

نگاهی به پیشینه، اهداف، دستاوردها و چشم‌انداز:

آزمایشگاه هوش مصنوعی علوم اسلامی و انسانی نور



*مهندس احمد ربیعیزاده

arabieei@noornet.net

اشاره

آنچه در این نوشتار می‌خوانید، برگرفته از سخنرانی مهندس احمد ربیعیزاده، مدیر آزمایشگاه هوش مصنوعی و علوم اسلامی دیجیتال نور است که در سلسله نشستهای تخصصی هفتنه پژوهش در سالن اجتماعات مرکز ارائه شده است. مهندس ربیعیزاده، ضمن اشاره به روند تکامل موضوع هوش مصنوعی در این مرکز، به بیان اهداف، دستاوردها و چشم‌اندازهای این مبحث در تولید محصولات نور پرداخت که بهروشنی نشانگر آن است، مرکز نور در طی بیش از سه دهه فعالیت خویش، روند روپردازی را در حوزه هوش مصنوعی علوم اسلامی و انسانی پشت سر گذاشته و برنامه‌های کاربردی و ابزارهای ارزشمندی را در این باره عرضه نموده و مسیر پژوهش در علوم و معارف اسلامی و انسانی را هموار ساخته است.

* مدیر آزمایشگاه هوش مصنوعی و علوم اسلامی دیجیتال نور.

توجهی در نزد رشد قابلیت‌های مبتنی بر هوش مصنوعی در مرکز شکل گرفت و پس از آن، سمینار پردازش هوشمند را داشتیم و هم‌اکنون در سال ۱۴۰۰، بیش از ۳۵ دستاوردهای هوش مصنوعی داریم که در نرم‌افزارها و پایگاه‌های مرکز نمود و بروز داشته که در ادامه بدانها خواهیم پرداخت.

هدف از ایجاد آزمایشگاه هوش مصنوعی

به طور کلی، هدف اصلی از ایجاد آزمایشگاه هوش مصنوعی، تکمیل زیستبوم مورد نیاز برای بهره‌برداری حداکثری از هوش مصنوعی در حوزه‌های علوم اسلامی و انسانی است. این هدف، شامل راهبردهای ذیل است:

۱. معرفی ظرفیت‌ها و دستاوردهای هوش مصنوعی در حوزه علوم اسلامی و انسانی؛
۲. تولید و عرضه پیکرهای استاندارد و سرویس‌های مورد نیاز در پردازش هوشمند محتوای اسلامی؛
۳. انتشار دستاوردها در فضای علمی، پژوهشی و آموزشی در قالب مقاله‌های پژوهشی، برای کمک در انتشار و جریان‌سازی فرآیند تولید و دستاوردها و نرم‌افزارهای مبتنی بر هوش مصنوعی؛
۴. تعامل فعال با مراکز پژوهشی، دانشگاه‌ها و مراکز فعال در حوزه هوش مصنوعی و علوم اسلامی و انسانی؛
۵. ارائه راهکارهای تخصصی پردازش هوشمند به مراکز، مؤسسات و پژوهشگاه‌های علوم اسلامی و انسانی.

پیشینه پردازش هوشمند در مرکز نور

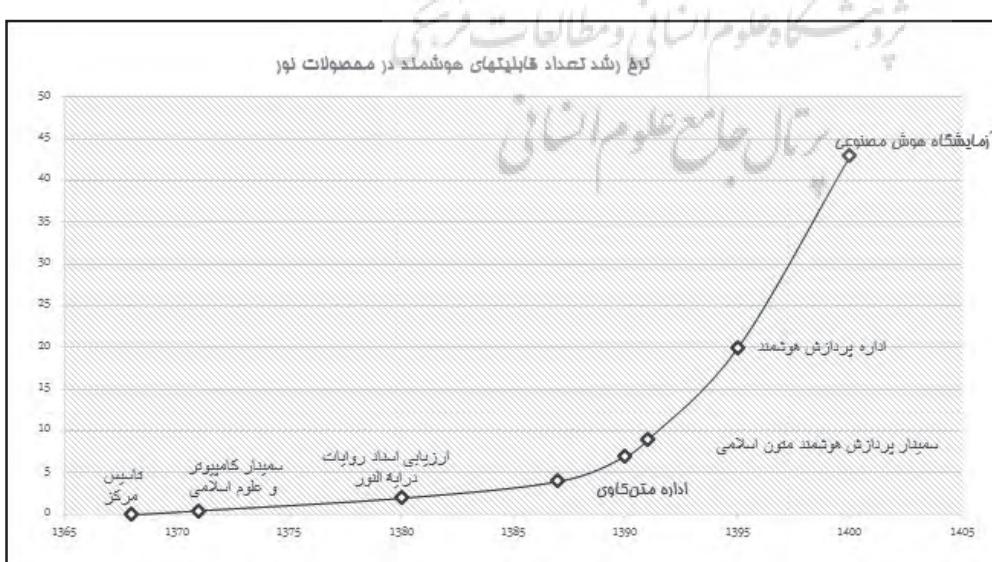
در خدمتستان هستیم با گزارش مختصراً از فعالیت‌های مرکز تحقیقات کامپیوتری علوم اسلامی که طی سه دهه گذشته در زمینه هوش مصنوعی انجام داده و منتج شده به اینکه ان شاء الله امر روز بتوانیم مراسم افتتاح آزمایشگاه هوش مصنوعی نور را رقم بزنیم.

در ابتدا، تاریخچه‌ای از روند تکامل پردازش هوشمند را در مرکز ارائه خواهیم داد و سپس، به اهداف تأسیس آزمایشگاه و نمونه‌های شاخصی که طی این سال‌ها داشتیم و در نهایت، به نقشه راه و چشم‌اندازی که برای آزمایشگاه هوش مصنوعی نور ترسیم شده، خواهیم پرداخت.

همان‌طور که مستحضر هستیم، از سال ۱۳۶۸ که با نظر مقام معظم رهبری مرکز تحقیقات کامپیوتری علوم اسلامی تأسیس شد، توجه فناوری‌های نوین پردازش محتوا، یکی از رسالت‌های اصلی بوده و با فلسفه وجودی مرکز گره خورده است. اوّلین محصول مرکز که مبتنی بر هوش مصنوعی تولید شد، پروژه اعتبارسنجی ماشینی اسناد و روایات بود که مرکز در دهه هشتاد بدان دست یافت. با به کارگیری این محصول در نرم‌افزار درایه‌النور، استاد روایات و احادیث با استفاده از هوش مصنوعی اعتبارسنجی شد و به صورت ماشینی مشخص می‌شد که چه سندی ضعیف و چه سندی معتبر است و بر اساس برآیند اعتبار روایانی که در سند وجود داشتند، تصمیم توسط ماشین پیشنهاد می‌شد. قبل از آن هم سمینار ملی کامپیوتر و علوم اسلامی در سال ۱۳۷۱ در مرکز برگزار شده بود.

پس از آن محصولات متعددی بر همان مبنای شکل گرفتند؛ تا اینکه در سال ۱۳۹۰ اداره‌ای تحت عنوان «اداره متن کاوی» در مرکز تأسیس شد که متولی پیشبرد دستاوردها در زمینه هوش مصنوعی و استفاده از قابلیت‌های مبتنی بر هوش مصنوعی در نرم‌افزارهای نور بود.

پس از تأسیس اداره متن کاوی، جهش قابل



گردید. این پیکره، ارتباط بین کلمات در جمله، خاص دیتای قرآن مشخص شد که بر اساس آن، ماشین می‌تواند از این روابط کنشگر و کنش‌پذیر، اموری چون: فاعل، مفعول، ظرف و... را شناسایی نماید.

۳. مدخل‌بایی هوشمند
کار دیگری که مبتنی بر موتور صرف انجام شد، مدخل‌بایی هوشمند بود. از این قابلیت، در پایگاه قاموس نور بهره‌برداری شده است. در اینجا پس از وارد شدن کلمه مورد نظر در پایگاه قاموس توسط کاربر، نزدیک‌ترین مداخل لغوی به پاسخ مورد نظر او پیشنهاد می‌شود.

۴. اعراب‌گذاری هوشمند
بر اساس این کار، به صورت ماشینی و خودکار، احادیثی را که قادرِ اعراب بودند، بر مبنای یادگیری ماشینی که روی پیکره اعراب‌خورده احادیث داشتیم، توانستیم این احادیث را با کیفیت بالای ۹۵ درصد اعراب‌گذاری کنیم.

۵. مشابه‌بایی احادیث
برای اولین بار، در نرم‌افزار «جامع‌الأحادیث»^{۳۵}، از این قابلیت رونمایی شد که در آن، کاربر با یک کلیک، تمام احادیثی را که شباهت بالایی با حدیث مورد نظر از حیث لفظی دارند، در اختیار خواهد داشت. از این قابلیت، در نرم‌افزار «جامع احادیث فرقین» بین کتب منابع حدیثی شیعه و اهل‌سنّت نیز استفاده شده است. این امکان، به حدیث‌پژوهانی که قصد بررسی

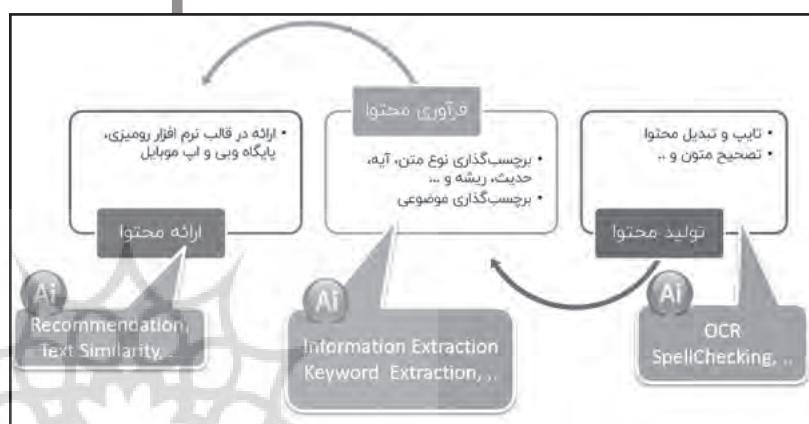
فرآیند تولید نرم‌افزارهای نور

فرآیند تولید نرم‌افزار در مرکز را می‌توان به طور خلاصه، به سه مرحله تقسیم‌بندی کرد:

۱. تولید محتوا شامل مراحل تایپ، تصحیح و OCR؛

۲. فرآوری و غنی‌سازی محتوای تولیدشده در مرحله اول: شامل برچسب‌گذاری، اعراب‌گذاری، مشخص شدن آیات و احادیث؛

۳. ارائه محتوای فرآوری شده در قالب نرم‌افزار.



دستاوردهای مرکز در حوزه هوش مصنوعی

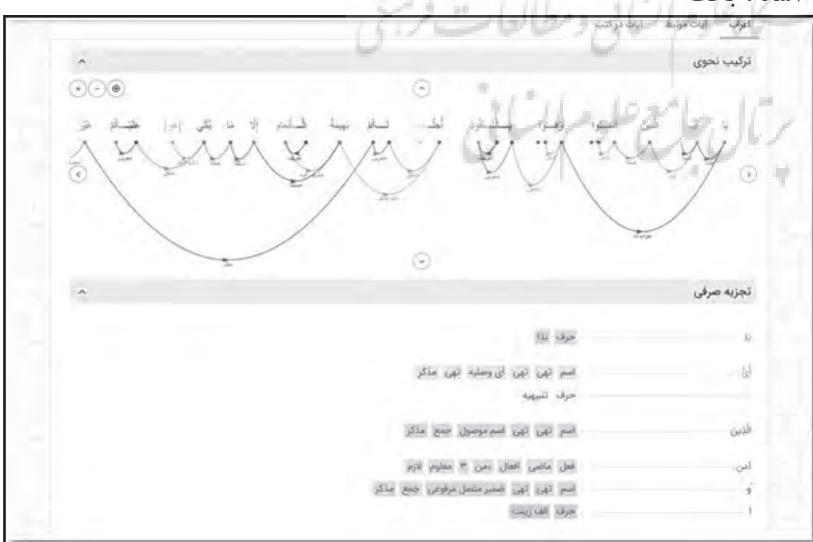
با استفاده از هوش مصنوعی توانسته‌ایم، کیفیت و سرعت فرآوری محتوا را به حداقل برسانیم که در ادامه، به نمونه‌های مشخصی از هر کدام از این موارد، اشاره خواهیم کرد.

۱. تجزیه صرفی ماشینی کلمات عربی

در حوزه زبان‌شناسی رایانشی، طی پروژه‌ای حدوداً ۴ ساله، بحث تجزیه صرفی ماشینی کلمات عربی انجام شد که به عنوان یکی از ابزارهای پایه تحلیل متن و فهم ماشین از متون به کار گرفته شد. با این ابزار که به صورت قانون محور توسعه داده شد، تقریباً تمامی قواعد زبان‌شناختی عربی، از جمله قواعد مربوط به اعلال‌ها، شناسایی ریشه، وزن، پیشوندها و پسوندهای هر کلمه، فراهم گردید و به عنوان سرویسی برای سایر ابزارها و خدمات به کار گرفته شد.

۲. پیکره استاندارد نحوی قرآن

این دستاورده، کار بعدی ما بود و در سطح نحوی که یک سطح فراتر از سطح صرفی است، تهییه



دیجیتال سازی و رقومی سازی متون، یک فرآیند هزینه برای سازمان‌ها در حوزه تولید محتواست. اگر این سازمان در ابعاد مرکز نور باشد که برای تولید نرم‌افزارهای متعدد نیازمند دیجیتال سازی متون بسیاری است، این هزینه بسیار بالاتر خواهد بود. در این حوزه، ما از ترکیب چند موتور OCR و نویسه‌خوان رومی استفاده نمودیم و چرخه‌ای ایجاد کردیم که با تجمیع حداکثری کیفیت این موتورها، بتوانیم زمان دیجیتال کردن متون را به حداقل برسانیم

نسخ مختلف حدیث را دارند، کمک شایانی خواهد کرد و جهشی در فرآیند حدیث‌پژوهی خواهد بود. از این قابلیت، در پایگاه «جامع الأحادیث» نیز استفاده شده است.

مبتنی بر تکنیک مشابه‌یابی، پایگاه سمیم نور نیز شکل گرفت؛ یعنی از این فناوری، برای اصالت‌سنگی مقالات و تولید محتوا استفاده شده است. کاربر مقاله مورد نظر خود را به سیستم سمیم نور وارد می‌کند و مقاله کاربر با انبوی از متون که پشتونه آن، ۵۷۰ هزار مقاله تمام‌من نورمگز، ۳۵ هزار جلد کتاب در پایگاه نورلا یپ و ۱۲۰ هزار پایان‌نامه است، مشابه‌ت‌سنگی می‌شود. این پشتونه ارزشمند، محتوای غنی را تشکیل داده و این موتور مشابه‌یاب، در کنار آن توanstه کیفیت بالایی در مشابه‌یابی مقالات در عرصه اصالت‌سنگی تولید علم به ارمغان آورد.

۶. مشابه‌یابی معنایی

ما پا را از سطح مشابه‌یابی لفظی فراتر گذاشته‌ایم و به سوی مشابه‌یابی معنایی در احادیث رفته‌ایم. گاهی اوقات دو حدیث مورد نظر کاربر، ممکن است از نظر محتوای لفظی تفاوت‌هایی با هم داشته باشند؛ اما ناظر به یک موضوع هستند و به مضمونی واحد از یک محتوا توسط مخصوص اشاره می‌کنند. سیستم بر اساس مشابه‌یابی معنایی، می‌تواند تمامی احادیث موضوع موردنظر کاربر را شناسایی کند و کلمات مرتبط و متراffد را استخراج نماید. سیستم توانسته به اینجا برسد که احادیثی را که به حدیث مورد نظر کاربر بسیار نزدیک هستند، در بین ۴۰۰ هزار حدیثی که هم‌اکنون در پایگاه جامع‌الأحادیث داریم، شناسایی کرده و به کاربر پیشنهاد دهد.

در زمینه برچسب‌گذاری موضوعی متون حدیثی و برچسب‌گذاری موضوعی متون فقهی هم کارهایی را انجام داده‌ایم. مашین با یادگیری خود توانسته احادیث و متون فقهی را از بین ابواب فقهی ۵۰ گانه، برچسب‌گذاری کند که این دستاورده، برای ارائه طبقه‌بندی شده متون فقهی و حدیثی به کاربر نهایی کمک می‌کند.

۷. برچسب‌گذاری آیات

کار دیگر در زمینه غنی‌سازی متون، برچسب‌گذاری آیات بوده است؛ بدین شکل که سیستم، متن خام را دریافت



۱۰. پیشنهاد هوشمند مقالات

شبیه تحلیل‌هایی که در زمینه عبارت‌های جستجو انجام شد، بر اساس باهم‌آیی دانلود یا بازدید دو مقاله در نورمگز، به این نتیجه رسیدیم که چون مقاله «الف» توسط کاربری که مقاله «ب» را دیده، در اینبوه کاربران به مراتب تکرار شده، مقاله دوم می‌تواند مفید باشد و این در حجم بالای لاغ کاربران می‌تواند ما را به سیستم‌های پیشنهاد‌دهنده Recommender system محظوا رهنمون کند که می‌تواند به توسعه پژوهش محققان کمک شایان نماید.

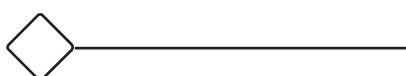
۱۱. ترجمه ماشینی

در حوزه ترجمه ماشینی، ترازبندی خودکار بین متن و ترجمه انجام شده است. در فرآیند پژوهش، بیشتر اوقات، کاربر با متن ترجمه‌ای مواجه است که با متن اصلی متناظر است. تشخیص اینکه دقیقاً کجا این متن ترجمه، به کجا متن اصلی اشاره می‌کند، گاهی برای پژوهشگر زمان بر و مشکل می‌شود. در مرکز قیلاً این کار



فرصت‌های مطالعاتی که می‌توانیم

با دانشگاه‌ها کلید بزیمیم، بحث
تعریف رقابت‌های عملی برای
نیازهای خاص، و در نهایت،
برگزاری رویدادها و کارگاه‌ها در
زمینه پردازش هوشمند محتوای
علوم اسلامی است که - ان شاء الله
- گوشهای از این دستاوردها در
پایگاه آزمایشگاه هوش مصنوعی
رونمایی خواهد شد و در آینده
نزدیک، محتواها و دستاوردهای
پیشتری را شاهد خواهیم بود



کرده و تمامی آیات و روایاتی را که در لابه‌لای این متن به کاررفته،
شناسایی می‌نماید و نشانی آنها را مشخص کرده، به ویگاه جامع
قرآنی یا پایگاه حدیثی ما لینک می‌دهد، تا کاربر بتواند پیرامون
آن حدیث، هر محتوایی را که در مرکز غنی‌سازی شده، از جمله
محتوای تفسیری، محتوای شرح حدیث و سایر محتواها را در
دسترس داشته باشد.

۸. لینک دھی پاورقی کتب

سیستم به صورت هوشمند، پاورقی ذیل هر صفحه را شناسایی کرده، اجزای آن را تحلیل نموده، اسم کتاب را شناسایی می‌کند و در نهایت، کاربر را به صفحه مورد نظر متصل می‌کند؛ بدون آنکه نیاز باشد کاربر، خودش به جستجوی محتوا پردازد. با توجه به پیچیدگی‌های این کار، مثل مواجه بودن با چاپ‌های مختلف یک کتاب، اعتبارسنجی محتوا در پشت صحنه توسط ماشین انجام شده و در صورتی که لینک محتوای نهایی وارد امتیاز و مشابهت لازم باشد، این محتوا به کاربر نمایش داده می‌شود.

در زمینه استخراج خودکار اعلام و اشخاص، موجودیت‌های اسمی متن، اعم از اشخاص و اسامی مکان‌ها و جای‌ها و همین‌طور، وقایع و رویدادها، توسط ماشین شناسایی و لینک شده که امیدواریم بهزودی از این قابلیت در پایگاه جامع تاریخ استفاده و رونمایی کنیم.

۹. یشنہاد آیات مرتبط

کار دیگر ما در حوزه هوش مصنوعی، پیشنهاد آیات مرتبط برای کاربران پایگاه جامع قرآنی است که بر اساس تحلیل باهم آمیخته شده باشد و آن را می‌توانیم از خرد جمعی مفسران استفاده کنیم؛ مثلاً اگر ۲۰ مفسر از دو آیه استفاده کرده‌اند، سیستم تحلیل می‌کند که آیه دوم می‌تواند برای کاربری که آیه اول را در نظر دارد، مفید باشد و به عنوان آیات مرتبط و بیشتر، به کاربر پیشنهاد شود.

مورد دیگری نیز در تحلیل باهم‌آیی‌ها انجام شد؛ به این شکل که با توجه به داشتن حجم عظیمی از لاغ کاربران و جست‌وجوی آنها در پایگاه نورمگز، ما توانستیم پیشنهاد عبارت جست‌وجو را برای کاربران داشته باشیم؛ برای مثال، سیستم در ازای جست‌وجوی «بیداری اسلامی» توسط کاربر، عبارات «بهار عربی»، «اخوان المسلمين»، «انقلاب مصر» را نیز به کاربر پیشنهاد می‌دهد. سیستم بدون اینکه معنای اینها را درک نموده باشد، صرفاً بر اساس تحلیل باهم‌آیی‌ها در حجم بالای لاغ کاربران می‌تواند این پیشنهادهای جذاب را به کاربران ارائه نماید.

سخن پایانی

دیجیتال سازی و رقومی سازی متون، یک فرآیند هزینه برای سازمانها در حوزه تولید محتواست. اگر این سازمان در ابعاد مرکز نور باشد که برای تولید نرم افزارهای متعدد نیازمند دیجیتال سازی متون بسیاری است، این هزینه بسیار بالاتر خواهد بود. در این حوزه، ما از ترکیب چند موتور OCR و نویسه خوان رومی استفاده نمودیم و چرخه ای ایجاد کردیم که با تحمیل حداکثری کیفیت این موتورها، بتوانیم زمان دیجیتال کردن متون را به حداقل برسانیم. در نهایت، در جاهایی که اغلاطی بین این سیستم‌ها مشترک هستند، شناسایی شده و برای اصلاح به تایپیست داده می‌شود و بقیه متن بدون نیاز به اصلاح انسانی، OCR شده و در چرخه فناوری قرار می‌گیرد. از این قابلیت، در پایگاه نورمگز استفاده کرده‌ایم و بیش از یک میلیون صفحه از مقالات نورمگز که صرفاً تصویری بودند، با استفاده از موتور OCR جست‌وجوپذیر شده‌اند و انبوهی از مقالات نورمگز به پایگاه اضافه شده است تا پژوهشگر بتواند در اینها هم جست‌وجو کند و درکنار آن، قابلیت‌های جانبی نیز به سیستم افزوده شده و کاربر می‌تواند عبارت مورد نظر خود را در تصویر، هایلایت و برجسته کند؛ بدون اینکه نیاز داشته باشد دنبال آن عبارت بگردد.

به صورت دستی انجام می‌شد؛ یعنی توسط پژوهشگر، مطالعه و جابه‌جا و صفحه‌به‌صفحه به متن اصلی لینک می‌شد. با ایجاد سیستم ماشینی، این کار به صورت خودکار انجام می‌شود و برنامه، نظریه‌به‌نظری جملات متن ترجمه را به یکایک جملات در متن اصلی متصل می‌کند و حتی بعضی موقع، یک جمله از ترجمه، به دو جمله از متن اصلی مرتبط است و یا بر عکس. شناسایی این موضوع، توسط ماشین صورت می‌گیرد.

در حوزه ترجمه ماشینی، کار دیگری انجام شده که ترجمه‌یابی احادیث است. یکسری از احادیث ترجمه شده‌اند؛ اما احادیثی هم داریم که ترجمه نشده‌اند. در این گونه موارد، سیستم بر اساس ترجمه ماشینی، این احادیث را ترجمه می‌کند؛ اما ممکن است کیفیت نهایی ترجمه ماشینی برای ارائه مستقیم به کاربر نهایی مطلوب نباشد. ما این سیستم ماشینی را با سیستم جست‌وجویی ترکیب کرده‌ایم تا بتواند ترجمه ماشینی را در متن ترجمه شده جست‌وجو نماید و با قرائتی، از جمله وجود متن اصلی حدیث، متن ترجمه نهایی را در کتاب فارسی روانی که یک مترجم، لایه‌لایی متن خود به حسب نیاز آن را ترجمه کرده، شناسایی می‌نماید و به کاربر به عنوان ترجمه صحیح ارائه می‌دهد. ان شاء الله، بهزادی از این قابلیت هم در پایگاه جامع الأحادیث رونمایی خواهد شد.



كتب متناظر عربی و فارسی، با تصویر و متن معادل آنها؛

۳. پیکره تحلیل صرفی زبان عربی که تقریباً بیش از ۳۲۰ هزار کلمه از قرآن کریم و کتاب شرائعالاسلام آمده شده؛ به همراه ۲۰ برچسب صرفی (ریشه کلمات، پیشوندها، پسوندها، وزن کلمه، قسم کلمه و...);

۴. داده محک پیراسته‌سازی کلمات عربی؛

۵. پیکره نمونه احادیث که بالغ بر ۱۵ هزار حدیث از کتاب شریف کافی استاندارد سازی شده؛

۶. پیکره موجودیت‌های اسمی زبان عربی و زبان فارسی که تقریباً ۶هزار حدیث برچسب‌گذاری شده و نیز اسمی اعلام و اشخاصی که درون این احادیث وجود دارند، توسط انسان برچسب‌گذاری شده؛

۷. پیکره مقالات علمی از پایگاه نورمگز که نمونه‌ای است حاوی حدود ۱۰ هزار مقاله به زبان فارسی، به همراه کلیدواژه، چکیده و سایر اقلام داده‌ای موردنیاز پژوهشگر.

گفتی است که تا به حال، بیش از ۱۰ مقاله در سطح ملی و بین‌المللی از کارها و دستاوردهای گروه متن کاوی نور - از این پس، «آزمایشگاه هوش مصنوعی» - منتشر کردایم که با کلید خوردن این آزمایشگاه و نگاه برون‌سازمانی و تعاملی حداکثری که به همراه خواهد داشت، این آمار چند برابر خواهد شد.

علاوه بر این مقالات، نیازمند تعاملات حداکثری با دانشگاه‌ها از طریق مشارکت در تعریف پایان‌نامه‌ها و اشتراک در پروژه‌ها هستیم؛ تا بتوانیم از دانش روزآمد و محققانی که به عنوان طرفیت بالقوه ما هستند، استفاده کنیم. به جای اینکه پژوهشگران روی دیتای فیس‌بوک یا موتورهای جستجوی گوگل، تحقیقات خود را آزمایش کنند، در حوزه علوم اسلامی، مبتنی بر این دادگانی که در اینجا آمده شده و ابزارها و سرویس‌هایی که به عنوان ابزارهای پایه در اختیار ایشان قرار می‌گیرد، این پژوهه‌ها را توسعه دهند.

همین‌طور، فرصت‌های مطالعاتی که می‌توانیم با دانشگاه‌ها کلید بزنیم، بحث تعریف رقابت‌های عملی برای نیازهای خاص، و در نهایت، برگزاری رویدادها و کارگاه‌ها در زمینه پردازش هوشمند محتوای علوم اسلامی است که - ان شاء الله - گوشهای از این دستاوردها در پایگاه آزمایشگاه هوش مصنوعی رونمایی خواهد شد و در آینده نزدیک، محتواها و دستاوردهای بیشتری را شاهد خواهیم بود. ■

در حوزه زبان‌شناسی رایانشی، طی پروژه‌ای حدوداً ۴ ساله، بحث تجزیه صرفی ماشینی کلمات عربی انجام شد که به عنوان یکی از ابزارهای پایه تحلیل متن و فهم ماشین از متن به کار گرفته شد. با این ابزار که به صورت قانون محور توسعه داده شد، تقریباً تمامی قواعد زبان‌شناختی عربی، از جمله قواعد مربوط به اعلام‌ها، شناسایی ریشه، وزن، پیشوندها و پسوندهای هر کلمه، فراهم گردید و به عنوان سرویسی برای سایر ابزارها و خدمات به کار گرفته شد

در حوزه پردازش محتوای صوتی، برای نخستین بار تحقیق و توسعه‌ای در این زمینه صورت گرفت و این قابلیت در نرم‌افزار «مقتل الحسین(علیه السلام)» ارائه گردید؛ به صورتی که کاربر بتواند در محتوای صوتی جست‌وجو کند.

در کنار این قابلیت‌ها، یک سری مجموع دادگان و پیکره‌گانی نیز آمده شد که برخی از اینها منتشر شده و برخی نیز از طریق همین رونمایی در پایگاه آزمایشگاه هوش مصنوعی نور ارائه خواهد شد. این موارد، برای تسهیل فرآیند متن کاوی و پردازش هوشمند در محتوای اسلامی است و اینکه جریان سازی کنیم و پژوهه‌های مشترکی را با نهادهای دانشگاهی و پژوهشگرانی که در این حوزه تخصص دارند، کلید بزنیم.

عنوانی پیکرگان بدین شرح است:

۱. پیکره موازی متن و ترجمه، شامل بیش از ۲۶۰ هزار تکه متن به زبان عربی و ترجمه معادل فارسی آنها؛

۲. پیکره موازی متن و تصویر متن که در سطح خط کلمات از