



طرح شهید احمدی روشن

گزارش نهایی

ششمین دوره طرح شهید احمدی روشن

*خواهشمند است به منظور یکدستی گزارش‌ها، برای نگارش متن از قلم Bzar14 معمولی و برای نگارش تیترها، از قلم Bzar12 سیاه استفاده

شود.*

فصل اول: معرفی هسته

۱. مشخصات هسته	
شماره هسته	۶۲۱۱۱
عنوان هسته	پرسش و پاسخ تصویری
استاد هسته	دکتر سید صالح اعتمادی

۲. معرفی اعضای هسته						
ردیف	نام و نام خانوادگی	مقطع تحصیلی	رشته تحصیلی	گرایش	دانشگاه	عکس (اختیاری)
۱	محمد جواد پیرهادی	کارشناسی ارشد	هوش مصنوعی		علم و صنعت	
۲	مطهره میرزایی	کارشناسی ارشد	نرم افزار		علم و صنعت	
۳	غزاله محمودی	کارشناسی ارشد	هوش مصنوعی		علم و صنعت	
۴	ملیحه حاجی حسینی	کارشناسی ارشد	هوش مصنوعی		علم و صنعت	
۵	ملیکا نوبختیان	کارشناسی	کامپیوتر		علم و صنعت	
۶	یاسمن لطف اللهی	کارشناسی	کامپیوتر		علم و صنعت	
۷	غزل زمانی نژاد	کارشناسی	کامپیوتر		علم و صنعت	
۸	شقایق مبشر	کارشناسی	کامپیوتر		علم و صنعت	

طرح شهید احمدی روشن (دوره ششم)

پرسش و پاسخ تصویری

	علم و صنعت		کامپیوتر	کارشناسی	محمد مهدی جاوید	۹
	علم و صنعت		کامپیوتر	کارشناسی	هادی شیخی	۱۰
	علم و صنعت		کامپیوتر	کارشناسی	یاسمین مدنی	۱۱
	علم و صنعت		کامپیوتر	کارشناسی	بابک بهکام کیا	۱۲
	علم و صنعت		کامپیوتر	کارشناسی	بیان دیوانی آذر	۱۳
	علم و صنعت		کامپیوتر	کارشناسی	باوان دیوانی آذر	۱۴

۳. برگزاری جلسات

ردیف	تاریخ برگزاری	نحوه برگزاری	موضوع مطرح شده (خلاصه)
۱	چهارشنبه ۲۴ فروردین	حضور <input type="checkbox"/> ؛ مجازی <input checked="" type="checkbox"/>	جلسه سرگروه های طرح جهت بررسی پیشرفت کارها: بررسی روند جمع آوری داده ها، پیگیری همکاری با دانشگاه علوم پزشکی، بررسی منابع خبری برای جمع آوری داده.
۲	چهارشنبه ۳۱ فروردین	حضور <input type="checkbox"/> ؛ مجازی <input checked="" type="checkbox"/>	جلسه سرگروه های طرح جهت بررسی پیشرفت کارها: بررسی عینک های هوشمند جهت برنامه نایینایان، بررسی سرور های ابر رایانه سیمرغ و شرایط استفاده از آن ها، بررسی داده ها و مدل های موجود، پیش بردن بات تلگرام.
۳	چهارشنبه ۷ اردیبهشت	حضور <input type="checkbox"/> ؛ مجازی <input checked="" type="checkbox"/>	جلسه سرگروه های طرح جهت بررسی پیشرفت کارها: بررسی و پیاده سازی تسک CLIP برای مسئله پرسش و پاسخ تصویری، بررسی یکی از مجموعه داده های چندزبانه، تمرکز بر مسابقات IMAGECLEF و پیش بردن مدل ها، بررسی استفاده بهینه از سرور ها و منابع سخت افزاری
۴	چهارشنبه ۱۴ اردیبهشت	حضور <input type="checkbox"/> ؛ مجازی <input checked="" type="checkbox"/>	جلسه سرگروه های طرح جهت بررسی پیشرفت کارها: بررسی عینک های هوشمند، بررسی نحوه تصویر برداری از میکروسکوپ برای مسئله پزشکی، آماده کردن مجموعه داده های کپشن تصاویر.

طرح شهید احمدی روشن (دوره ششم)

پرسش و پاسخ تصویری

جلسه با حضور اعضای طرح جهت بررسی مسابقه IMAGECLEF و انتقال تجربه اعضای گروه به سایر اعضای طرح	حضور <input checked="" type="checkbox"/> ؛ مجازی <input type="checkbox"/>	دوشنبه ۱۹ اردیبهشت	۵
جلسه سرگروه های طرح جهت بررسی پیشرفت کارها: بررسی داده های موجود روی سرور ها، تکمیل بات تلگرام.	حضور <input type="checkbox"/> ؛ مجازی <input checked="" type="checkbox"/>	چهارشنبه ۲۱ اردیبهشت	۶
جلسه با حضور اعضای طرح به مناسبت کسب رتبه نخست مسابقات IMAGECLEF	حضور <input checked="" type="checkbox"/> ؛ مجازی <input type="checkbox"/>	سه شنبه ۲۷ اردیبهشت	۷
جلسه سرگروه های طرح جهت بررسی پیشرفت کارها: بررسی کدهای موجود در گیت هاب، بررسی مدل های Generative VQA، پیش بردن فرانت و اپلیکیشن، بررسی جمع آوری دیتاست از طریق طراحی بازی.	حضور <input type="checkbox"/> ؛ مجازی <input checked="" type="checkbox"/>	چهارشنبه ۲۸ اردیبهشت	۸
جلسه سرگروه های طرح جهت بررسی پیشرفت کارها: پیگیری سناریو برای بات تلگرام، تمرکز بر روی مقاله مسابقه IMAGECLEF، بررسی مدل های مبتنی بر ترنسفورمر.	حضور <input type="checkbox"/> ؛ مجازی <input checked="" type="checkbox"/>	چهارشنبه ۴ خرداد	۹
جلسه سرگروه های طرح جهت بررسی پیشرفت کارها: بررسی کارهای انجام شده شامل کار بر روی بک و فرانت وب و اپ، تکمیل و ارسال مقاله، صحبت با دانشگاه علوم پزشکی برا گرفتن داده، دانلود داده WIT و کار بر روی Generative کردن مدل ها با استفاده از RNN و Transformer ها.	حضور <input type="checkbox"/> ؛ مجازی <input checked="" type="checkbox"/>	چهارشنبه ۱۱ خرداد	۱۰
جلسه سرگروه های طرح جهت بررسی پیشرفت کارها: انجام اصلاحات مقاله مسابقه، مصاحبه با خبرگزاری آنا، بررسی مقالات موجود، بررسی جمع آوری داده از طریق crawl، پیگیری سایت آزمایشگاه.	حضور <input type="checkbox"/> ؛ مجازی <input checked="" type="checkbox"/>	چهارشنبه ۱ تیر	۱۱
جلسه سرگروه های طرح جهت بررسی پیشرفت کارها: بررسی سناریو ها برای پروژه گروه ها، پیش بردن پیاده سازی مدل ها، پیش بردن برنامه نایبایان، بررسی مدل های ارائه شده برای آموزش مدل های با یادگیری بدون نظارت، بررسی مدل های گرافی.	حضور <input type="checkbox"/> ؛ مجازی <input checked="" type="checkbox"/>	چهارشنبه ۲۶ تیر	۱۲
جلسه سرگروه های طرح جهت بررسی پیشرفت کارها: بررسی پیشنهاد ها برای تفکیک برنامه نایبایان و افراد بینا، بررسی مدل های آفلاین برای تبدیل متن به صوت، پیش بردن سایت آزمایشگاه، پیش بردن سناریو ها، بررسی مدل های از پیش آموزش دیده.	حضور <input type="checkbox"/> ؛ مجازی <input checked="" type="checkbox"/>	یکشنبه ۲ مرداد	۱۳
جلسه سرگروه های طرح جهت بررسی پیشرفت کارها: پیاده سازی مدل شرح نویسی تصاویر پزشکی برای زبان فارسی، آماده کردن API مدل ها، جمع آوری داده از طریق crawl، پیشرفت سایت آزمایشگاه.	حضور <input type="checkbox"/> ؛ مجازی <input checked="" type="checkbox"/>	یکشنبه ۹ مرداد	۱۴
جلسه سرگروه های طرح جهت بررسی پیشرفت کارها: ساخت مجموعه داده پرسش و پاسخ تصویری به دو زبان فارسی و انگلیسی، جمع آوری داده های بیماری های پوستی از طریق همکاری با دانشگاه علوم پزشکی تهران، جمع آوری داده برای مسئله پرسش و پاسخ تصویری و شرح نویسی تصاویر.	حضور <input type="checkbox"/> ؛ مجازی <input checked="" type="checkbox"/>	یکشنبه ۲۳ مرداد	۱۵
جلسه سرگروه های طرح جهت بررسی پیشرفت کارها: پیاده سازی مدل های هوش مصنوعی برای مجموعه داده جمع آوری شده برای	حضور <input type="checkbox"/> ؛ مجازی <input checked="" type="checkbox"/>	یکشنبه ۱۳ شهریور	۱۶

طرح شهید احمدی روشن (دوره ششم)

پرسش و پاسخ تصویری

مسئله پرسش و پاسخ تصویری در حوزه پزشکی، برگزاری جلسه با گروه دانشگاه علوم پزشکی، پیشبرد بات تلگرام و برنامه نابینایان.			
جلسه سرگروه های طرح جهت بررسی پیشرفت کارها: جمع بندی نهایی کارها جهت ارائه نهایی و بررسی نواقص موجود.	حضور □ ؛ مجازی ■	یکشنبه ۲۰ شهریور	۱۷

- لازم به ذکر است این جلسات بخشی از جلسات تشکیل شده بوده است و جلسات متعدد دیگری توسط هر کدام از گروه ها با اعضای گروه و استاد راهنما در دوران پیشرفت پروژه به صورت مجزا برگزار شده است.

فصل دوم: موضوع فعالیت هسته

شرح تفصیلی موضوع فعالیت هسته مشتمل بر سرفصل های:

۱-۲. معرفی موضوع فعالیت هسته؛

هسته تشکیل شده شامل چند زیربخش می باشد که در ادامه به صورت مجزا به بررسی هریک پرداخته می شود:

۱. **دستیار نابینایان:** هدف از این پروژه، طراحی یک اپلیکیشن (دستیار) هوشمند برای افراد نابینا و کم بینا است که می توانند با استفاده از تلفن همراه هوشمند از اطراف عکس گرفته و علاوه بر دریافت شرح شنیداری عکس، در رابطه با عکس سوال کرده و با فضای اطراف خود ارتباط برقرار کنند. برای پیاده سازی این دستیار هوشمند از دو مدل هوش مصنوعی استفاده می کنیم:

➤ مدل تولید شرح تصویر^۱

➤ مدل پرسش و پاسخ تصویری^۲

وظیفه مدل اول تولید خودکار توصیف مناسب برای تصویر ورودی و وظیفه مدل دوم پاسخ به پرسش های مطرح شده در رابطه با عکس ورودی است. هدف ما در این پروژه پیاده سازی این دو مدل با استفاده از مدل های یادگیری عمیق است.

۲. **تصاویر پزشکی:** هوش مصنوعی در عرصه های مختلفی در حال پیشرفت است. یکی از حوزه هایی که هوش

مصنوعی می تواند تاثیر به سزایی در آن داشته باشد، حوزه پزشکی است. با پیشرفت هوش مصنوعی در حوزه پزشکی می

¹ Image Captioning

² Visual Question Answering

طرح شهید احمدی روشن (دوره ششم)

پرسش و پاسخ تصویری

توان بسیاری از امکانات درمانی را با هزینه کم و دسترسی بالا در اختیار بیماران قرار داد. همچنین این علم می تواند به پزشکان در تشخیص بهتر و سریع تر نیز کمک کند. هدف از این پروژه طراحی مدلی است که بتواند با دریافت یک تصویر پزشکی و یک سوال مرتبط با آن، پاسخ مناسب را تولید کند. این برنامه می تواند به سوالات بیماران پیرامون تصاویر پزشکی شان پاسخ دهد. همچنین به بررسی مدل های طبقه بندی بیماری های پوستی، تشخیص مفاهیم پزشکی از روی تصاویر و همچنین شرح نویسی تصاویر پزشکی نیز پرداخته شده است.

۳. **اخبار جعلی:** حجم قابل توجهی از اخبار و پیام های جعلی در شبکه های اجتماعی به وسیله زیرنویس های اشتباه برای تصاویر منتشر می شود. به عنوان مثال اخیرا تصویری از کشتار مردم عراق تحت عنوان کشتارهای آبان ماه در شبکه های اجتماعی پخش شده است. با توجه به ماهیت تصویری/متنی محتوای جعلی، شبکه های عصبی عمیق موفق در پرسش و پاسخ تصویری، می توانند در تشخیص اخبار جعلی نیز موثر باشند.

۲-۲. پیشینه پژوهشی موضوع در داخل و خارج از کشور؛

۱. **دستیار نایبایان:** در سال های اخیر مدل های یادگیری عمیق مختلفی برای حل دو مسئله شرح تصویر و پرسش و پاسخ تصویری ارائه شده است. اما پس از سال ۲۰۱۹، با معرفی شبکه های از پیش آموزش دیده متن-تصویر^۳ بهبود چشمگیری رخ داد به طوری که این دو مسئله بهترین نتیجه خود را با استفاده از این شبکه ها بدست آوردند. هدف این مدل ها بازنمایی مشترک داده های تصویری و داده های زبانی است. شبکه های LXMERT، ViLBERT، VisualBERT، OSCAR، UNITER، VILLA، ERNIE-ViL، UNIMO، VinVL، SimVLM، VLMO و OFA نمونه ای از این شبکه ها هستند. جدول ۱ عملکرد سه شبکه از پیش آموزش دیده متن-تصویر را که بیشترین دقت را بر روی دادگان VQA v2 بدست آوردند را نمایش می دهد.

مدل	سال انتشار	دقت
OFA	۲۰۲۲	۸۱/۹۸
VLMO	۲۰۲۱	۸۱/۳۰

جدول ۱: دقت شبکه های از قبل آموزش دیده بر روی دادگان VQA v2

جدول ۲ عملکرد سه شبکه از پیش آموزش دیده متن-تصویر را که بیشترین مقدار BLEU Score را بر روی دادگان COCO بدست آوردند را نمایش می دهد.

مدل	سال انتشار	دقت
OSCAR	۲۰۲۰	۴۱/۷
VinVL	۲۰۲۱	۴۱/۰
SimVLM	۲۰۲۱	۴۰/۰۶

³ Vision-and-Language Pretrained Model

جدول ۱: عملکرد شبکه‌های از پیش آموزش دیده بر روی دادگان COCO

۲. **تصاویر پزشکی:** موضوع هوش مصنوعی در پزشکی موضوع به روزی است و در حال حاضر در جهان مورد توجه قرار گرفته است. در حوزه های مختلفی مثل تولید دارو، تشخیص بیماری، تولید ربات های جراح، درمان بیماری و غیره تلاش شده است از کاربرد های هوش مصنوعی در پزشکی استفاده شود اما هنوز جای پیشرفت بسیاری در این حوزه وجود دارد. از سال ۲۰۰۳ رویداد CLEF به برگزاری مسابقاتی در حوزه هوش مصنوعی در پزشکی پرداخته است که سهم مهمی در گسترش اهمیت این حوزه داشته است.

۳. **اخبار جعلی:** در رابطه با این موضوع مقالات آکادمیک متعددی در داخل و خارج از کشور موجود است. البته تعداد مقالات با محوریت مجموعه داده‌های به زبان غیر از فارسی بیشتر است. در رابطه با محصول قابل استفاده در زبان فارسی نیز می‌توان به <http://khabarsaanj.ir> و <http://194.225.229.223:9002> اشاره کرد. اما در رابطه با تشخیص موضع در رابطه با موضوع (target base) محصول مشابهی در داخل از کشور موجود نمی‌باشد.

۳-۲. اهمیت موضوع برای کشور؛

۱. **دستیار نابینایان:** آمار نابینایان و کم‌بینایان حدود ۱۵۰ هزار نفر [گزارش](#) شده است. با توجه به عدم امکان بهره‌گیری این افراد از بسیاری از سرمایه‌های طبیعی کشور، سرمایه‌گذاری کشور در تکنولوژی هوش مصنوعی که باعث ارتقاء سطح کیفی زندگی این افراد بشود، از حقوق این افراد به شمار آمده و دارای اهمیت بالایی می‌باشد. علاوه بر این، تاکنون هیچ تحقیقی در مورد مسئله پرسش و پاسخ تصویری در زبان فارسی انجام نشده است. بنابراین ارائه دادگان مناسب برای این مسئله و آموزش یک مدل برای آن مسیر جدیدی را برای سایر محققین باز خواهد کرد.

۲. **تصاویر پزشکی:** با در نظر گرفتن توجه جامعه جهانی به این حوزه، نیاز است در داخل کشور هم تلاش هایی برای پیشرفت این حوزه صورت گیرد. تا کنون ایران توانسته است به موفقیت های خوبی دست یابد به عنوان مثال تولید ربات جراح سینا یکی از افتخارات مهمی است که دانشمندان ایران به آن دست یافته اند با این حال هنوز زمینه های مختلفی در پزشکی هستند که هوش مصنوعی می تواند به آن ها کمک کند. در صورت پیشرفت در این حوزه، امکان دسترسی آسان و کم هزینه بیماران به برنامه هایی که می تواند به سوالات شان پاسخ دهد فراهم می شود. همچنین پزشکان می توانند از این برنامه ها برای تشخیص بهتر و سریع تر بیماری ها بهره ببرند.

۳. **اخبار جعلی:** با توجه به گسترش روزافزون شبکه‌های اجتماعی و سرعت پخش اطلاعات در این بستر، نیاز به وجود یک راهکار هوشمند و خودکار برای بررسی صحت اطلاعات می‌باشد. همچنین اطلاعات موجود در شبکه‌های اجتماعی داده‌های مفیدی از موضع افراد در مورد موضوعی خاص را می‌تواند مشخص کند و آماری به ما ارائه دهد که در آینده در تصمیم‌گیری‌های کلان به ما کمک می‌کند. به عنوان مثال با توجه به داده‌های جمع‌آوری شده در رابطه با موضوع واکسن کرونا، می‌توان آماری از افراد موافق و مخالف به دست آورد. سپس نظرات مخالف را جمع‌آوری کرده و دلایل مخالفت را بررسی کرد. در نهایت تصمیم‌های مورد نیاز را برای اقناع مردم گرفت. علاوه بر تصمیم‌گیری‌های سیاسی،

طرح شهید احمدی روشن (دوره ششم)

پرسش و پاسخ تصویری

می‌توان نظرات مردم در رابطه با کالایی خاص را جمع‌آوری کرده و به شرکت مربوطه بازخورد داد. بنابراین سرمایه‌گذاری در این حوزه از اهمیت ویژه‌ای برخوردار است.

۴-۲. محدوده فعالیت‌های هسته.

۱. **دستیار نابینایان:** برای آموزش مدل تولید شرح تصویر، نیاز به تعداد زیادی تصویر و شرح آن تصاویر را داریم. طبق پژوهش‌هایی که انجام دادیم، به تازگی دادگان [WIT](#) منتشر شده است که شامل تصاویر و شرح آن است که از سایت ویکی‌پدیا جمع‌آوری شده است و بیش از ۱۰۸ زبان دنیا را پوشش می‌دهد. در این دادگان حدود ۵۰۰ هزار زوج تصویر و شرح آن موجود است که در سه بخش آموزش، ارزیابی و تست ارائه شده است. ما آن بخش از این دادگان را که مربوط به زبان فارسی است را جدا کردیم. علاوه بر این، از ترجمه دادگان‌های مشهور شرح تصویر مانند [Flicker30k](#)، [MS-CoCo](#) و [Conceptual Caption](#) بهره بردیم که با استفاده از Google Translate API ترجمه ماشینی شده است و با روش Multilingual Similarity Check بهترین ترجمه‌ها انتخاب شده است. (ترجمه دیتاست coco-flicker را در این [لینک](#) و ترجمه دیتاست Conceptual Caption را در این [لینک](#) می‌توانید مشاهده کنید. این ترجمه‌ها توسط آقایان سجاد ایوبی و نوید کنعانی تهیه شده است).

یکی از مشکلات بزرگ این مجموعه دادگان‌ها این است که نشان دهنده محیط و فرهنگ غرب است و نمی‌توانند محیط و فرهنگ ایران را به نمایش بگذارند به همین دلیل هفت دسته بندی مختلف شامل غذا، ماشین، فضای خانه، فضای خارج از خانه، مراسمات، اشخاص و ورزش را در نظر گرفتیم و برای هر دسته ۱۰۰ تصویر که نشان دهنده ی فرهنگ و محیط ایرانی باشد را از سایت pinterest استخراج کردیم. سپس [ابزار تحت وبی](#) را توسعه دادیم و از ۴ نیروی انسانی درخواست کردیم که برای هر تصویر توصیفی را با توجه به [راهنمایی](#) که تهیه نموده‌ایم وارد کنند. دادگانی که برای شرح تصویر جمع‌آوری کردیم بخشی از بنچمارک ParsVQA-Caps است که برای اولین بار دادگانی را برای مسئله پرسش و پاسخ تصویری و شرح تصویر در زبان فارسی در این پژوهش معرفی می‌شود. جدول ۳ تعداد تصاویر و شرح مربوط به هر دادگان را نمایش می‌دهد.

دادگان	تعداد تصاویر	تعداد شرح
WIT-fa	۴۳۴۴۹۵	۴۹۳۱۸۴
Flickerfa	۳۲,۴۰۴	۳۲,۴۰۴
cc3mfa	۲۴۲,۰۴۴	۲۴۲,۰۴۴
coco-flicker-farsi	۱۲۴,۱۸۸	۱۲۴,۱۸۸
ParsVQA-Caps	۷۱۳	۲۴۴۰
مجموع	۸۳۳۸۴۴	۸۹۴۲۶۰

جدول ۲: تعداد تصاویر و شرح آن‌ها در دادگان‌های موجود در زبان فارسی

با استفاده از دادگان‌های موجود مدلی را که از CLIP-fa به عنوان کدگذار و از GPT2 به عنوان کدگشا استفاده می‌کند را پیاده‌سازی و آموزش دادیم. نتایج این مدل در جدول ۴ گزارش شده است. در جدول ۵ چند نمونه از شرح‌های تولید شده مدل در مقایسه با شرح‌هایی که انسان تولید کرده است آورده شده است.

Method		Word-based			Embedding-based	
Encoder	Decoder	BLEU	METEOR	ROUGE-L	Average Score	BERT Score
CLIPfa	GPT2	20.20	-	50.30	-	-

جدول ۳: نتایج مدل شرح تصویر بر روی دادگان جمع‌آوری شده

طرح شهید احمدی روشن (دوره ششم)

پرسش و پاسخ تصویری

تصویر	شرح‌های تولیدشده توسط انسان	شرح‌های تولیدشده توسط مدل ما
	<ol style="list-style-type: none"> ۱. یک خیابان باریک و بارانی که یک طرفش ماشین‌ها پارک کرده‌اند. ۲. چند خودرو در حاشیه یک سمت خیابان با درختان بلند پارک کرده‌اند. ۳. رنگ برگ‌های درختان بلند به زرد و نارنجی تبدیل شده‌است. ۴. یک خیابان که تعدادی ماشین در آن جا پارک شده‌اند. 	ماشین‌های پارک شده در یک روز پاییزی در خیابان
	<ol style="list-style-type: none"> ۱. یک زن و مرد پیر در خانه‌ای قدیمی کنار یکدیگر نشسته و هندوانه می‌خورند. ۲. پیرزن با لباس آبی در حال هندوانه خوردن کنار پیرمرد است. ۳. پیرزن و پیرمرد روستایی کنار هم نشسته‌اند و هندوانه می‌خورند. ۴. یک زن و مرد روستایی در اتاق به رخت‌خواب‌ها تکیه داده‌اند. 	یک پیرزن و یک پیرمرد در کنار هم نشسته‌اند و با هم صحبت می‌کنند.
	<ol style="list-style-type: none"> ۱. زنان زیادی با چادر در حال خواندن نماز هستند. ۲. تعداد زیادی خانم با چادرهای رنگی در کنار هم نماز می‌خوانند. ۳. تعداد زیادی خانم در صفوف نماز در حال خواندن نماز جماعت هستند. ۴. زن‌های زیادی با چادر در صف نماز ایستاده‌اند. 	تعداد زیادی از زنان در حال نماز خواندن هستند.

جدول ۴: سه نمونه از شرح‌های تولید شده توسط مدل در مقایسه با شرح‌های تولید شده توسط انسان

برای جمع‌آوری داده برای آموزش مدل دوم، [ابزار تحت وبی](#) را طراحی کردیم و ۱۵۰۰۰ تصویر را به صورت رندوم از دادگان MS-COCO انتخاب کردیم. حدود ۱۱۰ قالب سوال در ۱۰ دسته‌بندی شامل حضور شی، تشخیص ورزش، استدلال مکان، درک احساسات، ویژگی‌های رنگ، شمارش اشیاء، تشخیص فعالیت، تشخیص اشیاء، جنس اشیاء، تشخیص جنسیت طراحی کردیم. این قالب‌ها در دو لحن رسمی و محاوره‌ای تهیه شده است. سپس برای هر تصویر ۲ سوال رسمی و ۲ سوال محاوره‌ای با استفاده از قالب‌های طراحی شده تولید شد و بر روی ابزار تحت وب قرار گرفت. از ۷ نیروی انسانی کمک گرفته شد تا به سوالات با توجه به تصویر پاسخ دهند. همچنین [راهنمایی](#) برای نحوه پاسخ به سوالات همراه با مثال در اختیار نیروهای انسانی قرار داده شد. پاسخی که توسط نیروی انسانی جمع‌آوری شده است کوتاه و در حد یک الی سه کلمه است. برای تولید پاسخ‌های بلند(جمله) الگوهایی طراحی شد تا به صورت اتوماتیک بتوانیم پاسخ‌های بلند را از روی پاسخ‌های کوتاه بسازیم. با توجه به اینکه تصاویر از یک دادگان خارجی برداشته شده است؛ نمی‌تواند به صورت کامل محیط اطراف ما را به تصویر بکشد. از این رو، از ۷۱۳ تصویری که از سایت pinterst برای مسئله شرح تصویر بدست آوردیم؛ استفاده کردیم و از نیروهای انسانی خواستیم که برای این تصاویر یک سوال رسمی و یک سوال محاوره‌ای تولید کنند به طوری که سخت و چالش برانگیز باشد و یک سوال رسمی با درجه سختی کمتر تولید کنند. جدول ۶ تعداد تصاویر و پرسش و پاسخ‌های موجود در این دادگان را برای هر دسته‌بندی از سوالات را نشان می‌دهد.

طرح شهید احمدی روشن (دوره ششم)

پرسش و پاسخ تصویری

دسته بندی	تعداد تصاویر		تعداد سوالات رسمی		تعداد سوالات محاوره‌ای		تعداد پاسخ ها	
	انسانی	الگو	انسانی	الگو	انسانی	الگو	انسانی	الگو
حضور شی	161	5109	130	2527	48	2583	178	5110
تشخیص ورزش	42	625	25	288	17	337	42	625
استدلال مکان	194	2187	160	1124	59	1063	219	2187
درک احساسات	20	464	14	234	6	230	20	464
ویژگی‌های رنگ	331	3638	214	1904	168	1735	382	3639
شمارش اشیاء	372	3668	303	1800	142	1868	445	3668
تشخیص فعالیت	81	1215	57	636	26	579	83	1215
تشخیص اشیاء	158	2090	112	871	58	1219	170	2090
جنس اشیاء	114	1831	73	910	46	921	119	1831
تشخیص جنسیت	114	1022	93	532	25	490	118	1022
سایر	327	0	280	0	120	0	400	0
مجموع	723	9872	1461	10826	715	11025	2176	21851

جدول ۵: تعداد تصاویر، سوال و پاسخ در دادگان جمع آوری شده با دو روش الگو و انسانی برای پرسش و پاسخ تصویری در زبان فارسی

برای حل مسئله پرسش و پاسخ تصویری یک مدل کدگذار و کدگشا طراحی و پیاده‌سازی کردیم و بر روی دادگان FSVQA در زبان انگلیسی و بر روی دادگان جمع آوری شده در زبان فارسی آموزش دادیم. از کدگذارهای [LXMERT](#) و [VisualBERT](#) و کدگشاهای [RNN](#)، [Attention RNN](#) و [Transformer](#) برای زبان انگلیسی و از کدگذار [CLIPFA](#) و کدگشا [Transformer](#) برای زبان فارسی استفاده کرده‌ایم. نتایج آزمایش‌ها بر روی دادگان FSVQA در جدول ۷ آورده شده است.

Method		Word-based			Embedding-based	
Encoder	Decoder	BLEU	METEOR	ROUGE-L	Average Score	BERT Score
LXMERT	1-LSTM	32.19	56.65	56.38	86.5	79.19
	2-LSTM	41.28	64.39	62.97	89.57	83.1
	3-LSTM	41.1	64.25	63.02	89.54	82.67
	1-GRU	39.37	62.99	61.5	89.63	82.07
	2-GRU	35.23	59.57	58.24	88.56	79.68
	1-BiLSTM	41.97	65.25	63.65	90.11	83.51
	2-BiLSTM	43.26	66.28	64.67	90.47	84.09
	3-BiLSTM	43.54	66.39	64.74	90.58	84.02
	1-BiGRU	40.89	64.15	62.66	89.88	82.46
	2-BiGRU	34.32	58.54	57.29	88.74	79.27
	3-BiGRU	27.88	54.82	54.55	86.66	74.72
	1-LSTM+Bahdanau attention	79.03	86.43	85.49	95.94	91.84
	1-LSTM+Luong(dot) attention	78.79	86.9	86.05	96.2	91.94
	1-LSTM+Luong(general) attention	79.54	86.96	86.25	96.11	91.9
	1-GRU+Bahdanau attention	71.4	83.26	82.62	95.42	89.32
	1-GRU+Luong(dot) attention	73.92	84.71	83.84	95.73	90.37
	1-GRU+Luong(general) attention	58.44	75.92	75.04	93.61	85.4
	1-GRU+Luong(concat) attention	36.53	60.62	59.48	88.88	79.99
	3-Transformer Decoder	86.73	91.18	90.6	90.2	95.01
	4-Transformer Decoder	85.98	90.91	90.33	90.13	94.79
VisualBERT	1-LSTM	18.62	38	38.35	84.83	69.2
	2-LSTM	18.92	38.62	38.42	85.3	70.4
	3-LSTM	19.3	39.33	38.8	85.87	71.35
	1-GRU	22.51	44.24	43.72	87.74	73.59
	2-GRU	21.96	45.16	44.93	88.21	73.26
	3-GRU	11.36	32.34	33.99	84.23	67.15
	1-BiLSTM	20.03	39.7	39.75	85.52	71.19
	2-BiLSTM	18.84	40.05	40.27	86.67	72.13
	3-BiLSTM	20.46	41.51	41.19	87.21	73.43
	1-BiGRU	22.72	45.66	45.26	88.42	74.32
	2-BiGRU	19.7	42.21	41.84	88.04	73.33
	3-BiGRU	15.76	36.74	36.37	85.93	69.14
	1-LSTM+Bahdanau attention	84.27	88.07	87.28	97.11	93.5

طرح شهید احمدی روشن (دوره ششم)

پرسش و پاسخ تصویری

1-LSTM+Luong(dot) attention	80.9	86.68	85.81	96.98	92.12
1-LSTM+Luong(general) attention	82.41	87.35	86.42	97.1	92.68
1-LSTM+Luong(concat) attention	82.9	87.71	86.9	97.17	93.11
1-GRU+Bahdanau attention	72.2	82.87	82.4	96.1	89.26
1-GRU+Luong(dot) attention	79.65	86.17	85.23	96.93	91.81
1-GRU+Luong(general) attention	72.4	82.15	81.21	96.17	89.22
1-GRU+Luong(concat) attention	29.64	54.3	54.01	89.13	70.5
3-Transformer Decoder	85.95	89.76	89.09	91.94	94.44
4-Transformer Decoder	85.99	89.78	89.16	91.95	94.52

جدول ۶: نتایج آزمایش‌ها بر روی دادگان FSVQA

نتایج آزمایش بر روی دادگان جمع‌آوری شده فارسی در جدول ۸ نشان داده شده است. خروجی مدل بر روی سه تصویر نیز در جدول ۹ آورده شده است.

Method		Word-based			Embedding-based	
Encoder	Decoder	BLEU	METEOR	ROUGE-L	Average Score	BERT Score
CLIPfa	3-transformer	16.59	53.75	-	43.37	87.35

جدول ۷: نتایج آزمایش بر روی دادگان جمع‌آوری شده برای پرسش و پاسخ تصویری در زبان فارسی

تصویر	سوال	پاسخ
	چنگال جنسش چیه؟	پاسخ درست: چنگال جنسش پلاستیک و فلزه. پاسخ مدل: چنگال جنسش پلاستیکه.
	قاب گوش‌های مرد چه رنگیه؟	پاسخ درست: رنگ قاب گوش‌های مرد در تصویر مشکیه. پاسخ مدل: تلفن همراه مشکیه است.
	در تصویر چند مرد وجود دارد؟	پاسخ درست: چهار مرد در تصویر وجود دارد. پاسخ مدل: دو مرد در تصویر وجود دارد.

جدول ۸: سه نمونه از خروجی مدل پرسش و پاسخ تصویری در زبان فارسی

طرح شهید احمدی روشن (دوره ششم)

پرسش و پاسخ تصویری

9

تصویر را اینجا رها کنید یا برای انتخاب تصویر کلیک کنید

سوال : سوالی هنوز پرسیده نشده است.

نتایج :

-

دریافت شرح سوال

پرسش سوال

برای انتخاب تصویر دیگر کلیک کنید.

سوال : سوالی هنوز پرسیده نشده است.

نتایج :

-

دریافت شرح سوال

دریافت پاسخ

برای انتخاب تصویر دیگر کلیک کنید.

سوال : این چیست

نتایج :

خرس عروسکی 96.09789276123047

دریافت شرح سوال

پرسش سوال

شکل ۱: نمونه‌ای از اپلیکیشن طراحی شده

پس از جمع‌آوری داده و آموزش مدل‌های مربوطه، اپلیکیشنی را طراحی کردیم که یک تصویر را از کاربر می‌گیرد و شرح تصویر و پاسخ سوالات کاربر را با استفاده از مدل‌های طراحی شده در خروجی نشان می‌دهد. برای دریافت سوال از یک سیستم تبدیل گفتار به متن استفاده کرده‌ایم تا امکان استفاده از آن برای افراد نابینا فراهم شود. شکل ۱ نمایی از این اپلیکیشن را نمایش می‌دهد.

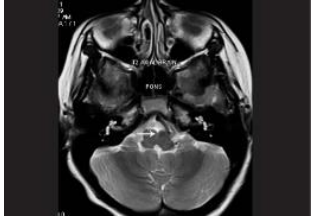
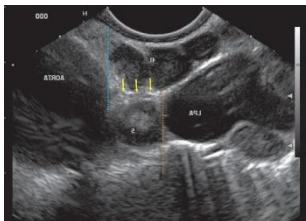
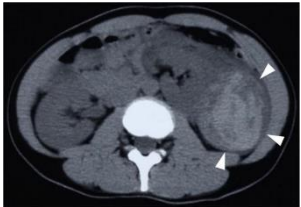
۲. **تصاویر پزشکی:** در بخش پزشکی نیاز به داده یکی از چالش‌های اصلی پیاده‌سازی مدل‌ها می‌باشد. با توجه به اینکه در زبان فارسی پژوهش خاصی در حوزه هوش مصنوعی در پزشکی صورت نگرفته است، نیاز به جمع‌آوری داده بسیار اهمیت پیدا کرده است. در این پژوهش برای جمع‌آوری داده از روش‌های مختلفی استفاده شده است. در بخش تولید مفاهیم پزشکی و شرح نویسی تصاویر پزشکی، از مجموعه داده منتشر شده در مسابقات جهانی ImageCLEF استفاده شد. این مجموعه داده شامل حدود ۹۰ هزار تصویر پزشکی است که برای هر تصویر شرح مناسب و مفاهیم مرتبط درج شده است. این مجموعه داده به زبان انگلیسی منتشر شده است. برای تولید مدل‌های فوق در زبان فارسی، از روش ترجمه مجموعه داده استفاده شد تا بتوان از ظرفیت این مجموعه داده غنی استفاده کرد.

در بخش مسئله پرسش و پاسخ تصویری پزشکی، با توجه به پیچیده شدن مسئله و ترکیب پردازش متن و تصویر، کار جمع‌آوری داده سخت‌تر می‌شود. رویکرد سوال و جوابی این مسئله نیازمند دقت بالایی است. تا کنون مجموعه داده‌های انگشت شماری در جهان برای این مسئله جمع‌آوری شده‌اند که اغلب تعداد داده کمی را شامل می‌شوند. در این پژوهش با استفاده از مجموعه داده ImageCLEF و مفاهیم مشخص شده برای هر تصویر، به صورت خودکار سوالات و پاسخ‌هایی تولید شد. این سوالات با محوری بررسی اندام‌های بدن، روش‌های تصویر برداری پزشکی، بررسی صفحات بدن و ناهنجاری‌ها و بیماری‌های موجود طراحی شده‌اند. در جدول زیر چند نمونه از این مجموعه داده نمایش داده شده است.

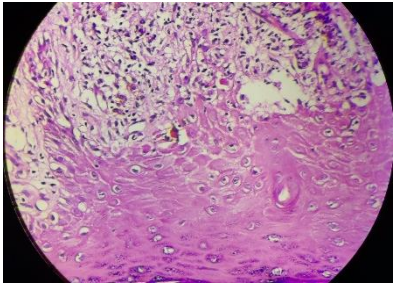
۱۲

طرح شهید احمدی روشن (دوره ششم)

پرسش و پاسخ تصویری

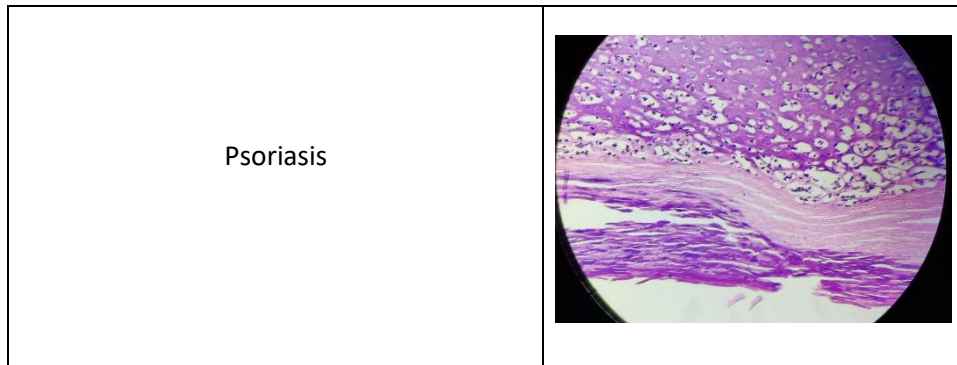
تصویر	سوال	پاسخ
	اندامی که در این تصویر نشان داده شده است چیست؟	مغز
	آیا این تصویر سونوگرافی است؟	بله
	چه ناهنجاری ای در تصویر مشاهده می شود؟	کیست

همچنین برای تشخیص و طبقه بندی بیماری های پوستی، همکاری ای با گروه تحقیقاتی دانشگاه علوم پزشکی تهران انجام شد بدین صورت که داده های پوستی توسط این گروه جمع آوری و در اختیار گروه هسته قرار گرفت تا آزمایش ها و پیاده سازی های لازم انجام گیرد. در جدول زیر دو نمونه از تصاویر بیماری های پوستی و طبقه بندی آن ها بر اساس الگو های موجود در تصاویر نشان داده شده است.

تصویر	بیماری
	Lichen planus

طرح شهید احمدی روشن (دوره ششم)

پرسش و پاسخ تصویری



تشخیص این بیماری ها بر اساس الگوهای موجود برای پزشکان و پرستاران کاری سخت است اما مدل پیاده سازی شده توانسته است با وجود تعداد کم تصاویر اولیه، به دقت حدودی ۸۲ درصد دست یابد. افزایش تعداد داده ها و الگوها می تواند در بهبود نتیجه بدست آمده تاثیرگذار باشد.

فصل سوم: فعالیت‌های انجام شده

شرح تفصیلی فعالیت‌های انجام شده در هسته مشتمل بر سرفصل‌های:

۳-۱. تقسیم کار فعالیت‌ها (شامل نحوه تقسیم‌بندی وظایف و جزئیات فعالیت هر یک از اعضای هسته در قالب جدول ذیل)؛

ردیف	نام و نام خانوادگی عضو	وظایف محول شده	فعالیت انجام شده
۱	محمدجواد پیرهادی	آموزش یک مدل به عنوان مدل پیش‌آموزش متن - تصویر	بررسی مدل‌های موجود - آزمایش مدل گرافی - آزمایش مدل OFA - به دست آوردن BPE برای زبان فارسی - آموزش مدل روی داده‌های فارسی (هنوز انجام نشده است)
۲	مطهره میرزایی	آموزش یک مدل به عنوان مدل پیش‌آموزش متن - تصویر	بررسی روش‌های جمع‌آوری داده - جمع‌آوری داده از طریق خزشگر (حدود ۹۰۰۰ داده) - پیش‌پردازش متون - بررسی مجموعه داده‌گان موجود فارسی جهت انجام فرآیند آموزش
۳	غزاله محمودی	مطالعه و تحقیق در رابطه با کارهای مرتبط انجام شده در موضوع اخبار جعلی و تشخیص موضع. بررسی کاربردهای مسئله در زندگی روزمره و تعریف سناریو. انجام پیاده‌سازی برای کاربردهای تعریف شده.	مطالعه و تحقیق در رابطه با کارهای مرتبط انجام شده در موضوع اخبار جعلی و تشخیص موضع. خواندن مقاله‌های مرتبط. بررسی کاربردهای مسئله در زندگی روزمره و تعریف سناریو. پیاده‌سازی برای کاربردهای تعریف شده. رفع اشکالات ربات تلگرام. انجام آزمایش‌های متعدد و بررسی مدل‌های یادگیری عمیق و یادگیری ماشین متعدد برای ارزیابی نتایج و رسیدن به بهترین مدل.
۴	ملیحه حاجی حسینی	مطالعه و تحقیق در حوزه پرسش و پاسخ تصویری در حوزه پزشکی، مدیریت گروه پزشکی طرح و گروه شرکت کننده در مسابقه ImageCLEF، پیگیری همکاری با گروه‌های فعال در حوزه پزشکی و هوش مصنوعی.	تهیه مجموعه داده دو زبانه (فارسی و انگلیسی) برای مسئله پرسش و پاسخ تصویری در حوزه پزشکی، پیگیری همکاری با دانشگاه علوم پزشکی تهران، انجمن مایا (فعال در حوزه هوش مصنوعی در پزشکی) دانشگاه شهید بهشتی، مرکز تحقیقات قلب و

طرح شهید احمدی روشن (دوره ششم)

پرسش و پاسخ تصویری

			عروق دانشگاه شهید بهشتی برای جمع آوری داده و تعریف پروژه مشترک. پیشبرد گروه شرکت کننده در مسابقه.
۵	ملیکا نوبختیان	فعالیت در گروه پزشکی طرح و پیاده سازی مدل ها، عضو گروه مسابقه ImageCLEF	پیاده سازی مدل تشخیص مفاهیم پزشکی، بررسی و جمع آوری داده برای پیاده سازی مدل از پیش آموزش دیده در حوزه استخراج ویژگی های مهم تصاویر پزشکی، شرکت در مسابقه جهانی ImageCLEF.
۶	یاسمن لطف اللهی	فعالیت در گروه پزشکی طرح و پیاده سازی مدل ها، عضو گروه مسابقه ImageCLEF.	پیاده سازی مدل تشخیص مفاهیم پزشکی، پیاده سازی مدل شرح نویسی تصاویر پزشکی به دو زبان فارسی و انگلیسی، شرکت در مسابقه جهانی ImageCLEF.
۷	غزل زمانی نژاد	آماده سازی بستری جمع آوری داده برای هر دو مسئله پرسش و پاسخ تصویری و شرح تصویر و پیاده سازی مدل تولید شرح تصویر	طراحی و پیاده سازی سایت برای سه سناریو مختلف برای جمع آوری داده، هماهنگی و آموزش برچسب زنانه برای جمع آوری داده، بازیابی داده جمع آوری شده توسط برچسب زنانه و اصلاح مشکلات احتمالی، پیاده سازی یک مدل برای مسئله شرح تصویری با استفاده از CLIP و GPT2، انجام آزمایش های متعدد بر روی مدل پیاده سازی شده و ارزیابی نتایج برای رسیدن به بهترین مدل، تحلیل خطای مدل جهت رفع مشکل
۸	شقایق مبشر	آماده سازی بستری جمع آوری داده برای هر دو مسئله پرسش و پاسخ تصویری و شرح تصویر و پیاده سازی مدل تولید شرح تصویر	طراحی و پیاده سازی سایت برای سه سناریو مختلف برای جمع آوری داده، هماهنگی و آموزش برچسب زنانه برای جمع آوری داده ، بازیابی داده جمع آوری شده توسط برچسب زنانه و اصلاح مشکلات احتمالی، پیاده سازی یک مدل برای مسئله شرح تصویری با استفاده از CLIP و GPT2، انجام آزمایش های متعدد بر روی مدل پیاده سازی شده و ارزیابی نتایج برای رسیدن به بهترین

طرح شهید احمدی روشن (دوره ششم)

پرسش و پاسخ تصویری

مدل، تحلیل خطای مدل جهت رفع مشکل			
<p>پیاده سازی اپلیکیشن ها</p> <p>تنظیم کردن وب سرور ها و SSL</p> <p>CI/CD</p> <p>مدیریت ساخت وبسایت آزمایشگاه</p> <p>نصب و تنظیم درایور ها / CUDA</p> <p>CUDNN و دیگر موارد مورد نیاز</p> <p>برای آموزش مدل ها</p> <p>ایجاد داکومنٹیشن برای روندها</p>	<p>پیگیری سرور های سخت افزاری مورد نیاز</p> <p>برای پیاده سازی مدل ها</p>	محمد مهدی جاوید	۹
<p>مطالعه و تحقیق در رابطه با کارهای مرتبط انجام شده در مسئله پرسش و پاسخ تصویری و خواندن مقاله های مرتبط، پیاده سازی مدل رمز گذار- رمز گشا جهت تولید پاسخ با دو رمز گذار مختلف LXMERT و VisualBERT و سه رمز گشا RNN، Attention RNN و Transformer، انجام آزمایش های متعدد بر روی مدل پیاده سازی شده و ارزیابی نتایج برای رسیدن به بهترین مدل، تحلیل خطای مدل جهت رفع مشکل</p>	<p>پیاده سازی مدل تولید پاسخ برای مسئله پرسش و پاسخ تصویری</p>	هادی شیخی	۱۰
<p>بررسی ابزار های مختلف برای استفاده بهتر از مدل ها، بررسی سمینار های موجود.</p>	بررسی روش های HCI	یاسمین مدنی	۱۱
<p>مطالعه و تحقیق در رابطه با کارهای مرتبط انجام شده در موضوع اخبار جعلی و تشخیص موضع. خواندن مقاله های مرتبط. بررسی کاربردهای مسئله در زندگی روزمره و تعریف سناریو. پیاده سازی برای کاربردهای تعریف شده. پیاده سازی ربات تلگرام. انجام آزمایش های متعدد و بررسی مدل های یادگیری عمیق و یادگیری ماشین متعدد برای ارزیابی نتایج و رسیدن به بهترین مدل.</p>	<p>مطالعه و تحقیق در رابطه با کارهای مرتبط انجام شده در موضوع اخبار جعلی و تشخیص موضع. بررسی کاربردهای مسئله در زندگی روزمره و تعریف سناریو. انجام پیاده سازی برای کاربردهای تعریف شده.</p>	بابک بهکام کیا	۱۲

طرح شهید احمدی روشن (دوره ششم)

پرسش و پاسخ تصویری

۱۳	بیان دیوانی آذر	طراحی رابط کاربری	طراحی رابط کاربری مدل های توسعه داده شده
۱۴	باوان دیوانی آذر	توسعه اپلیکیشن اندروید	توسعه اپلیکیشن اندروید دستیار نابینایان

۲-۳. مواد و روش های به کار گرفته شده؛

برای حل این مسائل از روش های یادگیری عمیق با ناظر استفاده شده است. مدل های مبتنی بر یادگیری عمیق توانایی خوبی در یادگیری ویژگی های مناسب دارند و می توانند در کیفیت مدل های ارائه شده تاثیر گذار باشند. البته مسئله مهم و تاثیر گذار در کیفیت مدل ها، داده های مسئله است. کیفیت و کمیت داده ها نقش به سزایی در کارایی مدل ها دارد از این رو جمع آوری و تهیه تعداد مناسبی داده از اصلی ترین گام ها در پیاده سازی مدل های هوش مصنوعی به شمار می رود. در ادامه به داده های جمع آوری شده در این هسته اشاره می شود.

۱. دستیار نابینایان:

در این قسمت یک مجموعه داده پرسش و پاسخ تصویری تهیه شده است که برای جمع آوری آن از دو روش استفاده شده است:

➤ بر اساس الگو: در این روش بر اساس ۱۱۰ الگو سوالات تولید شده اند سپس سوالات و تصاویر در اختیار افراد قرار گرفتند تا پاسخ مناسب و کوتاه شان درج شود. پاسخ های بلند هم از روی الگو تولید شده اند.

➤ بر اساس نیروی انسانی: در این بخش تصاویر در اختیار نیروهای انسانی قرار گرفتند و افراد سوالات و پاسخ های مناسب کوتاه و بلند برای تصاویر را درج کردند.

از مهم ترین ویژگی های این مجموعه داده، وجود سوالات کوتاه و بلند، پشتیبانی از زبان رسمی و محاوره و استفاده از تصاویر متناسب با فرهنگ ایران است.

علاوه بر جمع آوری مجموعه داده پرسش و پاسخ تصویری، یک مجموعه داده شرح نویسی تصاویر هم فراهم شده است که برای جمع آوری آن از دو روش استفاده شده است:

➤ استفاده از فضای وب: جمع آوری تصاویر و شرح آن ها از فضای وب که شامل حدود ۹۰۰۰ تصویر و شرح آن هاست.

➤ استفاده از نیروی انسانی: تعدادی تصاویر در اختیار نیروهای انسانی قرار داده شد که باید برای هر تصویر شرح مناسب را درج می کردند. برای هر تصویر چهار شرح توسط چهار شخص متفاوت درج شده است. در این بخش هم از تصاویر متناسب با فرهنگ ایرانی استفاده شد.

۲. تصاویر پزشکی:

در حوزه مسئل پزشکی، کمبود داده مناسب و کافی از گلوگاه های مهم محسوب می شود چرا که جمع آوری این نوع داده نیاز به تخصص پزشکی دارد و حتی در صورت وجود تخصص، به دلیل رعایت حریم خصوصی بیماران دسترسی به تصاویر پزشکی اغلب سخت و طاقت فرسا است. با این وجود و با توجه به اهمیت این حوزه تلاش هایی برای جمع آوری داده برای چند مسئله صورت گرفته است.

➤ ترجمه مجموعه داده تشخیص مفاهیم پزشکی به زبان فارسی از روی مجموعه داده مسابقه ImageCLEF.

➤ ترجمه مجموعه داده شرح نویسی تصاویر پزشکی به زبان فارسی از روی مجموعه داده مسابقه ImageCLEF.

➤ تولید مجموعه داده دو زبانه (انگلیسی-فارسی) مسئله پرسش و پاسخ تصویری در حوزه پزشکی شامل ۵۵۷۰۴۸ سوال و پاسخ پیرامون ۹۰۲۴۱ تصویر پزشکی.

➤ جمع آوری داده های مربوط به بیماری های پوستی Lichen planus و Psoriasis از طریق همکاری با دکتر واسعی از دانشگاه علوم پزشکی تهران.

با استفاده از این داده ها، مدل هایی توسعه داده شدند که توانستند به نتایج قابل قبولی در مسئله طرح شده دست یابند.

۳. اخبار جعلی: برای آموزش مدل صحت سنجی اخبار از مجموعه داده Sem Eval 2016 استفاده شد. ابتدا این مجموعه داده به زبان فارسی ترجمه شد سپس برای توسعه مدل مورد استفاده قرار گرفت.

۳-۳. آزمایش های انجام شده؛

۱. دستیار نابینایان: در این بخش یک مدل کدگزار-کدگشا برای مسئله پرسش و پاسخ تصویری پیاده سازی شده است. همچنین برای مسئله شرح نویسی تصاویر هم از معماری کدگزار-کدگشا استفاده شد که نتایج و نمونه های این مدل ها در قسمت های قبل ارائه شده است.

۲. **تصاویر پزشکی:** برای هر کدام از داده های معرفی شده در قسمت قبل مدل هایی توسعه داده شد که توانستند به نتایج خوبی دست یابند. در بخش تشخیص مفاهیم پزشکی به دقت حدودی ۳۵ درصد، در بخش شرح نویسی تصاویر پزشکی به امتیاز ۴۴.۸ درصد، در بخش مسئله پرسش و پاسخ تصویری در حوزه پزشکی به دقت حدود ۷۰ درصد و در بخش تشخیص بیماری های پوستی به دقت حدود ۸۲ درصد دست یافتیم.

۳. **اخبار جعلی:** در این بخش آزمایش های متعددی بر روی شبکه های BERT, BiCond, Adversarial صورت گرفت تا شبکه با بهترین نتایج انتخاب و مورد استفاده قرار گیرد.

۴-۳. چالش ها و موانع موجود؛

از مهم ترین چالش های پیش رو محققین حوزه هوش مصنوعی، عدم دسترسی به داده های مناسب و با کیفیت زبان فارسی است. اغلب پژوهش های انجام شده به زبان انگلیسی هستند که داده های مناسبی برایشان جمع آوری شده است. عدم وجود داده مناسب روند پیاده سازی مدل های هوش مصنوعی را با مشکل مواجه می کند. از دیگر چالش های پیش رو، عدم وجود بستر های سخت افزاری مناسب در ایران است. پیاده سازی مدل های هوش مصنوعی نیازمند سخت افزار های خاصی هستند که اغلب به دلیل هزینه های بالا به راحتی در دسترس قرار نمی گیرند. وجود اعتبار لبزنت و اختصاص آن به دانشجویان طرح تا حدی این مشکل را حل کرد و امکان دسترسی به منابع سخت افزاری مناسب و قوی را فراهم کرد.

۵-۳. دستاوردهای هسته؛

۱. **دستیار نابینایان:** در این پژوهش برای اولین بار بنچمارک مسئله پرسش و پاسخ تصویری و شرح نویسی تصاویر در زبان فارسی منتشر شد. از دیگر دستاوردهای این بخش توسعه اپلیکیشن هوشمند برای افراد نابینا هست که می توانند شرح تصاویر و پاسخ سوالات شان پیرامون تصاویر را دریافت کنند. همچنین مدل های پیاده سازی شده شرح نویسی تصاویر و مسئله پرسش و پاسخ تصویری به زبان فارسی برای اولین بار به صورت دمو در دسترس قرار گرفته اند و قابلیت استفاده دارند.

۲. **تصاویر پزشکی:** در این بخش ابزاری برای استفاده آسان کاربران از مدل های توسعه داده شده که در قسمت قبل معرفی شدند، طراحی شد. همچنین مجموعه داده های مناسبی در چند شاخه پزشکی جمع آوری شد که می تواند زمینه ساز تحقیقات بعدی در این مسائل باشد. همچنین مدل توسعه داده شده برای مسئله شرح نویسی تصاویر پزشکی توانست رتبه نخست مسابقات جهانی ImageCLEF در شاخه

حاشیه نویسی تصاویر پزشکی را کسب کند. بر اساس شواهد موجود، این اولین بار است که گروهی از ایران در این مسابقات حضور می یابند و موفق به کسب رتبه نخست می شوند. در مقاله [IUST NLPLAB at ImageCLEFmedical Caption Tasks 2022](#) جزییات مدل های پیاده سازی شده برای این مسئله ارائه شده است. همچنین به واسطه موفقیت در این مسابقه، نماینده ای از گروه، به کنفرانس CLEF کشور ایتالیا دعوت شد و به ارائه دست آورد های علمی گروه پرداخت.

۳. اخبار جعلی:

در این پروژه، ما یک [ربات تلگرام](#) توسعه داده ایم تا هر کسی بتواند به راحتی موضع هر خبری که در تلگرام دیده است را نسبت به یک موضوع دلخواه بدست آورد. از این ربات در صنعت نیز می توان استفاده کرد، به این صورت که می توان دیدگاه مشتریان را به صورت اتوماتیک ارزیابی کرد و محصول را بهبود داد. این ربات قابلیت جست و جو متن خبر و تصویر را در جهت صحت سنجی با رویکرد multi-modal دارد. تشخیص موضع متن ارسال شده در رابطه با موضوع مدنظر از دیگر کارکردهای این ربات می باشد.

۳-۶. نوآوری های صورت گرفته؛

از جمله نوآوری های صورت گرفته، تمرکز بر روی پیاده سازی مدل های هوشمند به زبان فارسی است. در حوزه زبان انگلیسی مدل ها و برنامه های متفاوتی توسعه داده شده اند که نیاز است در حوزه زبان فارسی هم اقداماتی صورت گیرد. در این هسته مدل ها و برنامه های متفاوتی از جمله اپلیکیشن نابینایان، مدل پرسش و پاسخ تصویری در دو حوزه عمومی و پزشکی، مدل شرح نویسی تصاویر در دو حوزه عمومی و پزشکی و همچنین توسعه بات تلگرام برای صحن سنجی اخبار به زبان فارسی توسعه داده شدند که می تواند گام مهمی در پیشرفت هوش مصنوعی در زبان فارسی باشد. همچنین در این هسته اولین ربات تلگرامی دوزبانه برای تشخیص موضع و جستجو متن و یا عکس خبر در جهت آگاهی بیشتر و جلوگیری از باور اخبار جعلی توسعه داده شد.

۳-۷. تفاوت این کار با پژوهش ها / محصولات پیشین.

براساس تحقیقات صورت گرفته، مدل های پیاده سازی شده در این هسته برای اولین بار هست که به زبان فارسی توسعه داده شده اند و این می تواند زمینه ساز مناسبی برای تحقیقات بعدی و پیشرفت های موثر در این حوزه در کشور باشد.

فصل چهارم: مطالعه بازار / مشتریان / ذی نفعان

شرح تفصیلی تلاش‌های گروه برای تجاری‌سازی (به نتیجه رساندن) دستاوردهای هسته مشتمل بر سرفصل‌های:

۴-۱. بررسی بازار / محصولات؛

باتوجه به گستردگی مدل‌های توسعه داده شده در هسته، محصولات متنوعی تولید شده‌اند که هر کدام می‌توانند بازار مخصوص به خود را داشته باشند. اپلیکیشن نابینایان می‌تواند برنامه خوبی برای افراد نابینا یا کم بینا جامعه باشد. همچنین مدل‌های پزشکی توسعه داده شده می‌توانند مورد استفاده بیماران، پزشکان و پرستاران قرار گیرند و باعث صرفه جویی در وقت و هزینه افراد شوند. براساس بررسی‌های صورت گرفته هیچ کدام از این برنامه‌ها در حال حاضر در ایران موجود نمی‌باشند و از این جهت می‌توانند بازار مناسبی را برای محصولات شان فراهم کنند.

۴-۲. نتایج و گزارش مذاکره با مشتریان / ذی نفعان دستاوردهای هسته؛

در بخش‌های مختلف تلاش شده است تا حد امکان مذاکراتی با گروه‌ها و ذی نفعان صورت بگیرد. به عنوان مثال انجمن نابینایان از جمله نهاد‌هایی است که می‌تواند در توسعه اپلیکیشن نابینایان نقش داشته باشد. در بخش پزشکی هم تلاش شده است تا حد امکان با گروه‌های پزشکی متفاوتی صحبت شود تا مدل‌های توسعه داده شده مورد استفاده قرار بگیرند. در زمینه بیماری‌های پوستی همکاری خوبی با دانشگاه علوم پزشکی تهران صورت گرفته است تا در آینده امکان استفاده از مدل تشخیص بیماری‌های پوستی جهت تشخیص سریع تر و دقیق تر برای پرستاران فراهم شود.

۴-۳. دسته‌بندی مشتریان؛

- جامعه نابینایان و کم بینایان کشور
- جامعه پزشکی و پرستاری
- بیماران و افراد مشکوک به بیماری
- کارآفرین‌ها و صاحبان کسب و کار

۴-۴. مدل کسب و کار هسته (Business Model)؛

۴-۵. ارزیابی وضعیت رقبا و مزیت رقابتی گروه؛

طرح شهید احمدی روشن (دوره ششم)

پرسش و پاسخ تصویری

➤ براساس بررسی های موجود محصول مشابهی با مدل های توسعه داده شده در هسته در کشور موجود نمی باشد و از این حیث این مدل ها می توانند شروع کننده رقابتی باشند که در ادامه توسط گروه های دیگر تکمیل می شود.

۴-۶. طرح تجاری کسب و کار (BP).

در بحث کسب و کار تلاش شده است تا کارکرد مدل ها در قالب تیزرهای مناسبی تهیه شود تا صاحبان کسب و کار بتوانند بهتر با اهداف و کارکرد های این مدل ها آشنا شوند. همچنین تلاش شده است با شرکت هایی مثل دادماتک که در زمینه هوش مصنوعی فعالیت دارند صحبت شود تا برای همکاری های بعدی و استفاده از این برنامه ها برنامه ریزی شود.

فصل پنجم: جمع بندی

جمع بندی فعالیت های هسته مشتمل بر سرفصل های:

۵-۱. نتایج فعالیت های هسته؛

از جمله خروجی های مناسب هسته، یادگیری کار گروهی و تعامل مناسب اعضای هسته برای رسیدن به هدف است. اعضای طرح در تمام مدت با برگزاری جلسات متعدد روند کار را پیگیری می کردند که علاوه بر پیشرفت کارها و پروژه ها، تعامل و کار گروهی زیر نظر استاد خبره را هم یاد گرفتند. همچنین مدل ها و برنامه های مناسبی که تا کنون در زبان فارسی موجود نبودند توسط گروه تولید شدند که این می تواند زمینه ساز تحقیقات آینده و پیشرفت هوش مصنوعی در کشور باشد.

۵-۲. اهداف محقق شده؛

در هر سه بخش دستیار نابینایان، تصاویر پزشکی و اخبار جعلی مدل ها و برنامه های مناسبی توسعه داده شدند که هم اکنون قابلیت استفاده را دارند.

۵-۳. اهداف محقق نشده؛

آموزش مدل های multi-modal در زمینه عمومی و پزشکی از دیگر اهداف هسته بود که به دلیل وجود برخی مشکلات نرم افزاری و سخت افزاری تاکنون آماده نشده است اما تلاش برای تحقق این هدف در آینده ادامه خواهد داشت.

۵-۴. برنامه برای رسیدن به اهداف محقق نشده؛

با وجود اتمام زمان هسته، تلاش ها و پیگیری ها برای پیشرفت کارهای باقی مانده و بهبود مدل های توسعه داده شده ادامه خواهد داشت تا نتایج تلاش تمام اعضای طرح محقق شود.

۵-۵. برنامه پیش روی هسته؛

- تلاش برای تحقق اهداف محقق نشده
- بهبود کیفیت مدل های توسعه داده شده
- همکاری با نهاد ها و گروه های مرتبط با برنامه های توسعه داده شده

۵-۶. هدف گروه؛

طرح شهید احمدی روشن (دوره ششم)

پرسش و پاسخ تصویری

پیشرفت مدل های توسعه داده شده از طریق تعریف موضوعات تحقیقاتی مرتبط با طرح و همکاری با گروه ها و نهاد های مرتبط.