز برای متون تهیه و ارائه شد.

در سطح سوم، یعنی ارائه اطلاعات، در نرم‌افزارهای نور این قابلیت را ایجاد کردیم تا شیوه دسترسی به محتوای کتب، تسهیل شود؛ مانند قابلیت کاربردی مشابه‌یابی متون که علاوه بر نرم‌افزارهای دستکتایی، در وبگاه‌هایی همچون نور لایب و نورمگز نیز استفاده می‌شود. گروه‌بندی احادیث، یکی دیگر از این دست امکانات است که به‌خصوص در حوزه چت‌بات‌ها خیلی به ما کمک کرد؛ زیرا می‌توانند از تکرار اطلاعات جلوگیری نمایند.

از جمله مقوله‌های مطرح در بحث هوش مصنوعی می‌توان به مواردی نظیر: اوس.آر، استخراج اطلاعات، ترجمه ماشینی و غیره اشاره کرد که همه اینها را می‌توانید در پایگاه آزمایشگاه هوش مصنوعی مرکز نور مشاهده کنید.

یکی دیگر از ابزارهای خوب در این عرصه، تحلیلگر صرفی نور است که قادر است، ویژگی‌ها و حالت‌های مختلف یک کلمه را از متن استخراج کند. اعراب‌گذاری هوشمند برای متون عربی اسلامی، از دیگر این قابلیت‌هاست که مبتنی بر چهارصد جلد کتاب اعراب‌خورده است و می‌تواند با دقت نود درصدی، متن شما را اعراب‌گذاری کند.

از دیگر قابلیت‌های مطرح در سطح ارائه اطلاعات، امکان برچسب‌گذاری آیات است. این سیستم می‌تواند آیات موجود در یک متن را بیابد. این قابلیت هوشمند، از رسم‌الخط‌های مختلف قرآن پشتیبانی می‌کند و شما را به وبگاه قرآن نور متصل می‌سازد. مشابه آنچه در مورد آیات گفتیم، برچسب‌گذاری خودکار احادیث را هم داریم که از امکانات هوشمند در حوزه ارائه اطلاعات حدیثی به شمار می‌رود؛ یعنی کاربر به کمک این فناوری می‌تواند احادیث داخل متن را شناسایی کند و برای اطلاعات کامل‌تر، شما را به پایگاه جامع الأحادیث هدایت نماید و در آنجا می‌توانید به اطلاعات پیرامونی یک حدیث، مانند: سند، ترجمه و یا شرح آن دسترسی داشته باشید. چنانچه در یک متن، احادیث جعلی وجود داشته باشد، به‌وسیله چت‌بات‌های حدیثی قابل شناسایی هستند. در واقع، اگر حدیثی در پایگاه داده و منابع اصلی کتابخانه حدیثی نور وجود نداشته باشد، گواه آن است که این روایت در منابع حدیثی وجود ندارد و احتمال جعلی‌بودنش وجود دارد.

قابلیت‌هایی که اشاره کردم، در پایگاه‌های نور هم استفاده می‌شود؛ مانند: حوزه‌نت، نور لایب و نورمگز. در واقع، همه واژگان به شکل ماشینی استخراج شده و از توانمندی‌های پژوهشی و فناوری‌های مرتبط





یکی از مباحث در روند تولیدات نور، بحث تبدیل اطلاعات است. در گذشته، زبان فارسی در کامپیوتر خیلی راحت قابل استفاده نبود. برای همین منظور، مرکز نور فرمت‌های خاص خودش را تعریف کرد. همین امر، نیازمند یک سری تغییرات بود و اساساً برای تبدیل اطلاعات، باید مقدماتاً کارهایی را انجام می‌دادیم تا به بهترین شکل، به این هدف برسیم. فرایند تصحیح متون هم به صورت هوشمند صورت گرفته است



با هوش مصنوعی برخوردار شده‌اند.

قابلیت دیگر، بحث برقراری پیوند بین متن و پاورقی‌ها در نورلایب است. در پایگاه نورلایب، خیلی از کتاب‌ها از طریق پاورقی، به منبع اصلی خودشان ارجاع می‌دهند. بدیهی است که یافتن این گونه ارجاعات، به شکل دستی و سنتی، خیلی سخت و زمان‌بر است؛ به خصوص اگر آن کتاب، چاپ‌های مختلفی داشته باشد. ما به کمک هوش مصنوعی، پاورقی‌ها را به متن اصلی متصل یا مرتبط کردیم تا کاربر یا محقق از مطابقت و صحت این ارجاعات، مطمئن شود. بیش از چهار میلیون پاورقی در نورلایب، به این شیوه «لینک» داده شده است.

همچنین، طبقه‌بندی متن و برچسب‌گذاری موضوعی، مانند موضوعات فقهی، قرآنی یا حدیثی، از دیگر ابزارهای ماشینی و هوشمند به شمار می‌رود.

استفاده از آیات مرتبط، یکی از قابلیت‌های هوشمند در خصوص استفاده بهینه از آیات قرآن است؛ یعنی وقتی شما در یک متن، با آیه‌ای خاص مواجه شوید، این فناوری به شما آیات مرتبط با آیه انتخابی را معرفی می‌کند؛ یعنی آیه‌ای که با هم ارتباط معنایی دارند، در دسترس قرار می‌گیرند. این قابلیت، در پایگاه قرآن نور وجود دارد.

بحث ترازبندی خودکار متن ترجمه هم از دیگر ابزارهای هوشمندی است که در محصولات خود تدارک دیده‌ایم. به کمک این قابلیت، ترجمه‌های موجود از یک کتاب یا متن، با هم مقایسه گردیده، مطابقت داده می‌شوند و ارتباط بین متن و ترجمه نیز مشخص می‌شود و کاربر می‌تواند به مقایسه و یا انتخاب ترجمه دلخواه خود اقدام نماید. این ابزار، در پایگاه آزمایشگاه هوش مصنوعی نور در دسترس قرار دارد.

بحث هم‌ترازی ترجمه‌های قرآن نیز از دیگر قابلیت‌های خوبی است که برای کاربران و قرآن‌پژوهان قابل استفاده است. این ابزار، در نرم‌افزار جامع التفاسیر نسخه چهارم موجود است. وقتی شما کلمه یا عبارتی از آیات قرآن را انتخاب کنید، در تمام ترجمه‌ها کاوش می‌کند و ترجمه‌های مرتبط با کلمه یا عبارت انتخابی شما را ارائه می‌نماید. بدیهی است که این توانمندی، نگرش خوبی به محقق می‌دهد تا پژوهش خود را هدایت کند. همچنین، شما می‌توانید مثلاً معادل‌های فارسی این کلمه را در ترجمه‌های مختلف موجود در تفاسیر قرآن مشاهده نمایید.

امکان دیگر، جست‌وجوپذیر کردن متون تصویری است. حدوداً نیمی از مقالات نورمگز و کتاب‌های نورلایب، به شکل اسکن شده یا تصویری است و قابلیت جست‌وجو نداشتند. این امر، مشکلاتی را از

نظر دسترسی کاربران به متن آنها ایجاد می‌کند. برای رفع این مشکل، بالای نود درصد این دست متون را به متن جست‌وجوپذیر تبدیل کرده‌ایم؛ ضمن اینکه بخش جست‌وجوشده، در متن تصویری نیز به شکل رنگی نمایش داده می‌شود.

قابلیت استخراج خودکار اعلام و اشخاص از متون تاریخی که در پایگاه تاریخ مرکز نور تعبیه شده، از دیگر ابزارهایی است که کاربران تاریخ‌پژوه را در مسیر پژوهش در متون تاریخی یاری می‌رساند. برای اطلاع بیشتر از سایر ابزارها و محصولات مرکز نور در حوزه AI می‌توانید به پایگاه آزمایشگاه هوش مصنوعی نور مراجعه کنید.»

تازه‌های نور در عرصه هوش مصنوعی

«در چند سال اخیر، تحولات بزرگی در زمینه هوش مصنوعی رخ داده است و تا حدودی در زندگی روزمره نیز ورود پیدا کرده است. اگرچه مرکز نور از حدود پانزده سال پیش به این عرصه ورود داشته، ولی لازم است که همواره خودمان را به‌روز کنیم و با تکنولوژی‌های و مدل‌های جدید، همگام شویم تا کیفیت تولیداتمان و نیز سهولت و سرعت دسترسی کاربران به آنها، افزایش یابد.

به‌طور کلی، مدل‌های زبانی بزرگ را در چهار سطح می‌توان بهبود داد: سطح اول، به مهندسی پرامپت مربوط است که راحت‌ترین روش است و بیشترین کاربرد را دارد. فقط شما باید یک سری اصول را به‌هنگام پرامپت‌نویسی رعایت کنید و در واقع، مسیر را برای مدل روشن کنید تا پاسخ بهتری را به شما عرضه کند.

در سطح دوم، باید از مدل‌های زبانی در سیستم استفاده کرد که به دانش برنامه‌نویسی نیاز دارد. در سطح بعدی که قدری پیچیدگی دارد، نیازمند دانش بیشتری هستیم که به نام «تنظیم دقیق» شناخته می‌شود. باید یک سری دیتاها را مطابق برخی استانداردها آماده کرد و عملاً نیازمند جی‌پی‌یوهای نسبتاً قوی است.

در سطح آخر هم آموزش مدل از صفر مطرح است که تنها در چند شرکت بزرگ دنیا وجود دارد؛ زیرا هزینه بسیار بالایی دارد و نیازمند دیتای خیلی زیادی است و فقط در انحصار چند شرکت است. در مقوله مهندسی پرامپت، باید به موارد زیر دقت کرد.

نخست، تعریف نقش برای هوش مصنوعی است؛ یعنی تعیین یک شخصیت یا تخصص برای مدل جهت دریافت پاسخ بهتر. مورد دیگر، تعیین وظیفه است؛ یعنی مشخص کردن دقیق کاری که مدل



باید انجام دهد. ویژگی بعدی پرامپت، تعیین محدودیت‌ها یا ساختار است که بر این اساس، باید قوانین، لحن، طول یا قالب خروجی را مشخص نمود. مطلب دیگر، بحث زمینه است؛ یعنی ارائه اطلاعات پیش‌نیاز یا سابقه مرتبط برای کمک به درک بهتر درخواست. مرحله بعدی، مثال است؛ یعنی ارائه یک یا چند جفت مثال (ورودی و خروجی مورد انتظار) برای آموزش الگو به مدل. ویژگی دیگر، زنجیره فکری است؛ به بیان دیگر، از مدل خواسته می‌شود تا مراحل استدلال خود را قبل از ارائه پاسخ نهایی توضیح دهد و در نهایت، باید به تقسیم وظیفه پرداخت؛ یعنی شکستن یک وظیفه پیچیده، به چندین زیروظیفه و هدایت مدل برای حل آنها به صورت متوالی. وقتی شما یک متن خیلی طولانی را به مدل بدهید، کیفیت آفت پیدا می‌کند؛ مثلاً اگر بخواهید یک کتاب بزرگ را ترجمه کنید، اگر همه آن را یکجا به مدل بدهید، کیفیت ترجمه کاهش می‌یابد؛ زیرا برخی قسمت‌ها را خلاصه یا حتی حذف می‌کند؛ اما اگر این را در بخش‌های کوچک به مدل بدهیم، آن را با کیفیت بهتری انجام خواهد داد.

در سطح دوم بحث، این موضوع مطرح است که گاهی مدل‌ها پاسخ‌های اشتباهی یا ناقص می‌دهند؛ در صورتی که اگر محتوای مرتبط را در اختیار مدل قرار دهیم، پاسخ بهتری دریافت خواهیم کرد؛ مثلاً از شخصی سؤال تخصصی می‌پرسید، او با حافظه ذهنی خودش پاسخ می‌خواهد؛ ولی زمانی که یک مطلب مشخص از یک کتاب را از او بپرسید، مطابق همان متن جزئی به شما جواب می‌دهد و دانش‌ها و اطلاعات دیگر خودش را در جواب دخالت نمی‌دهد؛ بدیهی است که دقت پاسخ، بالاتر می‌رود و کیفیت کار، توسعه پیدا می‌کند.

به‌روبودن اطلاعات سیستم، خیلی مهم است؛ روزآمد کردن دیتابیس، کم‌هزینه است و خیلی به‌صرفه‌تر و راحت‌تر از آموزش دادن به مدل است و میزان اطمینان و شفافیت آن هم بیشتر است و در واقع، انعطاف بالایی دارد. دلیل کمتر بودن هزینه، آن است که شما با مدل‌های کوچک‌تر هم می‌توانید پاسخ‌های خوبی بگیرید و لازم نیست حتماً مدل خیلی بزرگی داشته باشید.

معماری مبتنی بر بازبازی، این است که در مرحله اول، اسناد را به یک‌سری بخش‌های کوچک تقسیم می‌کنید و بعد به یک مدل تعبیه متن می‌دهید که اینها را به یک بردار عددی تبدیل کند. سپس، آنها را در یک دیتابیس ذخیره می‌کند. این، فاز اول کار است. وقتی کاربر سؤال خودش را مطرح می‌کند، یک جست‌وجوی معنایی در دیتاهایی که ذخیره شده، صورت می‌گیرد و مطالب مرتبط با سؤال کاربر



قابلیت دیگر، بحث برقراری پیوند بین متن و پاورقی‌ها در نورلایب است. در پایگاه نورلایب، خیلی از کتاب‌ها از طریق پاورقی، به منبع اصلی خودشان ارجاع می‌دهند. بدیهی است که یافتن این گونه ارجاعات، به شکل دستی و سنتی، خیلی سخت و زمان‌بر است؛ به خصوص اگر آن کتاب، چاپ‌های مختلفی داشته باشد. ما به کمک هوش مصنوعی، پاورقی‌ها را به متن اصلی متصل یا مرتبط کردیم تا کاربر یا محقق از مطابقت و صحت این ارجاعات، مطمئن شود

استخراج می‌شود و با توجه به اطلاعاتی که بازیابی شده، پاسخ نهایی کاربر داده خواهد شد. بنابراین، اگر سیستم مطالب خوبی را بازیابی کند، نتیجه خوبی را هم ارائه خواهد کرد.

برای دریافت یک پاسخ مطلوب، باید مدل را بهبود بخشید؛ برای مثال، طبق یک مدل، کاربران به نتایج جست‌وجو امتیازدهی می‌کنند و در نتیجه، پاسخ‌هایی که بیشتر با سؤال کاربر ارتباط دارند، در ابتدا قرار می‌گیرند و این امر، باعث افزایش کیفیت می‌شود.

بنابراین، ما هم مدل‌ها را توسعه دادیم و هم قابلیت‌هایی به آنها اضافه کردیم تا کیفیت و دقت بالاتری داشته باشند؛ از جمله می‌توانم به پایگاه جامع الأحادیث اشاره کنم که در آن، قابلیت بازیابی معنایی احادیث گنجانده شده است و تقریباً سه ماه پیش از این فناوری رونمایی شد. همچنین، در نرم‌افزار جامع الأحادیث نسخه چهار نیز قابلیت مشابهت‌یابی معنایی احادیث وجود دارد.

در این سیستم، مدل زبانی دیگری وجود دارد که درخواست کاربر را توسعه می‌دهد تا به شکلی مناسبی بازیابی شود؛ چون ممکن است سؤال کاربر، به همین شکل، مناسب بازیابی نباشد و نیازمند یک سری حذف، اضافه یا توضیح بیشتر باشد تا سیستم بازیابی، راحت‌تر کار کند و حتی گاهی لازم است در فرایند بازیابی، متن را به زبان عربی هم تبدیل کنیم تا نتیجه و خروجی کار، راحت‌تر به دست بیاید.

همچنین از امکان «گروه‌بندی احادیث» هم استفاده نموده‌ایم؛ چون ما در متون روایی، احادیث تکراری بسیاری داریم. بنابراین، تنوع احادیث، در بازیابی‌ها بهتر خواهد بود و پاسخ جامع‌تری در دسترس کاربر قرار می‌گیرد. مطلب دیگر اینکه بخش‌بندی احادیث از نظر موضوعی، کمک می‌کند که مطالب مرتبط با حدیث، راحت‌تر بازیابی شود؛ البته ما علاوه بر بازیابی ماشینی، از ارزیابی انسانی هم بهره برده‌ایم تا خروجی بهتری داشته باشیم.

محصول و خدمت دیگری که به‌تازگی عرضه کرده‌ایم، «سامانه گفت‌وگو با تفاسیر قرآن» است. در این سامانه، اگر سؤال شما به یک مفسر خاصی مربوط باشد و یا مبتنی بر یک کتاب تفسیری مشخص باشد و یا به یک قرن یا مذهب خاصی مرتبط باشد، سامانه فقط روی همان مفسر یا کتاب یا قرن یا مذهب موردنظر کاربر جواب می‌دهد؛ یعنی محدوده سؤال کاربر را تشخیص می‌دهد و فقط در همان مورد یا موارد، به بازیابی می‌پردازد و جواب مناسب ارائه می‌کند.

در این سامانه، ما محتوای تفاسیر را به دو زبان فارسی و عربی بازیابی می‌کنیم و جوابی که به کاربر داده می‌شود، بر اساس زبان رابط کاربر تنظیم شده است.

بحث بعدی، ابزار «رتبه‌بندی» است. برای اینکه نتایج جست‌وجو تعدیل شود و مطالب تکراری کمتری داشته باشیم، این قابلیت را فعال کرده‌ایم؛ به‌طوری‌که در سامانه گفت‌وگو با تفاسیر، خود سیستم پاسخ‌هایی را که به هم شبیه هستند، حذف می‌کند و یکی را به‌عنوان جواب نهایی انتخاب می‌نماید؛ به‌بیان دیگر، ما مطالب بازبازی شده متنوعی را در اختیار مدل نهایی قرار می‌دهیم که این کار، در دسترسی به نتیجه بهتر تأثیرگذار است؛ ضمن اینکه در گفت‌وگو با تفاسیر، تاریخچه گفت‌وگو هم حفظ می‌شود و کاربر می‌تواند سؤال خودش را ادامه دهد و جزئیات بیشتری را مطالبه کند.

همچنین، ما ابزاری داریم که می‌تواند مثلاً آیه و حدیث را در داخل متن شناسایی کند. چنانچه در متن پاسخ، آیات و احادیث شناسایی شود، برای کاربر هایلایت شده و به پایگاه قرآن یا جامع‌الاحادیث لینک می‌شود.

علاوه‌براین، یک‌سری متادیتاهایی در مورد هر پاسخی که از تفسیر بازبازی شده وجود دارد، و مدل می‌تواند بفهمد که مثلاً این متن، مربوط به چه آیه‌ای از قرآن کریم است. اینها به بازبازی بهتر کمک می‌کند و مطالب مرتبط با سؤال کاربر، راحت‌تر نمایش داده می‌شود.

گاهی کاربر به دنبال آیه خاصی است و می‌خواهد از طریق جست‌وجوی ساده آن را بیابد. ما در این سامانه، بین جست‌وجوی دستی و ساده با جست‌وجوی هوشمند جمع کرده‌ایم تا کاربر استفاده مطلوب‌تری از برنامه ببرد.

به‌منظور ارزیابی سامانه، بیش از ۱۵۰ پرسش و پاسخ تفسیری را آماده کردیم و به کمک مدل‌های زبانی بزرگ آن را ارزیابی نمودیم. سؤالات، به دو شکل ساده و مشکل، دسته‌بندی شده بود. نتیجه این ارزیابی، تأیید ۶۸ درصدی پاسخ‌های سامانه بود. وقتی پاسخ‌ها را بررسی کردیم، متوجه شدیم گاهی سیستم ارزیابی ماشینی دچار اشتباه شده و مطالب اضافه‌تری را اشاره کرده که جزء پاسخ اصلی نبوده است که ما این را به‌عنوان امتیاز منفی در نظر گرفته بودیم. در نهایت، برای اینکه به دقت و ارزیابی بهتری برسیم، به سراغ ارزیابی انسانی رفتیم تا پاسخ‌های دقیق‌تری داشته باشیم. نتیجه این ارزیابی انسانی، تأیید ۷۷ درصدی پاسخ‌ها بود؛ یعنی ارزیابی انسانی، بهتر از ارزیابی ماشینی عمل می‌کرد.

همچنین مدل‌های زبانی بزرگ جدیدتر را آزمایش کردیم که دیدیم دقت آن افزایش پیدا کرد؛ هرچند هزینه این نوع مدل‌ها، خیلی بالاست؛ ولی مدل‌هایی وجود دارد که هزینه کمتری دارد و نتیجه نسبتاً خوبی می‌دهد و ما می‌توانیم مدل مناسب کار خودمان را پیدا کنیم.

دستاوردهای فناورانه دیگر مرکز نور در حوزه علوم اسلامی و انسانی دیجیتال، تولید گزارش ماشینی برای رویدادهای تاریخی است که در پایگاه جامع تاریخ سامان یافته است.

پیش‌تر ذیل هر رویداد از منابع مختلف تاریخی، نشانی آن مشخص بود و یک متن مرتبط با آن هم در

استفاده از آیات مرتبط، یکی از قابلیت‌های هوشمند در خصوص استفاده بهینه از آیات قرآن است؛ یعنی وقتی شما در یک متن، با آیه‌ای خاص مواجه شوید، این فناوری به شما آیات مرتبط با آیه انتخابی را معرفی می‌کند؛ یعنی آیاتی که با هم ارتباط معنایی دارند، در دسترس قرار می‌گیرند. این قابلیت، در پایگاه قرآن نور وجود دارد



اختیار کاربر قرار می‌گرفت. برخی از این منابع، گاهی عربی بود و گاهی نیز فارسی. بعضی از آنها قدیمی بود و بعضی هم جدیدتر. برخی برای اهل تسنن بود و برخی از آنها نیز برای شیعه بود. در اینها، مطالب تکراری هم زیاد است؛ یعنی کتابی که یک واقعه را بیان کرده، ممکن است همان مطلب، در منبع دیگری هم باشد. اگر خود پژوهشگر بخواهد همه اینها را مطالعه و بررسی کند، زمان زیادی را باید صرف نماید.

برای رفع این مشکل و تسهیل کار، منابع تاریخی را به یک مدل هوش مصنوعی دادیم و آن هم جمع‌بندی و خلاصه کرد و به منبع اصلی ارجاع داد. در واقع، به وسیله مدل‌های زبانی بزرگ، بیش از ۱۸۳ هزار درباره رویداد را بر اساس ۲۷۱ هزار گزارش تاریخی تولید کردیم.

البته در این مسیر با مشکلات و چالش‌های بسیاری مواجه بودیم که به برخی از مهم‌ترین آنها اشاره می‌کنم:

- تنوع منابع از نظر: زبان، لحن، رویکرد و مذهب؛
- رعایت ملاحظات دینی و مذهبی، مانند: پرهیز از به‌کاربردن تعبیر دور از شأن در نام شخصیت‌ها و یا قائل‌شدن وزن و اهمیت واقعی برای رویداد (نه اغراق و نه کم‌توجهی)؛
- ناهمگونی گزارش‌ها برای هر واقعه تاریخی؛ از نظر کمی (تعداد گزارش‌ها و حجم مطالب در هر گزارش) و هم از نظر کیفی (دست‌اول بودن یا نبودن منابع)؛
- جلوگیری از توهّم در مدل و ارجاع به منبع هر مطلب؛
- متناسب‌بودن حجم خروجی (با اهمیت و میزان داده تزییق‌شده)؛
- ساختار منظم و منسجم (تیترو، بدنه، ارجاع‌دهی درون‌متنی و نیز عددی، شیوه به‌کاربردن آیات و روایات و عبارات تشریف و...).

کار دیگری که در حوزه هوش مصنوعی انجام دادیم، ترجمه چکیده مقالات در پایگاه نورمگز است. باید اضافه کنیم که ترجمه متن همه مقالات، در دست انجام است؛ البته با حفظ مواردی نظیر فرمت‌های داخل متن و صفحه‌بندی مقاله. این قابلیت، از مقالات طولانی تا ۱۰۰ صفحه نیز پشتیبانی می‌کند. ان‌شاءالله این دستاورد کاربردی، امسال انجام خواهد شد و در پایگاه نورمگز ارائه می‌شود.»

کار دیگری که در حوزه هوش مصنوعی انجام دادیم، ترجمه چکیده مقالات در پایگاه نورمگز است. باید اضافه کنم که ترجمه متن همه مقالات، در دست انجام است؛ البته با حفظ مواردی نظیر فرمت‌های داخل متن و صفحه‌بندی مقاله. این قابلیت، از مقالات طولانی تا ۱۰۰ صفحه نیز پشتیبانی می‌کند. ان شاءالله این دستاورد کاربردی، امسال انجام خواهد شد و در پایگاه نورمگز ارائه می‌شود

چشم‌اندازها

«در بخش پایانی عرایضم، به مهم‌ترین برنامه‌ها و فعالیت‌های آتی مرکز نور در راستای دستاوردهای فناورانه در زمینه علوم اسلامی و انسانی دیجیتال اشاره می‌کنم:

۱. در پایگاه نورلایب در صدد هستیم تا ربات «گفت‌وگو با کتب» را فعال کنیم.
۲. در پایگاه نورمگز دو اقدام را در دستور کار داریم: یکی، ترجمه همه مقالات و ارائه به سه زبان: فارسی، عربی و انگلیسی با قابلیت جست‌وجوست و دوم، ارائه ربات گفت‌وگو با مقالات است.
۳. در پایگاه سمیم نیز این کارها را در دست اقدام داریم: «مشابه‌یابی معنایی» و «تشخیص محتوای تولیدشده توسط مدل‌های زبانی بزرگ» که هنوز آن را شروع نکرده‌ایم؛ اینکه آیا مقاله ارائه‌شده توسط محقق، به‌وسیله هوش مصنوعی تولید شده است یا خیر؟ یا چند درصد از متن توسط هوش مصنوعی نوشته شده است؟
۴. سامانه گفت‌وگو با احادیث راه‌اندازی شد؛ اما قصد داریم آن را ارتقا دهیم؛ مثلاً قابلیت‌های جدیدی را که در سامانه گفت‌وگو با تفاسیر هست، به این ابزار هم اضافه کنیم؛ مانند: بهبود جست‌وجوی معنایی، اضافه کردن امکان ادامه گفت‌وگو و نیز ارتقای گفت‌وگو با حدیث با افزودن قابلیت‌هایی همچون امکان فیلتر هوشمند و اشتراک‌گذاری پاسخ‌ها.
۵. در پایگاه تاریخ، اقدامات ذیل را به سامان خواهیم رساند:
 - استخراج رویدادها از متون تاریخی؛
 - استخراج اعلام اشخاص و اماکن و ارائه به صورت فهرست در پایگاه تاریخ؛
 - ارائه ربات گفت‌وگو با وقایع تاریخی.
۶. از دیگر برنامه‌های آتی ما، تنظیم دقیق مدل‌های تعبیه متن و مدل‌های زبانی بزرگ بر اساس محتوای علوم اسلامی است.
۷. تولید دستیار جامع هوشمند علوم اسلامی، در افق دستاوردهای فناورانه نور قرار دارد؛ یعنی یکپارچه‌سازی همه ربات‌های گفت‌وگو که تاکنون تولید شده و یا تولید خواهند شد.
۸. از دیگر اقدامات ما در آزمایشگاه هوش مصنوعی نور، انتشار برخی کدها و داده‌ها در سطح عمومی است؛ مانند:
 - انتشار برخی کدها در GitHub؛
 - انتشار برخی مجموعه داده‌ها در Hugging Face؛
 - آماده‌سازی جدول رتبه‌بندی مدل‌ها در حوزه علوم اسلامی و انتشار آن در Hugging Face. ■