دانشکده علوم و فنون نوین

گروه بین رشته ای فناوری (بخش علوم و فناوری شبکه)

گزارش مرحله دوم پروژه

احراز هویت غیرحضوری متقاضیان خدمات الکترونیک انتظامی بر مبنای سنجه‌های بیومتریکی

توسط:

هادی ویسی

آذر ۱۴۰۰

فهرست

خلاصه اجرایی **۷**

فصل ۱ پیاده‌سازی سامانه تشخیص زنده بودن و تطبیق چهره **۱۰**

۱- ۱ مکان‌یابی چهره در تصویر ۱۱

۱- ۲ سنجش کیفیت و نور در تصویر ۱۲

۱- ۳ تطبیق چهره ۱۳

۱- ۴ تشخیص زنده بودن غیرتعاملی (Passive) ۱۳

۱- ۵ تشخیص زنده بودن تعاملی مبتنی بر تشخیص گفتار ۱۴

۱- ۶ تشخیص زنده بودن تعاملی مبتنی بر پلک زدن ۱۶

۱- ۷ ماژول‌های ترکیبی ۱۷

۱ -۷ -۱ تشخیص زنده بودن تعاملی (گفتار/پلک زدن) و غیرتعاملی ۱۷

۱ -۷ -۲ تشخیص زنده بودن تعاملی (گفتار/پلک زدن) و غیرتعاملی و تطبیق چهره با تصویر مرجع ۱۸

۱ -۷ -۳ تشخیص زنده بودن غیرتعاملی با ویدئو و تطبیق چهره با تصویر مرجع ۱۸

۱ -۷ -۴ تشخیص زنده بودن غیرتعاملی با یک تصویر و تطبیق چهره با تصویر مرجع ۱۸

فصل ۲ تست و ارزیابی سامانه **۱۹**

۲- ۱ ارزیابی سامانه مکان‌یابی چهره ۱۹

۲- ۲ ارزیابی سامانه تطبیق چهره ۲۱

۲ -۲ -۱ دادگان‌های ارزیابی تطبیق چهره ۲۱

۲ -۲ -۲ نتایج ارزیابی ۲۲

۲- ۳ ارزیابی سامانه تشخیص زنده بودن ۲۵

۲ -۳ -۱ تشخیص زنده بودن غیرتعاملی ۲۵

۲ -۳ -۲ تشخیص زنده بودن تعاملی ۲۷

فصل ۳ نحوه‌ی استفاده از APIها و ماژول مدیریت و دسترسی کاربران **۲۹**

۳- ۱ ساختار پروژه ۲۹

۳- ۲ دسترسی به ماژول‌های هسته‌ی مرکزی ۳۰

۳ -۲ -۱ ویژگیهای مشترک سرویسها ۳۱

۳ -۲ -۲ تطبیق چهره دو تصویر ۳۳
۳ -۲ -۳ تشخیص زنده بودن (غیرتعاملی) ۳۴
۳ -۲ -۴ تشخیص زنده بودن (تعاملی: پلک زدن) ۳۵
۳ -۲ -۵ سرویس تشخیص زنده بودن (تعاملی: تشخیص گفتار) ۳۶
۳ -۲ -۶ تطبیق چهره و تشخیص زنده بودن (غیرتعاملی) ۳۸
۳ -۲ -۷ تطبیق چهره و تشخیص زنده بودن تعاملی (تشخیص گفتار) ۳۹
۳ -۲ -۸ تطبیق چهره و تشخیص زنده بودن تعاملی (پلک زدن) ۴۱
۳- ۳ کدهای خطا ۴۳
۳- ۴ استعلام از سامانه‌های دیگر ۴۶

فصل ۴ نصب سرویس‌ها در سرور کارفرما ۴۷
۴- ۱ نصب و راه‌اندازی سرویس‌ها ۴۷
۴- ۲ گزارش انجام کار ۴۸
۴ -۲ -۱ اولین جلسه (۲۶ – ۴ - ۱۴۰۰) ۴۸
۴ -۲ -۲ دومین جلسه (۱۱ - ۷ - ۱۴۰۰) ۴۸
۴ -۲ -۳ سومین جلسه (۲۶ – ۷ - ۱۴۰۰) ۴۸
۴ -۲ -۴ چهارمین جلسه (۲۸ – ۷ - ۱۴۰۰) ۴۸
۴ -۲ -۵ پنجمین جلسه (۲۴ – ۰۸ - ۱۴۰۰) ۴۹

فهرست شکل‌ها

شکل ۱–۱ نمای کلی زیربخش‌های سامانه احراز هویت غیرحضوری ۱۰
شکل ۱–۲ نمونه تصاویر با کیفیت و نور نامناسب که توسط ماژول سنجش کیفیت سامانه تشخیص داده شده‌اند ۱۲
شکل ۱–۳ نمونه تصاویر تقلبی ۱۴
شکل ۱–۴ نمایی از روند تشخیص زنده بودن غیرتعاملی ۱۴
شکل ۱–۵ نمونه‌هایی از جملات مورد استفاده در تشخیص زنده بودن تعاملی با استفاده از تشخیص گفتار ۱۵
شکل ۱–۶ ساختار تشخیص زنده بودن تعاملی با استفاده از تشخیص گفتار ۱۶
شکل ۱–۷ ساختار تشخیص زنده بودن تعاملی با استفاده از تشخیص پلک زدن ۱۷
شکل ۲–۱ چند نمونه از تصاویر استفاده شده برای آزمایش‌های اولیه الگوریتم‌های مکان‌یابی چهره ۲۰
شکل ۲–۲ نمونه‌ای از فریم‌های استخراج شده از دادگان SIW ۲۱
شکل ۲–۳ نمونه‌ای از تصاویر LFW ۲۲
شکل ۲–۴ نمونه‌ای از تصاویر IRANCELEB ۲۲
شکل ۲–۵ نمودار DET مقایسه سه مدل تطابق چهره بر روی LFW پاک‌سازی شده ۲۳
شکل ۲–۶ نمودار DET مقایسه سه مدل تطابق چهره بر روی IRANCELEB ۲۴
شکل ۲–۷ نمونه‌ای از تصاویر زنده‌ی استفاده شده برای ارزیابی تشخیص زنده بودن غیرتعاملی ۲۶
شکل ۲–۸ نمونه‌ای از تصاویر جعلی استفاده شده برای ارزیابی تشخیص زنده بودن غیرتعاملی ۲۶

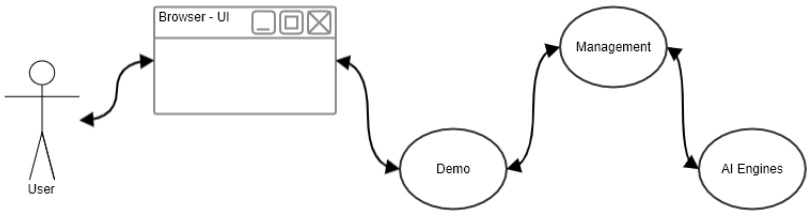
فهرست جدول‌ها

جدول ۲–۱ میانگین درصد اطمینان درستی و زمان برای پردازش هر تصویر ۲۰
جدول ۲–۲ ارزیابی SepidSystem-Face-Verification-V3 بر روی مجموعه‌داده‌ها ۲۴
جدول ۲–۳ مقدار EUC و EER مربوط به مدل‌های تطبیق چهره بر روی مجموعه‌داده‌های مختلف ۲۴
جدول ۲–۴ مقدار دقت (Accuracy) مربوط به مدل‌های تطبیق چهره بر روی مجموعه‌داده‌های مختلف ۲۵
جدول ۲–۵ اطلاعات دادگان استفاده شده برای ارزیابی تشخیص زنده بودن غیرتعاملی (دادگان نامتوازن) ۲۵
جدول ۲–۶ نتایج تشخیص زنده بودن غیرتعاملی بر روی مجموعه داده‌ی نامتوازن ۲۵
جدول ۲–۷ اطلاعات دادگان استفاده شده برای ارزیابی تشخیص زنده بودن غیرتعاملی (دادگان متوازن) ۲۶
جدول ۲–۸ نتایج تشخیص زنده بودن غیرتعاملی بر روی مجموعه داده‌ی متوازن ۲۷
جدول ۲–۹ ارزیابی سرویس تشخیص گفتار با معیار نرخ خطای کلمه ۲۷
جدول ۲–۱۰ ارزیابی سرویس تشخیص پلک زدن با معیار دقت (Accuracy) ۲۸
جدول ۳–۱ سطوح حساسیت در تشخیص و سخت‌گیری در تصمیم‌گیری ۳۱
جدول ۳–۲ جزئیات مربوط به وضعیت خروجی و نتیجه ۳۱
جدول ۳–۳ فیلدهای خروجی ناموفق - با بدنه Form-Data ۳۲
جدول ۳–۴ شرح فیلدهای شی Error ۳۲
جدول ۳–۵ جزئیات نحوه فراخوانی API نطبیق چهره – با بدنه Form-Data ۳۳
جدول ۳–۶ ورودی های سرویس تطبیق چهره - با بدنه Form-Data ۳۳
جدول ۳–۷ جزئیات نحوه فراخوانی API نطبیق چهره – با بدنه JSON ۳۳
جدول ۳–۸ ورودی های سرویس تطبیق چهره - با بدنه JSON ۳۳
جدول ۳–۹ خروجی موفق سرویس تطبیق چهره ۳۳
جدول ۳–۱۰ جزئیات نحوه فراخوانی API تشخیص زنده بودن (غیرتعاملی)– با بدنهForm-Data ۳۴
جدول ۳–۱۱ ورودی های سرویس تشخیص زنده بودن (غیرتعاملی) - با بدنهForm-Data ۳۴
جدول ۳–۱۲ جزئیات نحوه فراخوانی API تشخیص زنده بودن (غیرتعاملی)– با بدنه JSON ۳۴
جدول ۳–۱۳ ورودی های سرویس تشخیص زنده بودن (غیرتعاملی) - با بدنهJSON ۳۴
جدول ۳–۱۴ خروجی موفق سرویس تشخیص زنده بودن (غیرتعاملی) ۳۴
جدول ۳–۱۵ جزئیات فراخوانی سرویس دریافت الگو در تشخیص زنده بودن با پلک زدن ۳۵
جدول ۳–۱۶ خروجی موفق سرویس دریافت الگو در تشخیص زنده بودن با پلک زدن ۳۵
جدول ۳–۱۷ جزئیات نحوه فراخوانی سرویس تشخیص زنده بودن پلک زدن – با بدنه Form-Data ۳۵
جدول ۳–۱۸ورودی های سرویس تشخیص زنده بودن پلک زدن- با بدنه Form-Data ۳۶
جدول ۳–۱۹ خروجی موفق سرویس تشخیص زنده بودن پلک زدن ۳۶
جدول ۳–۲۰ جزئیات فراخوانی سرویس دریافت الگو در تشخیص زنده بودن با گفتار ۳۶
جدول ۳–۲۱ خروجی موفق سرویس دریافت الگو در تشخیص زنده بودن با گفتار ۳۶
جدول ۳–۲۲ جزئیات نحوه فراخوانی سرویس تشخیص زنده بودن تعاملی با تشخیص گفتار– با بدنه Form-Data ۳۷
جدول ۳–۲۳ ورودی های سرویس تشخیص زنده بودن تعاملی با تشخیص گفتار- با بدنهForm-Data ۳۷
جدول ۳–۲۴ خروجی موفق سرویس تشخیص زنده بودن تعاملی با تشخیص گفتار ۳۷
جدول ۳–۲۵ جزئیات نحوه فراخوانی سرویس تطبیق چهره و تشخیص زنده بودن (غیرتعاملی)– با بدنه Form-Data ۳۸
جدول ۳–۲۶ ورودی های سرویس تطبیق چهره و تشخیص زنده بودن (غیرتعاملی)- با بدنه Form-Data ۳۸

جدول ۳–۲۷ جزئیات نحوه فراخوانی سرویس تطبیق چهره و تشخیص زنده بودن (غیرتعاملی)– با بدنهJSON ۳۸
جدول ۳–۲۸ ورودی های سرویس تطبیق چهره و تشخیص زنده بودن (غیرتعاملی)- با بدنهJSON ۳۸
جدول ۳–۲۹ خروجی موفق سرویس تطبیق چهره و تشخیص زنده بودن (غیرتعاملی) ۳۸
جدول ۳–۳۰ جزئیات فراخوانی سرویس دریافت الگوی تشخیص گفتار ۳۹
جدول ۳–۳۱ قالب خروجی موفق سرویس دریافت الگوی تشخیص گفتار ۴۰
جدول ۳–۳۲ جزئیات نحوه فراخوانی سرویس تطبیق چهره و تشخیص زنده بودن تعاملی تشخیص گفتار ۴۰
جدول ۳–۳۳ رودی‌های سرویس تطبیق چهره و تشخیص زنده بودن تعاملی تشخیص گفتار (بدنه FORM-DATA) ۴۰
جدول ۳–۳۴ خروجی موفق سرویس تطبیق چهره و تشخیص زنده بودن تعاملی تشخیص گفتار ۴۱
جدول ۳–۳۵ جزئیات فراخوانی سرویس دریافت الگوی پلک زدن ۴۲
جدول ۳–۳۶ قالب خروجی موفق سرویس دریافت الگوی پلک زدن ۴۲
جدول ۳–۳۷ جزئیات نحوه فراخوانی سرویس تطبیق چهره و تشخیص زنده بودن تعاملی پلک زدن ۴۲
جدول ۳–۳۸ ورودی‌های سرویس تطبیق چهره و تشخیص زنده بودن تعاملی پلک زدن (بدنه FORM-DATA) ۴۲
جدول ۳–۳۹ خروجی موفق سرویس تطبیق چهره و تشخیص زنده بودن تعاملی پلک زدن ۴۳
جدول ۳–۴۰ کد خطاها و جزئیات خطاهای عمومی ۴۴
جدول ۳–۴۱ کد خطاهای تطبیق چهره و تشخیص زنده بودن غیرتعاملی ۴۵
جدول ۳–۴۲ کد خطاهای تشخیص زنده بودن تعاملی ۴۵

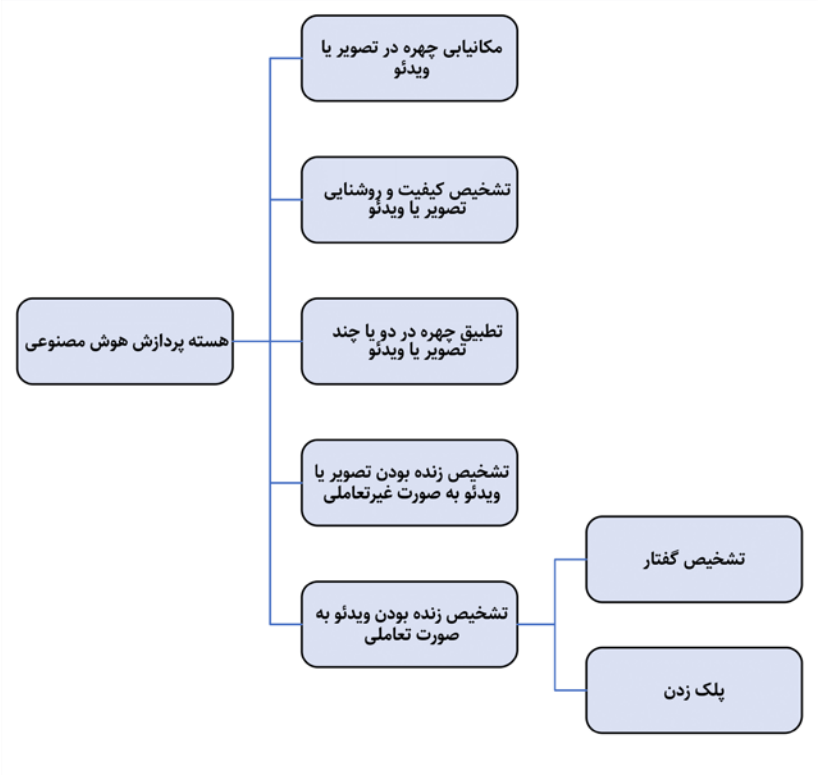
خلاصه اجرایی

سامانه احراز هویت غیرحضوری موضوع قرارداد در این مرحله، روی سرورهای ناجا نصب و راه‌اندازی شد و سرویس‌های آن مورد ارزیابی قرار گرفت. این سامانه شامل سه لایه هوش مصنوعی، مدیریت و دمو است که درخواست‌ها توسط لایه دمو دریافت شده و پس از بررسی توسط لایه مدیریت (برای مدیریت درخواست‌ها و کنترل دسترسی)، به لایه هوش مصنوعی برای پردازش اصلی ارسال می‌شوند. لایه پردازش هوش مصنوعی با بهره‌گیری از مدل‌های یادگیری عمیق، فعالیت‌های پردازش تصویر و گفتار را انجام می‌دهد و نتیجه برگردانده می‌شود. در شکل ۱ نمای کلی از این سامانه دیده می‌شود.



شکل ۱ نمای کلی زیربخش‌های سامانه احراز هویت غیرحضوری

وظایف هسته پردازش هوش مصنوعی که کار اصلی احراز هویت بر عهده آن است، را می‌توان در پنج دسته قرار داد که این دسته‌ها در شکل ۲ آورده شده است.



شکل ۲ وظایف هسته پردازش هوش مصنوعی

هر کدام از سرویس‌ها و ماژول‌های پیاده‌سازی شده در گزارش به طور کامل توضیح داده شده است. ماژول‌های مکان‌یابی چهره، تطبیق چهره و تشخیص زنده بودن به وسیله‌ی دادگان معتبر در این حوزه مورد آزمایش و ارزیابی قرار گرفته است و نتایج آن‌ها به همراه معرفی دادگانِ مورد آزمایش در فصل دوم آورده شده است. در جدول ۱ اطلاعات ارزیابی سرویس‌های هوش مصنوعی سامانه برای مدل نهایی استفاده شده در پروژه به طور خلاصه آورده شده است.

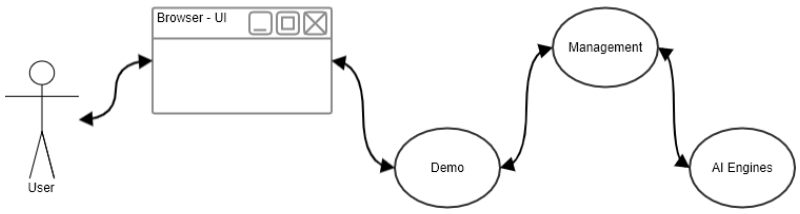
جدول ۱ نتایج ارزیابی سرویس‌های هوش مصنوعی

سرویس تطبیق چهره	EER	دقت
SepidSystem-Face-Verification-V3	0.003289	0.9950
سرویس تشخیص زنده بودن (غیر تعاملی)	**EER**	**دقت**
SepidSystem-Passive-Liveness-V1*V2 (Fusion2)	0.1714	0.847
سرویس تشخیص زنده بودن (تعاملی – گفتار)	**WER**	
بر روی جملات eKYC	0.019	
سرویس تشخیص زنده بودن (تعاملی – پلک زدن)	**دقت**	
بر روی ۴۱۵۸۷ فریم از ویدئوها	99.1	

در این گزارش، ساختار سامانه به طور کامل توضیح داده می‌شود و لایه‌های پروژه مورد بررسی قرار خواهد گرفت. سرویس‌ها و نحوه‌ی دسترسی آن‌ها با جزییات پیاده‌سازی ذکر می‌گردد. نتایج ارزیابی سرویس‌های هوش مصنوعی در یک فصل مجزا به تفصیل شرح داده می‌شود و در نهایت نصب سرویس‌ها در سرور کارفرما و گزارش جزئیات کار (در نصب سه نسخه مختلف از سامانه) بیان شده است.

فصل ۱ پیاده‌سازی سامانه تشخیص زنده بودن و تطبیق چهره

سامانه احراز هویت غیرحضوری این پروژه شامل دو سرویس اصلی تطبیق چهره[1] و تشخیص زنده بودن[2] است که پیاده‌سازی پروژه آن شامل یک هسته مرکزی پردازش هوش مصنوعی، یک لایه مدیریت دسترسی و یک لایه دمو است. درخواست‌های احراز هویت توسط لایه دمو از کاربر دریافت شده و پس از بررسی اولیه توسط لایه مدیریت، در صورت دارا بودن شرایط اولیه (از جمله رعایت سطح دسترسی)، به لایه هسته پردازشی مرکزی (هوش مصنوعی) ارسال می‌شوند. پردازش مرکزی با بهره‌گیری از مدل‌های یادگیری عمیق، فعالیت‌های پردازش تصویر و گفتار را انجام می‌دهد. در شکل ۱–۱ نمای کلی از این سامانه دیده می‌شود.



شکل ۱–۱ نمای کلی زیربخش‌های سامانه احراز هویت غیرحضوری

در این بخش، هسته پردازش هوش مصنوعی این سامانه مورد بررسی قرار می‌گیرد. وظایف و زیرماژول‌های این بخش را می‌توان در پنج دسته قرار داد:

[1] Face Matching (Face Verification)
[2] Liveness Detection

- مکان‌یابی چهره[1] در تصویر یا ویدئو و تعیین نقاط کلیدی چهره[2]
- تشخیص کیفیت و روشنایی تصویر یا ویدئو
- تطبیق چهره در دو یا چند تصویر یا ویدئو
- تشخیص زنده بودن تصویر یا ویدئو به صورت غیرتعاملی یا passive (با بررسی‌هایی مانند آنالیز کیفیت، روشنایی، ویژگی‌های رنگی، بازتاب نور و پیش‌بینی تصویر عمقی)
- تشخیص زنده بودن ویدئو به صورت تعاملی یا active (بررسی انجام‌شدن تعامل خواسته‌شده از کاربر) که خود شامل حالت‌های زیر است:
  - پلک زدن
  - تشخیص گفتار

سرویس‌های این سامانه معمولا با بهره‌گیری از چند ماژول به‌صورت همزمان پاسخ درخواست ارسالی را آماده می‌کند. در بخش‌های آینده، هریک از این دسته‌ها و همین‌طور ماژول‌های ترکیبی مورد بررسی قرار می‌گیرد.

۱-۱ مکان‌یابی چهره در تصویر

مکان‌یابی چهره یک فناوری رایانه‌ای مبتنی بر هوش مصنوعی است که برای یافتن و شناسایی چهره انسان در تصاویر دیجیتال استفاده می‌شود. فناوری مکان‌یابی چهره را می‌توان در زمینه‌های مختلف، از جمله امنیت و زیست‌سنجی به کار برد تا نظارت و ردیابی افراد را در زمان واقعی ارائه دهد. مکان‌یابی چهره از تکنیک‌های پایه‌ای بینایی ماشین تا شبکه‌های عصبی مصنوعی پیچیده، برای پیدا کردن مکان چهره بهره گرفته است و اکنون نقش مهمی را به عنوان اولین گام در بسیاری از برنامه‌های کلیدی ایفا می‌کند، از جمله ردیابی چهره، تجزیه و تحلیل چهره و بازشناسی چهره. مکان‌یابی چهره تأثیر قابل توجهی بر نحوه انجام عملیات در این برنامه‌ها را دارد.

پیشرفت‌های عمده در روش مکان‌یابی چهره در سال ۲۰۰۱ رخ داد، چارچوب Viola-Jones یکی از معروف‌ترین آن‌ها می‌باشد. برای بهبود تشخیص چهره، الگوریتم‌های دیگری، مانند شبکه عصبی R-CNN وSSD برای کمک به بهبود فرآیندها توسعه یافته‌اند. لازم به ذکر است که در سال‌های اخیر با توجه به اهمیت فناوری زیست‌سنجی در مسائل امنیتی و تجاری، کارهای قابل توجهی بر روی زیست‌سنجی چهره توسط محققان انجام شده است. همانطور که بیان شد به دلیل اهمیت کاربرد مکان‌یابی چهره، این فناوری نیز پیشرفت‌های قابل ملاحظه‌ای داشته است. با توجه به تحقیقات صورت گرفته بهترین مدل‌ها برای تشخیص چهره انتخاب و بهبود یافته‌اند.

با توجه به این که مکان‌یابی چهره در تصویر تقریباً در تمامی سرویس‌های دیگر سامانه احراز هویت غیرحضوری در این پروژه مورد استفاده قرار می‌گیرد، این بخش به عنوان یکی از بخش‌های پیش‌نیاز در هسته هوش مصنوعی محسوب می‌شود. این سامانه در حال حاضر مجهز به چهار مدل مکان‌یابی چهره است که سه مورد از آن‌ها به صورت فعال در سرویس‌های مختلف مورد استفاده قرار می‌گیرند:

- SepidSystem-FaceDetection-V1: این مدل برای کاربردهایی که در آن تصویر ورودی کنترل‌شده باشد قابل قبول بوده و به لطف پیچیدگی پردازشی پایین آن، سرعت بسیار بالایی دارد. این مدل با بهره‌گیری از ویژگی‌های HOG تصویر، مکان چهره‌ها در تصویر را محاسبه می‌کند.

[1] Face Detection
[2] Facial Landmarks

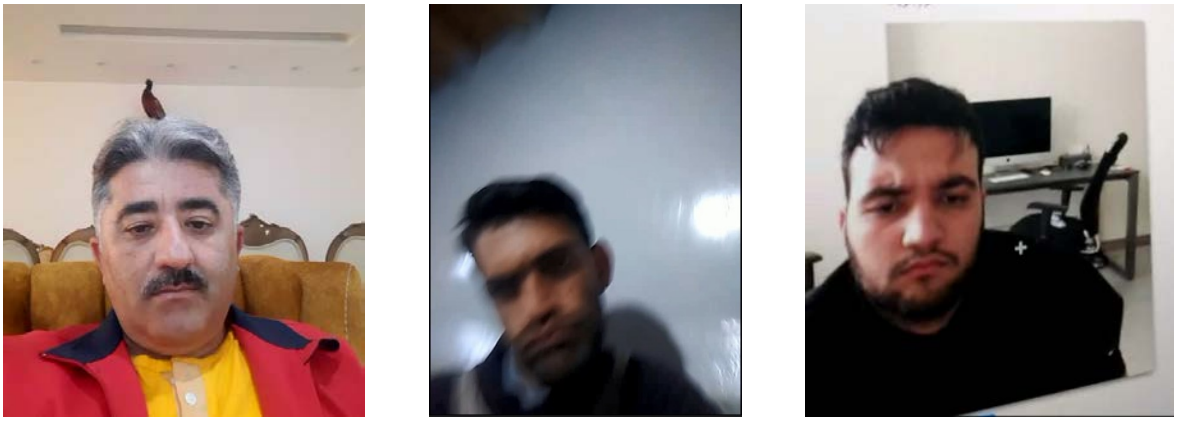
- SepidSystem-FaceDetection-V2: این مدل برای استخراج ویژگی از تصویر از معماری شبکه عصبی عمیق VGG-16 بهره می‌برد و بر روی دادگان AFLW آموزش دیده است. با اضافه شدن مدل‌های مکان‌یابی چهره بهتر، در حال حاضر در سرویس‌های ارائه‌شده به‌کار نمی‌رود.
- SepidSystem-FaceDetection-V3: این مدل با داشتن سرعت پردازش نسبتا بالا، دقت مناسبی را نیز ارائه می‌دهد و در اکثر موارد گزینه مناسبی برای استفاده در کاربردهای احراز هویت از راه دور است. این مدل بر پایه شبکه عصبی و از لایه‌های پیچشی سبک برای استخراج ویژگی استفاده شده است.
- SepidSystem-FaceDetection-V4: دقت مکان‌یابی چهره این مدل بسیار بالا بوده ولی نسبت به دو مدل قبلی پیچیدگی محاسباتی بیشتری داشته و سرعت کمتری دارد. در این مدل استخراج ویژگی با استفاده از مدل پایه ResNet-50 صورت می‌گیرد و در مواردی که تصویر یا ویدئو ورودی کنترل‌شده نباشد، می‌توان آن را مورد استفاده قرار داد.

در حال حاضر از مدل SepidSystem-FaceDetection-V4 در سرویس تشخیص چهره، از SepidSystem-FaceDetection-V3 در سرویس‌های تشخیص زنده بودن و تطبیق چهره و از SepidSystem-FaceDetection-V1 در سرویس‌های پلک زدن و لب‌خوانی استفاده می‌شود. نتایج ارزیابی و مقایسه این مدل‌ها در فصل دو مورد بررسی قرار خواهد گرفت.

بنا بر نوع تصاویر ورودی سامانه، دقت موردنیاز و توان سخت‌افزاری بستری که سامانه در آن اجرا می‌شود، می‌توان ماژول‌های مختلفی را برای استفاده در سرویس‌ها فعال‌سازی کرد.

۱- ۲ سنجش کیفیت و نور در تصویر

با توجه به این که تصاویر با کیفیت یا وضعیت نوری نامناسب (مانند تصاویر تار و تصاویر با رنگ‌های غیرواقعی) باعث افت دقت ماژول‌های تشخیص چهره و تشخیص زنده‌بودن می‌شود؛ آنالیز کیفیت و نور تصاویر و ویدئوها به صورت ضمنی موجب افزایش کارایی ماژول‌های تشخیص چهره و زنده بودن خواهد شد. تشخیص نور و کیفیت در حال حاضر با استفاده از تبدیل فضای رنگی، استفاده از تبدیل فوریه و روش‌های آماری و با سرعت بسیار بالا در این سامانه صورت می‌پذیرد. در شکل ۱–۲ نمونه تصاویر با کیفیت و نور نامناسب که توسط سامانه تشخیص داده‌شده‌اند، ملاحظه می‌شود.



نور نامناسب　　کیفیت نامناسب　　تصویر مورد تایید

شکل ۱–۲ نمونه تصاویر با کیفیت و نور نامناسب که توسط ماژول سنجش کیفیت سامانه تشخیص داده شده‌اند

۱- ۳ تطبیق چهره

در این سامانه سه مدل مجزا برای تطبیق چهره ارزیابی و پیاده‌سازی شده است:

- SepidSystem-Face-Verification-V1: در این مدل تطبیق چهره، قسمت استخراج ویژگی از تصویر بر پایه معماری Inception ResNet-v1 بوده و بر روی دادگان VGGFace2 آموزش دیده است. اندازه قالب[1] استخراج‌شده برای هر تصویر چهره در این مدل بدون فشرده‌سازی ۲۰۴۸ بایت است.
- SepidSystem-Face-Verification-V2: استخراج ویژگی از تصویر در این مدل با معماری ResNet-34 انجام می‌شود و بر روی مجموعه دادگان MS-Celeb-1M آموزش دیده شده است. اندازه قالب استخراج‌شده برای هر تصویر چهره در این مدل بدون فشرده‌سازی ۲۰۴۸ بایت است.
- SepidSystem-Face-Verification-V3: استخراج ویژگی در این مدل بر پایه ResNet-50 صورت می‌گیرد. این مدل بر روی مجموعه دادگان Glint360K آموزش دیده شده که یکی از کامل‌ترین دادگان موجود این حوزه است و روی داده چهره ایرانی بهسازی شده است. اندازه قالب چهره استخراج‌شده برای هر تصویر چهره در این مدل بدون فشرده‌سازی ۵۱۲ بعد است.

در حال حاضر مدل SepidSystem-Face-Verification-V3 برای تطبیق چهره در سامانه احراز هویت مورد استفاده قرار می‌گیرد.

۱- ۴ تشخیص زنده بودن غیرتعاملی (Passive)

دسته‌بندی‌های مختلفی از روش‌های تشخیص زنده بودن وجود دارد که هر کدام از جنبه‌های مختلفی این کار را کرده‌اند. به طور مثال یکی از این دسته‌بندی‌ها، تقسیم روش‌ها به دو دسته‌ی تعاملی و غیرتعاملی می‌باشد. این نوع دسته‌بندی در دنیای تجارت و شرکت‌ها بیشتر کاربرد دارد. نیازمندی‌های کاربران از سامانه، افزایش درصد اطمینان سامانه، سیاست شرکت‌ها و نوع مشتری‌های آنان سبب شده که نحوه‌ی تعامل کاربر با سامانه در انتخاب روش تشخیص زنده بودن اهمیت بسیاری پیدا کند. از این رو الگوریتم‌های موجود در بازار بیشتر بر پایه‌ی این نوع دسته‌بندی می‌باشند. روش‌های تعاملی از کاربر می‌خواهد تا طبق یک سناریوی خاص عمل کند که ممکن است از یک پلک زدن ساده تا گفتن جملات خاص متفاوت باشد. در روش‌های غیرتعاملی، سناریویی وجود ندارد و تنها با دریافت یک تک عکس یا ویدیو از فرد، باید زنده بودن آن را تشخیص داد. روش‌های غیرتعاملی به دلیل عدم ارتباط با کاربر سریع‌تر می‌باشند و زمان تعامل و یا اجرای فرایند مورد نظر در این نوع روش‌ها وجود ندارد. بررسی‌های تجربه رابط کاربری در این حوزه نشان داده است که کاربران در بسیاری از موارد تمایلی به تعامل و اجرای یک کار خاص را ندارند و هم‌چنین به دلیل اینکه کاربران این حوزه می‌توانند از تمام اقشار جامعه باشند و صرفا قشر باتجربه از این نوع سامانه‌ها استفاده نمی‌کنند، ممکن است در حین اجرای عمل خواسته شده از کاربر، اشتباهات زیادی رخ دهد که این موضوع، هم بار پردازشی سامانه و هم زمان اجرای فرایند را بیشتر می‌کند و در نهایت برای کاربر تجربه‌ی راحتی را ایجاد نمی‌کند.

برای تشخیص زنده بودن غیرتعاملی در این سامانه ابتدا با استفاده از ماژول‌های سنجش کیفیت تصویر، کیفیت تصویر چهره بررسی می‌شود و سپس نتیجه نهایی با استفاده از همجوشی[2] دو مدل مجزا در سطح امتیاز مشخص می‌شود. در شکل ۱–۴ نمایی از این روند مشاهده می‌شود. این دو مدل مجزا عبارتند از:

- SepidSystem-Passive-Liveness-V1: شامل مدلی است که استخراج ویژگی در آن بر پایه ResNet-18 صورت می‌گیرد و با پیش‌بینی تصویر عمقی، انعکاس نور و ساختار کانال‌های رنگی در تصویر؛ زنده یا تقلبی بودن تصویر را پیش‌بینی

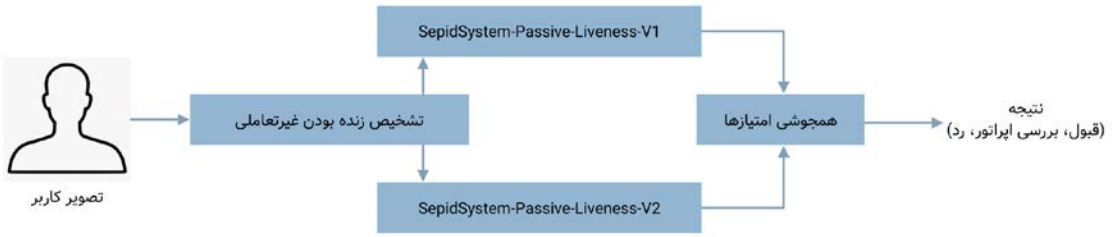
[1] Template
[2] Fusion

می‌کند. این شبکه بر روی یکی از بزرگترین دادگان در این حوزه آموزش داده شده است که از نظر انواع تصاویر جعلی بسیار غنی می‌باشد و شامل انواع مختلف از جمله کاغذهای چاپ‌شده، مانیورها و ماسک‌های سه‌بعدی است (شکل ۱–۳).

تصویر چاپ شده | بازنمایش از روی صفحه نمایش | چاپ شده برش داده شده | ماسک

شکل ۱–۳ نمونه تصاویر تقلبی

- SepidSystem-Passive-Liveness-V2: این مدل بر پایه نسخه سبک‌شده‌ی معماری MobileFaceNet بوده و مدل تشخیص زنده بودن، بر مبنای کاغذهای چاپ‌شده، مانیتور و ماسک سه بعدی چهره آموزش دیده است.



شکل ۱–۴ نمایی از روند تشخیص زنده بودن غیرتعاملی

۱- ۵ تشخیص زنده بودن تعاملی مبتنی بر تشخیص گفتار

روش‌های غیرتعاملی در تشخیص زنده بودن از امنیت بیشتری برخوردار هستند، به این دلیل که سناریوی تشخیص زنده بودن به کاربر گفته نمی‌شود. در سامانه احراز هویت غیرحضوری، مدل SepidSystem-Speech-Recognition-V1 که وظیفه تشخیص گفتار فارسی را به عهده دارد، سرویس تشخیص زنده بودن تعاملی با استفاده از گفتار را ارائه می‌دهد. برای استفاده از بازشناسی گفتار در تشخیص زنده بودن تصویر، تعدادی جمله اختصاصی طراحی شده است که دارای ویژگی‌های زیر است:

- کوتاه باشند
- کلمات آسان و پر تکرار باشند
- همه کلمات در واژگان سامانه بازشناسی گفتار وجود داشته باشند و کلمات خارج از واژگان (OOV)[1] نداشته باشند
- ساختار دستوری جملات بر اساس ساختار نحوی استاندارد فارسی باشد

[1] Out Of Vocabulary

این ویژگی‌ها از این جهت مد نظر بوده است که هم همه افراد (به ویژه افراد عمومی که ممکن دارای تحصیلات و دانش کم باشند) بتوانند از این سرویس استفاده کنند و هم دقت تشخیص سامانه بالا باشد. نمونه‌هایی از جملات در شکل زیر مشاهده می‌شود.

هزینه تبلیغات برای این کار بسیار زیاد است
افراد دانشگاهی همواره در جامعه اثرگذار هستند
ایران برای تبادل دانش مهندسی برق آمادگی دارد
باید تلاش کنیم علم و فرهنگ در کشور رشد کند
کشاورزی در ایران یک صنعت مهم است
ساختن فیلم و بهره‌گیری از هنر رسانه می‌تواند از افسردگی جامعه جلوگیری کند
کارت هوشمند ملی مهم‌ترین سند برای احراز هویت است
تیم‌های استقلال و پرسپولیس دو تیم محبوب قدیمی هستند
نوروز یک عید باستانی برای همه ایرانیان است

شکل ۱–۵ نمونه‌هایی از جملات مورد استفاده در تشخیص زنده بودن تعاملی با استفاده از تشخیص گفتار

بهره‌گیری کاربر از سرویس تشخیص گفتار در سه مرحله انجام می‌شود:

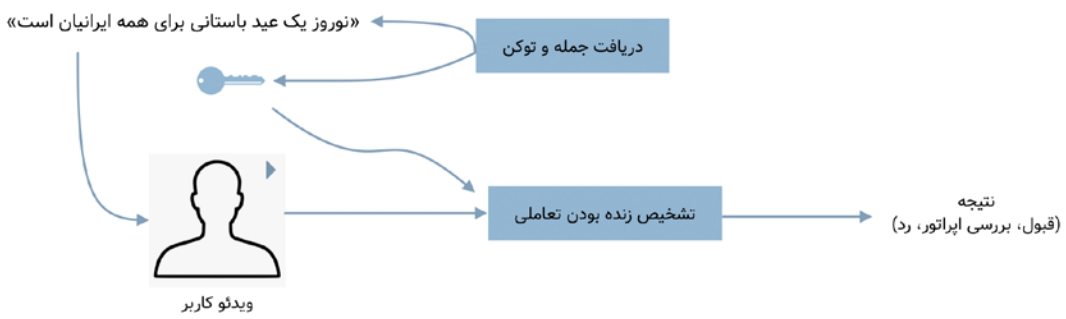
۱- یک جمله فارسی به همراه یک توکن یکتا از سامانه دریافت می‌شود (اولین درخواست). این جمله از بین جملات ساده و روان در موضوعات مختلف از متن اخبار فارسی استخراج شده است. در این مرحله توکن، جمله و زمان درخواست برای پردازش‌های بعدی در پایگاه داده سامانه ذخیره می‌شوند.

۲- از کاربر خواسته می‌شود که در یک ویدئو جمله‌ی مشخص‌شده را بخواند.

۳- ویدئو به همراه توکن دریافتی برای صحت‌سنجی به سامانه ارسال می‌شود (دومین درخواست). در این مرحله، سامانه با توجه به توکن دریافتی، مدت‌زمان گذشته از درخواست اول را محاسبه می‌کند. در صورت قابل قبول بودن میزان فاصله زمانی و در صورتی که توکن قبلا مورد استفاده قرار نگرفته باشد، جمله متناظر با توکن دریافتی از پایگاه داده استخراج می‌شود و با استفاده از ماژول محاسبه امتیاز (بخش بعد) نتیجه نهایی (قبول، بررسی اپراتور یا رد) اعلام می‌شود.

برای افزایش کارایی این سرویس، در آن پایگاه‌داده‌ای از جفت‌های «جمله برای نمایش» و «جمله برای مقایسه با گفتار کاربر» درنظر گرفته شده است. به عنوان مثال در جمله‌ای که برای کاربر نمایش داده می‌شود، علائم نگارشی یا ارقام برای خوانایی بیشتر درنظر گرفته شده است، اما در جمله متناظر آن که برای مقایسه با گفتار کاربر استفاده می‌شود، علائم نگارشی حذف شده و همه ارقام به حروف نوشته شده‌اند تا معادل خروجی مدل به ازای گفتار کاربر باشد. به منظور محاسبه امتیاز در این قسمت، جمله خروجی مدل SepidSystem-Speech-Recognition-V1 که با پردازش صوت کاربر به دست آمده است با «جمله برای مقایسه با گفتار کاربر» مقایسه می‌شود و از معیار MED برای محاسبه شباهت دو جمله استفاده می‌شود. بر روی میزان شباهت به دست آمده با در نظر گرفتن میزان حساسیت سرویس (که هنگام فراخوانی دومین درخواست دریافت شده)، آستانه‌های مناسب اعمال می‌شود و نتیجه تشخیص زنده بودن (قبول، بررسی اپراتور یا رد) مشخص خواهد شد.

در شکل ۱–۶ نمای کلی این فرایند ملاحظه می‌شود.



شکل ۱–۶ ساختار تشخیص زنده بودن تعاملی با استفاده از تشخیص گفتار

موتور تشخیص گفتار فارسی مورد استفاده در سامانه مبتنی بر یادگیری عمیق و آموزش یافته با حدود ۱۰۰۰ ساعت گفتار فارسی است که گفتار مورد استفاده از گوینده‌های فارسی زبان در سراسر کشور با تنوع لهجه، جنسیت، سن، تحصیلات و با جملات محاوره/رسمی تهیه شده است.

۱- ۶ تشخیص زنده بودن تعاملی مبتنی بر پلک زدن

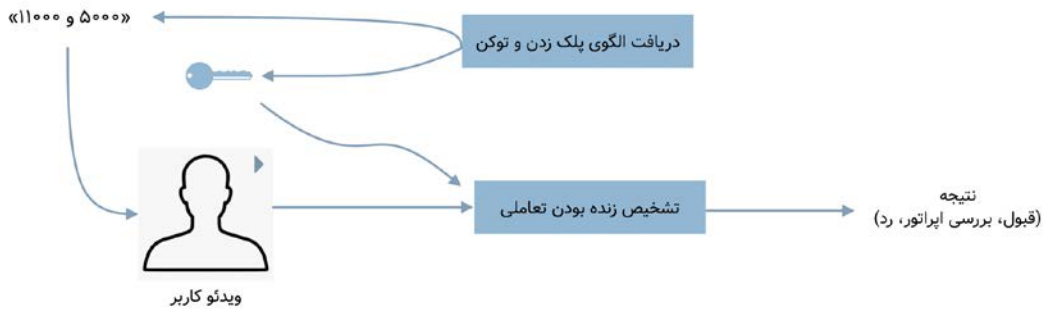
مدل SepidSystem-Blink-Detection-V1 وظیفه تشخیص زمان‌های پلک زدن در ویدئو را به عهده دارد. هسته هوش مصنوعی با تکیه بر این مدل، سرویس تشخیص زنده بودن تعاملی با استفاده از پلک زدن را ارائه می‌کند. بهره‌گیری کاربر از این از این سرویس در سه مرحله انجام می‌شود:

۱- یک الگوی پلک زدن (زمان‌های پلک زدن) به همراه یک توکن از سامانه دریافت می‌شود (اولین درخواست). این الگو به صورت تصادفی و به گونه‌ای انتخاب می‌شود که با پلک زدن ناخودآگاه انسان تفاوت داشته باشد. این الگو شامل چند عدد صحیح است که زمان‌هایی (به میلی‌ثانیه) که کاربر باید پس از آن پلک بزند را مشخص می‌کند. در این مرحله توکن، جمله و زمان درخواست برای پردازش‌های بعدی در پایگاه داده سامانه ذخیره می‌شوند.

۲- از کاربر خواسته می‌شود که در یک ویدئو در زمان‌های مشخص‌شده پلک بزند. به عنوان مثال اگر الگو به صورت «۵۰۰۰ و ۱۱۰۰۰» باشد، کاربر لازم است در ویدئو دو بار، یکی پس از ثانیه ۵ (حداکثر تا دو ثانیه پس از آن) و دیگری پس از ثانیه ۱۱ (حداکثر تا دو ثانیه پس از آن) پلک بزند.

۳- ویدئو به همراه توکن دریافتی برای صحت‌سنجی به سامانه ارسال می‌شود (دومین درخواست). در این مرحله، سامانه با توجه به توکن دریافتی، مدت‌زمان گذشته از درخواست اول را محاسبه می‌کند. در صورت قابل قبول بودن میزان فاصله زمانی و در صورتی که توکن قبلا مورد استفاده قرار نگرفته باشد، الگوی پلک زدن متناظر با توکن دریافتی از پایگاه داده استخراج می‌شود و با استفاده از ماژول محاسبه امتیاز (بخش بعد) نتیجه نهایی (قبول، بررسی اپراتور یا رد) اعلام می‌شود.

در شکل ۱–۷ نمای کلی این فرایند ملاحظه می‌شود.



شکل ۱–۷ ساختار تشخیص زنده بودن تعاملی با استفاده از تشخیص پلک زدن

به منظور محاسبه امتیاز در این قسمت، با استفاده از مدل SepidSystem-Blink-Detection-V1 زمان پلک‌های کاربر مشخص می‌شود. میزان شباهت پلک‌های زده شده با الگوی خواسته شده با در نظر گرفتن این دو مورد محاسبه می‌شود:

- پلک زدن کاربر در زمان‌های خواسته نشده
- پلک نزدن کاربر در خارج از زمان‌های خواسته شده

بر روی میزان شباهت به دست آمده با در نظر گرفتن میزان حساسیت سرویس (که هنگام فراخوانی دومین درخواست دریافت شده)، آستانه‌های مناسب اعمال می‌شود و نتیجه تشخیص زنده بودن (قبول، بررسی اپراتور یا رد) مشخص خواهد شد.

۱– ۷ ماژول‌های ترکیبی

در فرآیندهای واقعی احراز هویت نیاز است ماژول‌های بیان شده به صورت ترکیبی استفاده شود، به عنوان مثال همراه با تشخیص زنده بودن به صورت تعاملی، تطبیق چهره نیز انجام شود. به این ترتیب، سرویس‌های ذکرشده در بخش‌های قبلی معمولا به صورت هم‌زمان برای تشخیص پایدارتر به‌کار می‌روند. در حال حاضر فرآیندهای ترکیبی زیر در سامانه احراز هویت پیش‌بینی شده‌اند که شامل موارد زیر هستند:

۱ –۷ –۱ تشخیص زنده بودن تعاملی (گفتار/پلک زدن) و غیرتعاملی

به منظور پایدارتر شدن سرویس‌های تشخیص زنده بودن تعاملی، در کنار در همگی آن‌ها تشخیص زنده بودن غیرتعاملی نیز به صورت موازی بر روی ویدئو دریافتی صورت می‌گیرد. به این ترتیب خروجی‌های سرویس‌های تشخیص زنده بودن تعاملی همگی شامل بخش‌های زیر هستند:

- پاسخ تشخیص زنده بودن تعاملی در ویدئو
- پاسخ تشخیص زنده بودن غیرتعاملی در ویدئو
- پاسخ کلی (همجوشی وضعیت دو بخش فوق)

۱-۷-۲ تشخیص زنده بودن تعاملی (گفتار/پلک زدن) و غیرتعاملی و تطبیق چهره با تصویر مرجع

در این حالت یک ویدئو و یک تصویر از کاربر دریافت می‌شود و تشخیص زنده بودن تعاملی و غیرتعاملی بر روی ویدئو دریافتی صورت می‌گیرد. به طور موازی تطبیق چهره نیز بین ویدئو و تصویر دریافتی انجام می‌شود. نتیجه نهایی این سامانه شامل این بخش‌ها است:

- پاسخ تشخیص زنده بودن تعاملی با گفتار در ویدئو
- پاسخ تشخیص زنده بودن غیرتعاملی در ویدئو
- پاسخ تطبیق چهره بین ویدئو و تصویر
- پاسخ کلی (همجوشی وضعیت دو بخش فوق)

۱-۷-۳ تشخیص زنده بودن غیرتعاملی با ویدئو و تطبیق چهره با تصویر مرجع

در این حالت یک ویدئو و یک تصویر مرجع دریافت می‌شود (تصویر مرجع می‌تواند تصویر کارت ملی باشد، یا تصویری باشد که از قبل در سامانه ثبت‌نام شده است و یا از منابع بیرونی مانند ثبت‌احوال دریافت شده است). در این حالت دو وظیفه بررسی زنده بودن ویدئو و تطبیق چهره کاربر در ویدئو با تصویر مرجع به صورت موازی در سامانه انجام می‌شود. در این حالت پاسخ نهایی شامل سه بخش است:

- پاسخ تشخیص زنده بودن غیرتعاملی در ویدئو
- پاسخ تطبیق چهره بین ویدئو و تصویر
- پاسخ کلی (همجوشی وضعیت دو بخش فوق)

۱-۷-۴ تشخیص زنده بودن غیرتعاملی با یک تصویر و تطبیق چهره با تصویر مرجع

در این حالت دو تصویر از ورودی دریافت می‌شود که فرض می‌شود یکی از آن‌ها تصویر زنده کاربر است (تصویر سلفی) و دیگری تصویری از وی که از قبل در سامانه موجود بوده یا به واسطه استعلام حاصل شده است. در این حالت برای تصویر اول بررسی تشخیص زنده بودن انجام می‌شود و به صورت همزمان تطبیق چهره بین دو تصویر صورت می‌گیرد. نتیجه این سرویس نیز مانند سرویس قبلی شامل سه بخش است:

- پاسخ تشخیص زنده بودن غیرتعاملی در تصویر زنده
- پاسخ تطبیق چهره بین دو تصویر
- پاسخ کلی (همجوشی وضعیت دو بخش فوق)

فصل ۲ تست و ارزیابی سامانه

همانطور که در فصل قبل اشاره گردید، سامانه‌ی eKYC دارای مدل‌های مختلفی می‌باشد که هر کدام دارای پیاده‌سازی منحصر به خود هستند. به منظور ارزیابی الگوریتم‌های سامانه، آزمایش‌های مختلفی بر روی دادگان متفاوتی صورت گرفت. هر کدام از الگوریتم‌ها بر روی مجموعه داده‌هایی که متشکل از دادگان معروف است، مورد آزمایش قرار گرفتند. در ادامه نتایج مربوط به این آزمایش‌ها و توضیحات داده‌های استفاده شده به تفکیک نوع خدمات آورده شده است.

۲-۱ ارزیابی سامانه مکان‌یابی چهره[1]

در این قسمت به مقایسه‌ی دو مدل برتر مکان‌یابی چهره که در سامانه پیاده‌سازی شده است، پرداخته می‌شود. قبل از بیان نتایج باید به این موضوع اشاره کنیم که اولین مرحله‌ی ارزیابی بر روی یک مجموعه ۱۰۰ عددی از تصاویری می‌باشد که در طی فرایند آزمایش‌های مختلف سامانه، به عنوان تصاویر چالشی از ویدیو و یا عکس‌های سلفی چهره‌های ایرانی جمع‌آوری شده است. در این نمونه تصاویر، چالش‌هایی مانند چند چهره‌ای، تار بودن تصویر، تصویر بی‌کیفیت، تصاویر دارای چرخش و تصاویر کارت‌های شناسایی مختلف گنجانده شده است. تصاویر به مدل‌ها داده و نتایج آن‌ها به صورت شهودی مقایسه شده است.

همانطور که بیان شد تصاویر انتخاب شده برای آزمایش، دارای چالش‌های مختلفی می‌باشند تا بتوان قدرت هر کدام از الگوریتم‌ها را شناسایی کرد. داده‌های ورودی به دو دسته تصویر و ویدیو تقسیم می‌شوند که داده‌ی ویدیو، فریم استخراج شده از ویدیو می‌باشد. نمونه‌ای از این تصاویر را در شکل ۲–۱ مشاهده می‌کنید.

[1] Face Detection

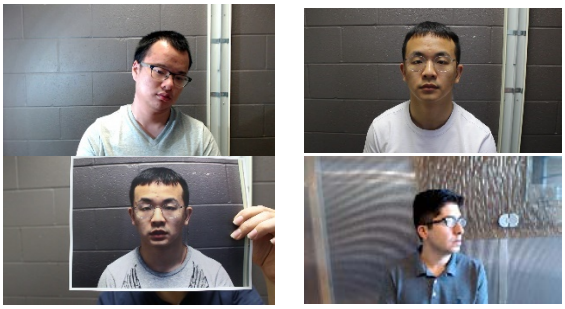
شکل ۲–۱ چند نمونه از تصاویر استفاده شده برای آزمایش‌های اولیه الگوریتم‌های مکان‌یابی چهره

با توجه به اینکه تصاویر مختلف در هر کدام از آزمایش‌ها، نتایج مختلفی را برای مکان‌یابی چهره ارائه می‌دهند، در جدول ۲–۱ میانگین درصد اطمینان و همچنین زمان تشخیص هرکدام از ارزیابی‌ها آورده شده است. باید به این نکته توجه شود که اعداد برای مقایسه در کنار هم قرار نگرفتند، چرا که در آزمایش SepidSystem-Face-Detection-V3 چهره در تصاویر با کیفیت پایین (اندازه چهره کوچک) تشخیص داده نشده‌اند (الگوریتم به گونه‌ای تنظیم شده است که دارای حداقل اندازه تصویر ورودی می‌باشد) و در نتیجه درصدی برای آن‌ها موجود نمی‌باشد. واضح است که درصد تشخیص درستی برای این نوع تصاویر پایین‌تر می‌باشد و برای همین SepidSystem-Face-Detection-V4 درصد پایین‌تری را به خود اختصاص داده‌اند.

جدول ۲–۱ میانگین درصد اطمینان درستی و زمان برای پردازش هر تصویر

عنوان تست	میانگین درصد اطمینان	میانگین زمان (ثانیه)
SepidSystem-Face-Detection-V4	۰.۹۸۳	۳.۲۲
SepidSystem-Face-Detection-V3	۰.۹۹۳	۱

در ادامه یک بررسی دیگر بر روی دو الگوریتم انجام شده است. این مقایسه بر روی یک مجموعه‌ی داده‌ی بزرگتر صورت گرفته است. همان‌طور که مشاهده می‌شود زمان الگوریتم‌های بیان شده متفاوت می‌باشد، اما کارایی آن‌ها تا حد بسیار خوبی با یکدیگر مشابه است. به منظور اطمینان از عملکرد مدل‌ها، یک مرحله آزمایش دیگر بر روی زیر مجموعه‌ای از فریم‌های استخراج شده از ویدیوهای دادگان SIW انجام گرفت. این فریم‌ها شامل چهره در موقعیت‌های مختلف می‌باشد. در شکل ۲–۲ نمونه‌ای از این تصاویر را مشاهده می‌کنید.



شکل ۲–۲ نمونه‌ای از فریم‌های استخراج شده از دادگان SIW

از آنجایی که اطلاعات کادر چهره[1](bbox) برای فریم های SIW که در دادگان خود آن موجود است، از نظر اندازه و هم‌چنین دقت تشخیص مناسب نبودند. امکان مقایسه bboxهای استخراج شده‌ی این تصاویر با داده مرجع وجود نداشت به همین منظور از همان مقایسه شهودی استفاده می‌گردد، خطاها مقایسه و تعداد آن‌ها بررسی می‌شود. تعداد فریم‌های استخراج شده در تصاویر زنده به ۱۲۴۳۵ عدد و در تصاویر جعلی به ۱۶۸۰۳ عدد می‌رسد. دسته‌بندی تصاویر به این دو گروه به این دلیل است که نوع آن‌ها بر تعداد خطاها تاثیر گذاشته است.

هر دو الگوریتم دارای قدرت کافی برای تشخیص چهره در شرایط استاندارد می‌باشند. کادرهایی که هر دو تصویر برای چهره استخراج می‌کنند مشابه یکدیگر می‌باشد و هر دو تقریبا فضای یکسانی از چهره را به عنوان خروجی ارائه می‌دهند. الگوریتم SepidSystem-Face-Detection-V4 در شرایطی که تصویر حالت‌های خاص دارد مانند کیفیت پایین، شرایط نوری نامناسب، چرخش سر و شرایط خاص دیگر بهتر عمل می‌کند اما در تصاویر که در آن‌ها چهره بخش بیشتری از تصویر را به خود اختصاص داده است و یا به عبارتی نزدیک دوربین است به خطا می‌خورد. سرعت SepidSystem-Face-Detection-V3 بیشتر است و زمان کمتری را برای پردازش نیاز دارد.

۲– ۲ ارزیابی سامانه تطبیق چهره

در این بخش به ارزیابی الگوریتم‌های تطبیق چهره در سامانه پرداخته می‌شود. لازم به ذکر است که به منظور ارزیابی دقیق از دادگان معروف و استانداردی که برای محک‌زنی این نوع الگوریتم‌ها در دنیا به کار می‌رود، استفاده شده است و به منظور ارزیابی کارایی سامانه روی تصاویر چهره ایرانی، از دادگان تست تهیه شده برای این کار استفاده شده است.

۲ –۲ –۱ دادگان‌های ارزیابی تطبیق چهره

دادگانی که برای ارزیابی الگوریتم‌های تطابق چهره استفاده شده است عبارتند از LFW پاک‌سازی شده، IranCeleb (چهره ایرانی)، مجموعه‌داده LFW استاندارد، مجموعه داده LFW با ۱ میلیون Non-match و مجموعه داده LFW با ۱۰ میلیون Non-match که در ادامه تشریح شده‌اند.

- **مجموعه‌داده LFW پاک‌سازی شده:** این مجموعه‌داده دارای 1200 تصویر است. به‌منظور ساخت این مجموعه‌داده اقدام به پالایش مجموعه‌داده LFW استاندارد شد (شکل ۲–۳). بدین صورت که تصاویر با بیش از یک چهره از داخل مجموعه‌داده حذف شدند. سپس اقدام به ساخت دسته‌های match و nonmatch شد. تعداد pairهای تولید شده برای هر کدام از دسته‌های match و nonmatch برابر با 600 جفت است.

[1] Bounding Box (bbox)


شکل ۲–۳ نمونه‌ای از تصاویر LFW

- **مجموعه‌داده IranCeleb**: این مجموعه‌داده شامل تصاویر سلبریتی‌های ایرانی است که دارای ۱۲۰۰ تصویر است (شکل ۲–۴). به‌منظور ساخت این مجموعه‌داده اقدام به پالایش مجموعه‌داده IranCeleb استاندارد شد. بدین صورت که تصاویر با بیش از یک چهره از داخل مجموعه‌داده حذف شدند. سپس اقدام به ساخت دسته‌های match و nonmatch شد. تعداد pairهای تولید شده برای هر کدام از دسته‌های match و nonmatch برابر با ۶۰۰ جفت است.


شکل ۲–۴ نمونه‌ای از تصاویر IranCeleb

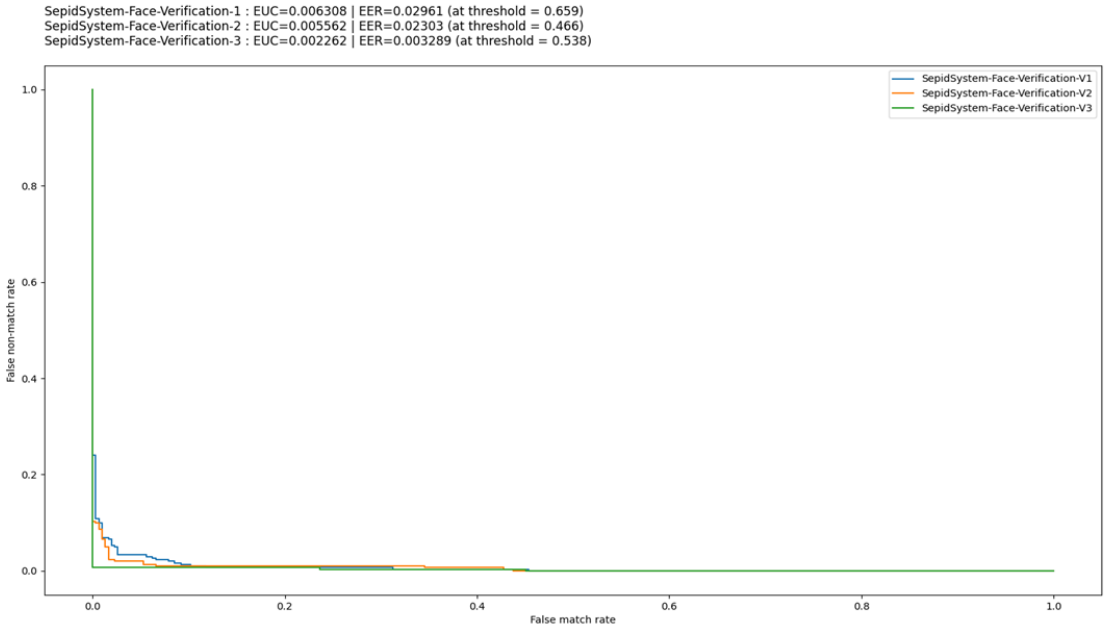
- **مجموعه‌داده LFW استاندارد:** این مجموعه‌داده دارای ۱۳۲۳۳ تصویر متعلق به ۵۷۴۹ شخصیت می‌باشد. ۳۰۰۰ جفت تصویر به‌عنوان match و ۳۰۰۰ جفت تصویر به‌عنوان non-match در نظر گرفته شده‌اند.
- **مجموعه داده LFW با ۱ میلیون Non-match**: این مجموعه‌داده همان مجموعه‌داده LFW استاندارد است با این تفاوت که ۱۲۱۱۲۸۵ جفت تصویر به‌عنوان non-match و ۲۴۲۲۵۷ به‌عنوان match در نظرگرفته شده است.
- **مجموعه‌داده LFW با ۱۰ میلیون Non-match**: این مجموعه‌داده همان مجموعه‌داده LFW استاندارد است با این تفاوت که ۱۰ میلیون جفت تصویر به‌عنوان non-match و ۲۴۲۲۵۷ به‌عنوان match در نظر گرفته شده است.

۲ -۲ -۲ نتایج ارزیابی

در ادامه نتایج آزمایش مدل‌های تطابق چهره برای دادگان بیان شده آورده شده است.

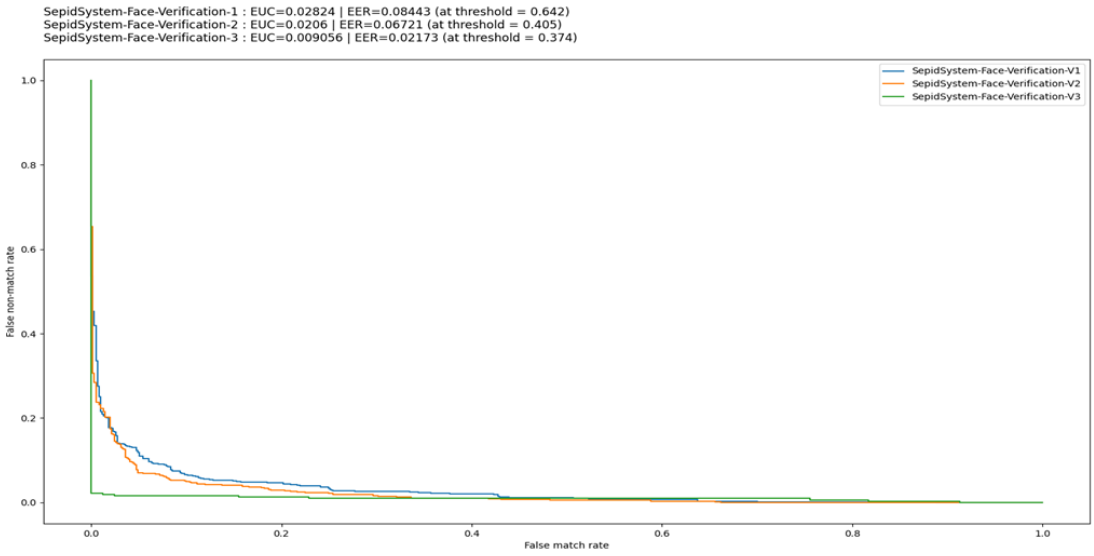
- **نتایج ارزیابی بر روی مجموعه‌داده LFW پاک‌سازی شده:** در این بخش سه مدل SepidSystem-Face-Verification-1 , SepidSystem-Face-Verification-2 و SepidSystem-Face-Verification-3 بر روی مجموعه‌داده

اعمال شده و نتایج قالب نمودار DET (شکل ۲–۵) مورد ارزیابی قرار گرفت که همانطور که مشخص است، کمترین خطا مربوط به SepidSystem-Face-Verification-3 با EUC=0.002 است.



شکل ۲–۵ نمودار DET مقایسه سه مدل تطابق چهره بر روی LFW پاک‌سازی شده

- **نتایج ارزیابی بر روی مجموعه‌داده IranCeleb**: در این بخش سه مدل , SepidSystem-Face-Verification-V2 SepidSystem-Face-Verification-V3 و SepidSystem-Face-Verification-V3 بر روی مجموعه‌داده اعمال شده و نتایج قالب نمودار DET (شکل ۲–۶) مورد ارزیابی قرار گرفت که همانطور که مشخص است، کمترین خطا مربوط به SepidSystem-Face-Verification-3 با EUC=0.009 است.



شکل ۲–۶ نمودار DET مقایسه سه مدل تطابق چهره بر روی iranCeleb

- **ارزیابی SepidSystem-Face-Verification-V3 بر روی مجموعه‌داده‌ها:** با توجه به اینکه مدل SepidSystem-Face-Verification-V3 دارای بهترین عملکرد بر روی دو مجموعه‌داده iranCeleb و LFW است، از این پس به منظور ارزیابی، تنها از این مدل استفاده می‌شود که در جدول ۲–۲ نتایج روی سایر دادگان‌ها ارائه شده است.

جدول ۲–۲ ارزیابی SepidSystem-Face-Verification-V3 بر روی مجموعه‌داده‌ها

عنوان دادگان	EER	EUC
LFW استاندارد	0.003	0.00091
LFW با ۱ میلیون non-match	0.0003791	0.0001193
LFW با ۱۰ میلیون non-match	0.0003792	0.0001189

- **جمع‌بندی و خلاصه نتایج:** نتایج ارزیابی الگوریتم‌های تطابق چهره در سامانه با معیارهای EUC و EER (جدول ۲–۳) و دقت (جدول ۲–۴) به طور خلاصه آورده شده است.

جدول ۲–۳ مقدار EUC و EER مربوط به مدل‌های تطبیق چهره بر روی مجموعه‌داده‌های مختلف

اسم مدل	LFW-600 Pair	IranCeleb	Standard LFW	LFW-1M Pair	LFW-10M Pair
SepidSystem-Face-Verification-V1	EUC=0.006308 EER=0.02961	EUC=0.02824 EER=0.08443	-	-	-
SepidSystem-Face-Verification-V2	EUC=0.005562 EER=0.02303	EUC=0.0206 EER=0.06721	-	-	-
SepidSystem-Face-Verification-V3	EUC=0.002262 EER=0.003289	EUC=0.009801 EER=0.02295	EUC=0.0009148 EER=0.003	EUC=0.000119 EER=0.0003791	EUC=0.0001189 EER=0.0003792

جدول ۲–۴ مقدار دقت (Accuracy) مربوط به مدل‌های تطبیق چهره بر روی مجموعه‌داده‌های مختلف

اسم مدل	LFW-600 Pair	IranCeleb	Standard LFW	LFW-1M Pair	LFW-10M Pair
SepidSystem-Face-Verification-V1	0.969	0.921	-	-	-
SepidSystem-Face-Verification-V2	0.979	0.939	-	-	-
SepidSystem-Face-Verification-V3	0.995	0.988	0.998	0.9998	0.9999

۲- ۳ ارزیابی سامانه تشخیص زنده بودن

۲ -۳ -۱ تشخیص زنده بودن غیرتعاملی

برای ارزیابی تشخیص زنده بودن غیرتعاملی، یک مجموعه داده شامل چند دادگان معروف در این حوزه و همچنین یک مجوعه داده ایرانی جمع‌آوری شده است که در تصاویر جعلی آن شامل انواع جعل است. اطلاعات این دادگان در جدول ۲–۵ آورده شده است.

جدول ۲–۵ اطلاعات دادگان استفاده شده برای ارزیابی تشخیص زنده بودن غیرتعاملی (دادگان نامتوازن)

مجموعه داده	تصاویر زنده	تصاویر جعلی
Celeba spoof	19923	47247
SIW	12118	13487
HKBU 3D	504	504
Iranian-Logs	500	-
LCC-fasd	1300	8240

در جدول ۲–۶ نتایج مدل مورد استفاده در سامانه با دو روش همجوشی (برای ادغام تصمیم دو مدل پایه این سرویس) آورده شده است.

جدول ۲–۶ نتایج تشخیص زنده بودن غیرتعاملی بر روی مجموعه داده‌ی نامتوازن

نوع داده	EER	EUC
SepidSystem-Passive-Liveness-V1*V2 (Fusion1)	0.125	0.05
SepidSystem-Passive-Liveness-V1*V2 (Fusion2)	0.116	0.046

در ادامه، یک آزمایش دیگر بر روی دادگان ارزیابی همگن انجام شده است تا نتایج این سرویس دقیق‌تر بیان گردد. در جدول ۲–۷ اطلاعات مربوط به دادگان همگن را مشاهده می‌کنید.

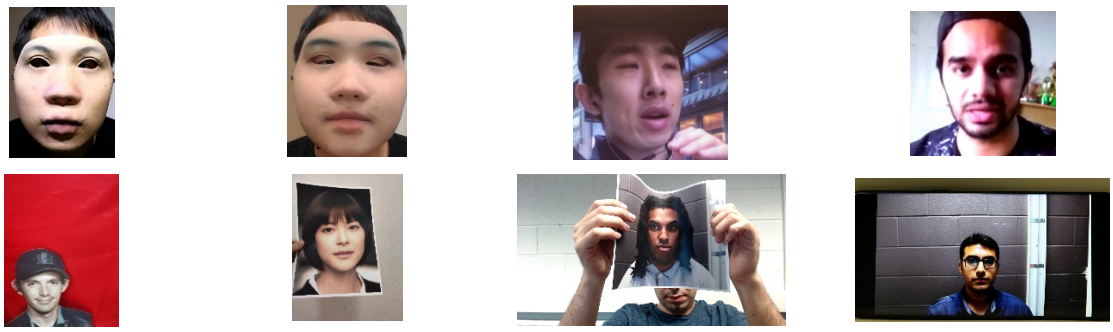
جدول ۲–۷ اطلاعات دادگان استفاده شده برای ارزیابی تشخیص زنده بودن غیرتعاملی (دادگان متوازن)

مجموعه داده	تصاویر زنده	تصاویر جعلی
Celeba spoof	10000	10000
SIW	10000	10000
HKBU 3D	504	504
Iranian-Logs	500	-
LCC-fasd	1200	1200

در شکل ۲–۷ نمونه‌ای از تصاویری زنده‌ای که در این آزمایش مورد استفاده قرار گرفته است، مشاهده می‌شود.

شکل ۲–۷ نمونه‌ای از تصاویر زنده‌ی استفاده شده برای ارزیابی تشخیص زنده بودن غیرتعاملی

در شکل ۲–۸ نمونه‌ای از تصاویر جعلی که برای آزمایش استفاده شده، آورده شده است.



شکل ۲–۸ نمونه‌ای از تصاویر جعلی استفاده شده برای ارزیابی تشخیص زنده بودن غیرتعاملی

در جدول ۲–۸، EER نتایج بر روی مجموعه داده‌ی همگن آورده شده است که در آن برای ادغام دوم که کارایی بهتری دارد، ارائه شده است.

جدول ۲–۸ نتایج تشخیص زنده بودن غیرتعاملی بر روی مجموعه داده‌ی متوازن

مدل‌ها	EER
SepidSystem-Passive-Liveness-V1*V2 (Fusion2)	0.1714

۲ -۳ -۲ تشخیص زنده بودن تعاملی

- **مبتنی بر تشخیص گفتار:** برای ارزیابی تشخیص زنده بودن تعاملی با روش تشخیص گفتار از دو دادگان تهیه شده برای زبان فارسی استفاده شد که خلاصه مشخصات آنها به صورت زیر است:
  - دادگان DeepMine: که شامل ۱۰۲۱ جمله فارسی است که توسط گوینده‌های زن و مرد (با نسبت حدود ۴۰٪ زن و ۶۰٪ مرد) با خواندن جملات مختلف جمع‌آوری شده است. در جمع آوری داده‌ها از یک اپ موبایل استفاده شده و جملات در شرایط واقعی و در محیط‌های متنوع ضبط شده است که در آنها نویز و جملات محاوره هم وجود دارد.
  - دادگان میکروفونی: که شامل ۱۷۸۰ جمله رسمی فارسی ضبط شده با صدای زن و مرد (با نسبت حدود ۳۰٪ زن و ۷۰٪ مرد) با استفاده از میکروفون‌های متصل به رایانه و در محیط‌های واقعی است.

علاوه بر دادگان‌های بزرگ بیان شده، یک ارزیابی روی تعدادی از گوینده‌ها (۵ نفر) که جملات طراحی شده برای تشخیص زنده بودن در این پروژه را خوانده‌اند صورت گرفته است که نتایج سرویس تشخیص گفتار روی این سه دادگان در جدول ۲–۹ آورده شده است که در آن معیار ارزیابی نرخ خطای کلمه (WER) است که بیانگر درصد کلمات تشخیص داده شده به صورت نادرست در جملات تست است. همانطور که مشخص است، کارایی سرویس روی جملات مورد استفاده برای تشخیص زنده بودن (به دلیل سادگی جملات و عدم وجود کلمات خارج از واژگان) بالاست.

جدول ۲–۹ ارزیابی سرویس تشخیص گفتار با معیار نرخ خطای کلمه

دادگان	WER
DeepMine	0.091
میکروفونی	0.048
جملات eKYC	0.019

مبتنی بر تشخیص پلک زدن: به منظور ارزیابی زنده بودن از طریق پلک زدن در سامانه احراز هویت از مجموعه داده eyeblink8 استفاده شد و کارایی الگوریتم پیاده‌سازی شده بر روی این مجموعه داده مورد ارزیابی قرار گرفت. این مجموعه داده دارای ۸ نمونه از ۴ شخصیت مختلف می‌باشد که از بین این چهار نفر، یکی از آنها عینکی است. ویدیوها در محیط خانه ضبط شده‌اند و شخص مستقیم روبروی دوربین نشسته است و مشابه شرایط مورد نیاز در کاربرد این پروژه است. در این مجموعه داده انتخاب شده برای ۴ نفر، در مجموع تعداد ۴۱۵۸۷ فریم برچسب‌گذاری شده‌اند که از تعداد حدود ۱۰۰۰ فریم چشم بسته بوده و در باقی موارد چشم باز بوده است. رزولوشن تصاویر این مجموعه داده ۶۴۰*۴۸۰ می‌باشد.

برای ارزیابی الگوریتم، ابتدا مکان چهره در تصاویر با الگوریتم مکان‌یابی چهره مشخص شد و سپس بسته بودن یا باز بودن چشم تشخیص داده شده است. در جدول زیر دقت (Accuracy) الگوریتم در تشخیص باز و بسته بودن برای هر کدام از ۴ نمونه به تفکیک و همچنین متوسط کارایی نهایی ارائه شده است.

جدول ۲–۱۰ ارزیابی سرویس تشخیص پلک زدن با معیار دقت (Accuracy)

عنوان	تعداد فریم	میزان دقت تشخیص پلک زدن (درصد)
نمونه ۱	۱۵۷۸۴	۹۹.۳
نمونه ۲	۱۱۱۸۲	۹۹.۲
نمونه ۳	۹۲۱۶	۹۸.۳
نمونه ۴	۵۴۰۵	۹۹.۳
متوسط		**۹۹.۱**

فصل ۳ نحوه‌ی استفاده از APIها و ماژول مدیریت و دسترسی کاربران

همان‌طور که بیان شد، ساختار پروژه به ترتیب دارای سه لایه هوش مصنوعی، مدیریت و دمو می باشد. در قسمت هوش مصنوعی مسائل مربوط به پردازش تصاویر و ویدئو ها است که به لایه مدیریت سرویس می‌دهند. در لایه مدیریت، مسائل مربوط به کنترل میزان مصرف و سطح دسترسی بررسی می‌شود. همان‌طور که از اسم لایه دمو برداشت می‌شود، این لایه صرفا برای ارائه دمو و آزمون سیستم توسط کاربر نهایی می‌باشد و به دلیل در خواست کارفرما و متناسب نیاز کسب و کار استفاده کننده پیاده‌سازی شده است.

۳-۱ ساختار پروژه

در این بخش زیرساخت پروژه و هر کدام از لایه‌های بیان شده در مقدمه بررسی می‌گردد.

- **زیر ساخت:** معماری سیستم بر اساس سرویس‌های مستقل و بر بستر داکر طراحی شده است و برای نصب بر روی سرور، به سرورهای لینوکسی نیاز است. به این ترتیب هر سرویس به صورت ایمیج‌های داکر پیاده‌سازی شده‌اند که از مزایای آن می‌توان به افزایش توزیع پذیری، استقلال سرویس‌ها نسبت به یکدیگر و محیط استقرار اشاره نمود. در حال حاظر سیستم عامل سرور اصلی توزیع Centos بوده که آخرین نسخه داکر بر روی آن نصب شده است.
- **لایه هوش مصنوعی:** این لایه شامل مدل‌های هوش مصنوعی و الگوریتم های مربوطه برای تشخیص هویت می‌باشد (در بخش‌های قبلی به تشریح این لایه پرداخته شده است). این لایه از بیرون سیستم قابل دسترسی نبوده و صرفا با لایه مدیریت در تعامل می‌باشد.
- **لایه مدیریت:** این لایه برای ارائه سرویس‌های هوش مصنوعی به سرویس‌های خارجی می‌باشد. وظایفی همچون احراز هویت و سنجش میزان استفاده از سرویس‌ها در این لایه انجام شود.

- **لایه دمو:** این لایه به درخواست ناجا و صرفا برای ارائه دمو پیاده‌سازی شده است. این لایه شامل یک اپلیکیشن بک‌اند برای ارائه سرویس و یک اپلیکیشن فرانت‌اند برای نمایش نحوه کارکرد سیستم می‌باشد. دلیل استفاده از اپلیکیشن بک‌اند به این دلیل می‌باشد که برای فراخوانی سرویس‌های هوش مصنوعی نیاز به شناسه یکتای ApiToken می‌باشد که عبارتی محرمانه است و نمی‌توان آن را در برنامه سمت کاربر ذخیره نمود. به این ترتیب، یک اپلیکیشن بک‌اند توسعه داده شده است که درخواست های سمت کاربر را بدون احراز هویت دریافت می‌نماید و در پس زمینه برنامه، شناسه یکتا را به درخواست اضافه کرده و برای برای سرویس‌های هوش مصنوعی ارسال می‌نماید. همچنین بنا به درخواست ناجا نیاز به آن بود تا برخی تصاویر به صورت پیش‌فرض در سیستم ذخیره شده باشند تا برای تست سیستم از آن ها استفاده شود. به دلیل آن که این نیازمندی خاص ناجا می باشد، ذخیره داده‌های اضافی مورد نیاز نیز در اپلیکیشن بک‌اند ذخیره می‌شوند. اپلیکیشن سمت فرانت هم برای ایجاد یک وب‌سایت که امکان تست سرویس‌های تطبیق چهره، تشخیص زنده بود غیر تعاملی و تعاملی (با حالت‌های مختلف) را می‌دهد، توسعه داده شده است.

۳- ۲ دسترسی به ماژول‌های هسته‌ی مرکزی

به منظور استفاده از سرویس‌های پروژه لازم است مقدار ApiToken را در هدر درخواست قرار دهید. این توکن محرمانه در اختیار شما قرار می‌گیرد و شناسه‌ای برای تشخیص دادن شما از سایر مشتریان است. همچنین رابط کاربری swagger این سرویس‌ها نیز در آدرس https://[ServerAddress]/swagger/index.html قابل دسترسی می‌باشد که می‌توان به راحتی از آن استفاده کرد. در این آدرس منظور از [ServerAddress] آدرس سروری است که سرویس‌ها روی آن نصب شده است. در ادامه به تشریح هر کدام از سرویس‌ها پرداخته شده است.

- **احراز هویت:** برای احراز هویت کاربران از ApiToken استفاده می‌شود. این شناسه یکتا و ثابت است که به ازای هر کاربر صادر می‌شود. به این ترتیب در هنگام فراخوانی سرویس‌ها از این شناسه برای تشخیص دسترسی کاربر استفاده می‌شود و با توجه به میزان دسترسی سرویس داده می‌شود. در حال حاضر تعدادی از آن صادر شده و در اختیار ناجا قرار گرفته است. برای صدور توکن‌های جدید Endpoint مشخصی قرار داده شده است که با استفاده از آن می‌توان توکن‌های جدید صادر نمود. همزمان با استفاده از این توکن، سرویس IdentityServer هم در پروژه در نظر گرفته شده است که در حال حاظر غیرفعال است چون در پروژه ناجا کاربردی ندارد. در صورت نیاز به توسعه پرتال‌های خاص، می‌توان از آن برای کنترل سطح دسترسی نیز استفاده نمود. در حال حاضر به جهت امنیت و دسترسی هیچ نیازی به این سرویس نمی‌باشد و صرفا در صورت بروز نیازمندی‌های جدید از آن استفاده می‌شود.
- **ملاحظات و محدودیت‌های عمومی:** برای بهره‌گیری هرچه بهتر از این سرویس‌ها لازم است در تصاویر و ویدئوهای ورودی شرایط زیر برقرار باشد:
    - در هریک از تصاویر یا ویدئوها، دقیقاً یک چهره وجود داشته باشد.
    - تصاویر و ویدئوها در شرایط نوری مناسب (نور از روبه‌رو) دریافت شوند؛ در حالتی که نور از پشت‌سر می‌تابد یا قسمتی از چهره سایه افتاده تصویر دریافت نشود.
    - برای دریافت نتیجه مطلوب اکیداً پیشنهاد می‌شود وضوح تصویر و فاصله‌ی چهره از دوربین به‌گونه‌ای باشد که ابعاد چهره در هر تصویر حداقل ۱۵۰ پیکسل باشد.
    - محدودیت‌های سرویس‌ها برای دریافت فایل‌ها عبارت است از:

- فرمت‌های قابل پشتیبانی برای تصویر عبارتند از: JPG، JPEG، PNG، BMP و TIF
- فرمت‌های قابل پشتیبانی برای ویدئو عبارتند از: WebM، MP4، MOV و AVI
- حداقل ابعاد قابل پذیرش برای تصویر ۱۰۰ پیکسل و برای ویدئو ۳۰۰ پیکسل (برای طول یا عرض) است.
- حداکثر ابعاد قابل پذیرش برای تصویر ۷۰۰۰ پیکسل و برای ویدئو ۲۰۰۰ پیکسل (برای طول یا عرض) است.

۳ -۲ -۱ ویژگی‌های مشترک سرویس‌ها

در این قسمت ویژگی‌های که در خروجی یا ورودی سرویس‌ها مشترک هستند معرفی می‌شوند.

- **میزان حساسیت سامانه:** همه‌ی سرویس‌های هوش مصنوعی مقداری ورودی SensitivityType را از ورودی دریافت می‌نمایند که نشان‌دهنده میزان حساسیت سامانه نسبت به داده ورودی می‌باشد و میزان سخت‌گیری روی تصمیم نهایی است. مقدار VeryHigh به مفهوم بیشترین سخت‌گیری می‌باشد که به طبع در تشخیص‌هایی با بیشینه حساسیت باید از آن استفاده شود. پیشنهاد می‌شود در کاربردهای عادی از میزان سخت گیری Normal استفاده شود. جدول جدول ۳–۱ مقادیر معتبر برای میزان سخت‌گیری را نشان می‌دهد.

جدول ۳–۱ سطوح حساسیت در تشخیص و سخت‌گیری در تصمیم‌گیری

مقدار فیلد	توضیح
VeryLow	حساسیت پایین، سخت‌گیری کم
Low	حساسیت نسبتاً پایین، سخت‌گیری نسبتاً کم
Normal	حساسیت متوسط، سخت‌گیری متوسط
High	حساسیت نسبتاً بالا، سخت‌گیری زیاد
VeryHigh	حساسیت بالا، سخت‌گیری زیاد

- **نتیجه احراز هویت موفقیت آمیز:** در هنگام فراخوانی سرویس‌های احراز هویت در صورت اجرای بدون خطای برنامه، خروجی برنامه شامل عبارت Status است که نشان‌دهنده نتیجه بازشناسی می‌باشد. در صورتی که مقدار این عبارت، Approved باشد به این معنی است که احراز هویت با موفقیت انجام شده است و در اطلاعات ورودی همخوانی وجود دارد. در صورتی که مقدار این عبارت Rejected باشد به این معنی است که احزار هویت موفقیت آمیز نبوده و بین اطلاعات ورودی همخوانی وجود ندارد. در نهایت، در صورت صورتی که نتیجه سامانه به Approved نزدیک باشد ولی شرایط لازم را نداشته باشد مقدار آن برابر OperatorCheck است که نشان می‌دهد مقادیر ورودی دارای همخوانی نسبی هستند و نیاز است تا توسط انسان یا سامانه (با سطح سخت‌گیری متفاوت) دوباره بررسی شوند. جدول ۳–۲ مقادیر مجاز Status را نشان می دهد.

جدول ۳–۲ جزئیات مربوط به وضعیت خروجی و نتیجه

مقدار فیلد	توضیح
Approved	احراز هویت موفق

مقدار فیلد	توضیح
OperatorCheck	نیاز به بررسی اپراتور
Rejected	رد شدن احراز هویت

- **نتیجه احراز هویت در صورت وقوع خطا:** در هنگام فراخوانی سرویس های احراز هویت در برخی موارد به دلیل مشکل در فایل‌های ورودی (ارسال ویدئو به جای تصویر) و یا در دسترسی نبودن سرویس‌های داخلی، خطایی رخ می‌دهد. سامانه در صورت وقوع هر گونه خطا، پیغامی با ساختار زیر با HTTP Status Code مناسب آن خطا بر می‌گرداند (جدول ۳–۳ و جدول ۳–۴).

```
{
  "__unauthorizedRequest": true,
  "__wrapped": true,
  "__traceId": "",
  "error": {
    "errorCode": "USER_NOT_FOUND",
    "message": "User not found.",
    "details": "",
    "source": ""
  }
}
```

جدول ۳–۳ فیلدهای خروجی ناموفق - با بدنه Form-Data

نام فیلد	نوع فیلد	توضیح
__unauthorizedRequest	Boolean	-
__wrapped	Boolean	-
__traceId	String	-
error	Error	-

جدول ۳–۴ شرح فیلدهای شی Error

نام فیلد	نوع فیلد	توضیح
errorCode	String	کد خطا
message	String	شرح خطا
details	String	جزئیات خطا
source	String	منبع خطا

۳ -۲ -۲ تطبیق چهره دو تصویر

در سرویس تطبیق چهره (Face Verification) دو تصویر (یکی مرجعی و دیگری آزمون) با یکدیگر مقایسه شده و یک نتیجه از بین حالات جدول ۳–۲ بازگردانده می‌شود. برای راحتی پیاده‌سازی، این سرویس با دو endpoint با نوع بدنه Form-Data و JSON پیاده‌سازی شده است. جزئیات نحوه‌ی فراخوانی این دو API در جدول ۳–۵ و جدول ۳–۶ آمده است.

جدول ۳–۵ جزئیات نحوه فراخوانی API نطبیق چهره – با بدنه Form-Data

عنوان	توضیح
نوع API	POST
قالب URL	https://[ServerAddress]/api/verification/face-by-image
هدر موردنیاز	ApiToken

جدول ۳–۶ ورودی های سرویس تطبیق چهره - با بدنه Form-Data

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	درجه حساسیت (جزئیات مربوط به این فیلد در جدول ۳–۱ شرح داده شده است)
FirstImage	Byte []	عکس اول
SecondImage	Byte []	عکس دوم

جدول ۳–۷ جزئیات نحوه فراخوانی API نطبیق چهره – با بدنه JSON

عنوان	توضیح
نوع API	POST
قالب URL	https://[ServerAddress]/api/verification/face-by-image-byte-array
هدر موردنیاز	ApiToken

جدول ۳–۸ ورودی های سرویس تطبیق چهره - با بدنه JSON

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	درجه حساسیت (جزئیات مربوط به این فیلد در جدول ۳–۱ شرح داده شده است)
FirstImage	Base64	عکس اول
SecondImage	Base64	عکس دوم

جدول ۳–۹ خروجی موفق سرویس تطبیق چهره

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	درجه حساسیت (جزئیات مربوط به این فیلد در جدول ۳–۱ شرح داده شده است)
status	String	وضعیت (جزئیات مربوط به این فیلد در جدول ۳–۲ شرح داده شده است)

در صورت ناموفق بودن درخواست، خطایی مطابق با بخش «خطاهای عمومی» و «خطاهای تطبیق چهره» برگشت داده می‌شود. لیست این خطاها به‌ترتیب در جدول ۳–۴۰ و جدول ۳–۴۱ قابل ملاحظه هستند.

۳ -۲ -۳ تشخیص زنده بودن (غیرتعاملی)

این سرویس با دریافت یک ویدئو، با بررسی تشخیص زنده بودن passive معتبر بودن یا نبودن ویدئو را به عنوان نتیجه باز می‌گرداند. برای راحتی پیاده‌سازی، این سرویس با دو endpoint با نوع بدنه Form-Data (جدول ۳–۱۰ جدول ۳–۱۱) و JSON (جدول ۳–۱۲ و جدول ۳–۱۳) پیاده‌سازی شده است.

جدول ۳–۱۰ جزئیات نحوه فراخوانی API تشخیص زنده بودن (غیرتعاملی)– با بدنهForm-Data

عنوان	توضیح
نوع API	POST
قالب URL	https://[ServerAddress]/api/verification/passive-liveness
هدر موردنیاز	ApiToken

جدول ۳–۱۱ ورودی های سرویس تشخیص زنده بودن (غیرتعاملی) - با بدنهForm-Data

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	درجه حساسیت (جزئیات مربوط به این فیلد در جدول ۳–۱ شرح داده شده است)
Video	Byte []	ویدئو ضبط‌شده از سمت کاربر

جدول ۳–۱۲ جزئیات نحوه فراخوانی API تشخیص زنده بودن (غیرتعاملی)– با بدنه json

عنوان	توضیح
نوع API	POST
قالب URL	https://[ServerAddress]/api/verification/passive-liveness-byte-array
هدر موردنیاز	ApiToken

جدول ۳–۱۳ ورودی های سرویس تشخیص زنده بودن (غیرتعاملی) - با بدنهJson

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	درجه حساسیت (جزئیات مربوط به این فیلد در جدول ۳–۱ شرح داده شده است)
Video	Base64	ویدئو ضبط‌شده از سمت کاربر

جدول ۳–۱۴ خروجی موفق سرویس تشخیص زنده بودن (غیرتعاملی)

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	درجه حساسیت (جزئیات مربوط به این فیلد در جدول ۳–۱ شرح داده شده است)
status	String	وضعیت (جزئیات مربوط به این فیلد در جدول ۳–۲ شرح داده شده است)

در صورت ناموفق بودن درخواست، خطایی مطابق با بخش «خطاهای عمومی» و « خطاهای تشخیص زنده بودن » برگشت داده میشود.لیست این خطاها در بخش پیوست به‌ترتیب در جدول ۳–۴۰ و جدول ۳–۴۲ قابل ملاحظه هستند.

۳ -۲ -۴ تشخیص زنده بودن (تعاملی: پلک زدن)

برای استفاده از ماژول تشخیص زنده پلک زدن، نیاز است ابتدا یک «الگو»ی تصادفی دریافت کنیم (جدول ۳–۱۵ و جدول ۳–۱۶). این الگو شامل لیستی از اعداد صحیح (به میلی‌ثانیه) است که مشخص می‌کند فرد در چه لحظاتی لازم است پلک بزند. شما به‌عنوان توسعه‌دهنده لازم است با دریافت این لیست در رابطه کاربری خود در لحظات مناسب (به مدت یک ثانیه) از کاربر بخواهید پلک بزند. به عنوان مثال اگر الگوی دریافتی شامل اعداد ۴۵۰۰ و ۸۸۰۰ باشد لازم است در دو بازه زیر از کاربر بخواهید چشم‌های خود را ببندد:

- ۴۵۰۰ میلی‌ثانیه پس از شروع ضبط ویدئو تا ۶۵۰۰ میلی‌ثانیه پس از شروع ضبط
- ۸۸۰۰ میلی‌ثانیه پس از شروع ضبط ویدئو تا ۱۰۸۰۰ میلی‌ثانیه پس از شروع ضبط

سپس لازم است «کد الگوی دریافتی» را به همراه ویدئو ضبط‌شده به ماژول تشخیص زنده بودن ارسال کنید. این ماژول یک نتیجه از بین حالات جدول ۳–۲ را باز می‌گرداند.

- **دریافت الگو:** جزئیات فراخوانی سرویس دریافت الگو و پاسخ آن در دو جدول زیر آورده شده است.

جدول ۳–۱۵ جزئیات فراخوانی سرویس دریافت الگو در تشخیص زنده بودن با پلک زدن

عنوان	توضیح
نوع API	GET
قالب URL	https://[ServerAddress]/api/verification/active-liveness-pattern
هدر موردنیاز	ApiToken

جدول ۳–۱۶ خروجی موفق سرویس دریافت الگو در تشخیص زنده بودن با پلک زدن

نام فیلد	نوع فیلد	توضیح
requestId	string	کد درخواست (مقدار این فیلد باید در مرحله بعدی به همراه ویدئو ارسال شود.)
type	string	نوع. دارای مقدار BlinkTimes
value	string	رشته‌ای است شامل اعداد صحیح که نمایانگر لحظه‌های موردنظر به میلی‌ثانیه است. به‌عنوان مثال اعداد ۴۵۰۰ و ۸۸۰۰ به صورت "4500,8800" بازگردانده می‌شود

- **فراخوانی سرویس تشخیص زنده بودن:** این سرویس با دریافت یک ویدئو، و یک «کد الگو» با بررسی زمان‌های پلک زدن، نتیجه معتبر بودن ویدئو را به عنوان نتیجه باز می‌گرداند. برای راحتی پیاده‌سازی، این سرویس با دو endpoint با نوع بدنه Form-Data (جدول ۳–۱۷ و جدول ۳–۱۸) و JSON پیاده سازی شده است.

جدول ۳–۱۷ جزئیات نحوه فراخوانی سرویس تشخیص زنده بودن پلک زدن – با بدنه Form-Data

عنوان	توضیح
نوع API	POST

عنوان	توضیح
قالب URL	https://[ServerAddress]/api/verification/active-liveness
هدر موردنیاز	ApiToken

جدول ۳–۱۸ورودی های سرویس تشخیص زنده بودن پلک زدن- با بدنه Form-Data

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	درجه حساسیت (جزئیات مربوط به این فیلد در جدول ۳–۱ شرح داده شده است)
RequestId	string	کد درخواست دریافتی از مرحله قبل
Video	Byte []	ویدیو

جدول ۳–۱۹ خروجی موفق سرویس تشخیص زنده بودن پلک زدن

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	درجه حساسیت (جزئیات مربوط به این فیلد در جدول ۳–۱ شرح داده شده است)
status	String	وضعیت (جزئیات مربوط به این فیلد در جدول ۳–۲ شرح داده شده است)

در صورت ناموفق بودن درخواست، خطایی مطابق با بخش «خطاهای عمومی» و « خطاهای تشخیص زنده بودن » برگشت داده میشود.لیست این خطاها به‌ترتیب در جدول ۳–۴۰ و جدول ۳–۴۲ قابل ملاحظه هستند.

۳ -۲ -۵ سرویس تشخیص زنده بودن (تعاملی: تشخیص گفتار)

- برای استفاده از ماژول تشخیص زنده بودن با تشخیص گفتار، نیاز است ابتدا یک «الگو» دریافت کنیم (**دریافت الگو:** جزئیات فراخوانی سرویس دریافت الگو و پاسخ آن در دو جدول زیر آورده شده است.

جدول ۳–۲۰ و جدول ۳–۲۱). این الگو شامل یک جمله کوتاه بوده که کاربر آن را روخوانی می‌نماید. شما به‌عنوان توسعه‌دهنده لازم است با دریافت این جمله در رابطه کاربری خود، از کاربر بخواهید جمله را روخوانی نماید و همزمانی با روخوانی کاربر، تصویر و صدای او را ضبط نمایید. سپس لازم است «کد الگوی دریافتی» را به همراه ویدئو ضبط‌شده به ماژول تشخیص زنده بودن با تشخیص گفتار ارسال کنید. به این ترتیب، ویدویی از کاربر تهیه می‌شود که در آن در حال بازخوانی یک متن است و سامانه با توجه به نحوه گفتار و همچنین متن ارسالی، زنده بودن شخص را تشخیص می‌دهد.

- **دریافت الگو:** جزئیات فراخوانی سرویس دریافت الگو و پاسخ آن در دو جدول زیر آورده شده است.

جدول ۳–۲۰ جزئیات فراخوانی سرویس دریافت الگو در تشخیص زنده بودن با گفتار

عنوان	توضیح
نوع API	GET
قالب URL	https://[ServerAddress]/api/verification/speech-liveness-pattern
هدر موردنیاز	ApiToken

جدول ۳–۲۱ خروجی موفق سرویس دریافت الگو در تشخیص زنده بودن با گفتار

نام فیلد	نوع فیلد	توضیح
requestId	string	کد درخواست (مقدار این فیلد باید در مرحله بعدی به همراه ویدئو ارسال شود.)
type	string	نوع. دارای مقدار SpeechRecognition
value	string	مقدار این قسمت همان جمله‌ای است که کاربر روخوانی می‌کند و همزمان تصویر او ضبط می‌شود .

- **فراخوانی سرویس تشخیص زنده بودن (تعاملی: تشخیص گفتار)**: در جدول‌های زیر بیان شده است.

جدول ۳–۲۲ جزئیات نحوه فراخوانی سرویس تشخیص زنده بودن تعاملی با تشخیص گفتار– با بدنه Form-Data

عنوان	توضیح
نوع API	POST
قالب URL	https://[ServerAddress]/api/verification/speech-liveness
هدر موردنیاز	ApiToken

جدول ۳–۲۳ ورودی های سرویس تشخیص زنده بودن تعاملی با تشخیص گفتار- با بدنهForm-Data

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	درجه حساسیت (جزئیات مربوط به این فیلد در جدول ۳–۱ شرح داده شده است)
RequestId	string	کد درخواست دریافتی از مرحله قبل
Video	Byte []	ویدیو

جدول ۳–۲۴ خروجی موفق سرویس تشخیص زنده بودن تعاملی با تشخیص گفتار

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	درجه حساسیت (جزئیات مربوط به این فیلد در جدول ۳–۱ شرح داده شده است)
status	String	وضعیت (جزئیات مربوط به این فیلد در جدول ۳–۲ شرح داده شده است)

در صورت ناموفق بودن درخواست، خطایی مطابق با بخش «خطاهای عمومی» و « خطاهای تشخیص زنده بودن » برگشت داده میشود. لیست خطاها به‌ترتیب در جدول ۳–۴۰ و جدول ۳–۴۲ قابل ملاحظه هستند.

محدودیت‌ها و ملاحظات سرویس عبارتند از:

- حداقل ابعاد ویدئو ۴۰۰ پیکسل (برای طول یا عرض) است.
- حداکثر ابعاد ویدئو ۲۰۰۰ پیکسل (برای طول یا عرض) است.
- حداقل طول ویدئو یک ثانیه است.
- حداکثر طول ویدئو ۱۵ ثانیه است.

۳ -۲ -۶ تطبیق چهره و تشخیص زنده بودن (غیرتعاملی)

در این بخش به معرفی API مربوط به تطبیق چهره (Face Verification) و تشخیص زنده بودن مبتنی بر تحلیل ویدئو (Liveness Detection) و نحوه استفاده از آن پرداخته می‌شود. این دو ماژول (تطبیق چهره و تشخیص زنده بودن) در قالب یک API به‌صورت یکپارچه قابل دسترسی است (جدول ۳–۲۵، جدول ۳–۲۶، جدول ۳–۲۷، جدول ۳–۲۸ و جدول ۳–۲۹) و در راهکارهای احراز هویت الکترونیکی مانند شناسایی مشتریان از راه دور (e-KYC) مورد استفاده قرار می‌گیرند. چهره استخراج‌شده از تصویر ارسالی به عنوان مرجع با چهره موجود در ویدئو مقایسه می‌شود.

جدول ۳–۲۵ جزئیات نحوه فراخوانی سرویس تطبیق چهره و تشخیص زنده بودن (غیرتعاملی)– با بدنه Form-Data

عنوان	توضیح
نوع API	POST
قالب URL	https://[ServerAddress]/api/verification/video-by-image
هدر موردنیاز	ApiToken

جدول ۳–۲۶ ورودی های سرویس تطبیق چهره و تشخیص زنده بودن (غیرتعاملی)- با بدنه Form-Data

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	درجه حساسیت (جزئیات مربوط به این فیلد در جدول ۳–۱ شرح داده شده است)
Image	Byte []	عکس مرجع جهت تطبیق چهره
Video	Byte []	ویدیو چهت بررسی لایونس و تطبیق چهره

جدول ۳–۲۷ جزئیات نحوه فراخوانی سرویس تطبیق چهره و تشخیص زنده بودن (غیرتعاملی)– با بدنهJson

عنوان	توضیح
نوع API	POST
قالب URL	https://[ServerAddress]/api/verification/video-by-image-byte-array
هدر موردنیاز	ApiToken

جدول ۳–۲۸ ورودی های سرویس تطبیق چهره و تشخیص زنده بودن (غیرتعاملی)- با بدنهJson

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	درجه حساسیت (جزئیات مربوط به این فیلد در جدول ۳–۱ شرح داده شده است)
Image	Base64	عکس مرجع جهت تطبیق چهره
Video	Base64	ویدیو چهت بررسی لایونس و تطبیق چهره

جدول ۳–۲۹ خروجی موفق سرویس تطبیق چهره و تشخیص زنده بودن (غیرتعاملی)

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	درجه حساسیت (جزئیات مربوط به این فیلد در جدول ۳–۱ شرح داده شده است)
status	StatusEnum	وضعیت درخواست
livenessStatus	StatusEnum	وضعیت درخواست تشخیص زنده بودن
verificationStatus	StatusEnum	وضعیت درخواست تطبیق چهره

در صورت ناموفق بودن درخواست، خطایی مطابق با بخش «خطاهای عمومی» و « خطاهای تشخیص زنده بودن» برگشت داده می‌شود. لیست خطاها به‌ترتیب در جدول ۳–۴۰ و جدول ۳–۴۲ قابل ملاحظه هستند.

۳ -۲ -۷ تطبیق چهره و تشخیص زنده بودن تعاملی (تشخیص گفتار)

در این سرویس، از کاربر خواسته می‌شود در یک ویدئو عبارتی مشخص (که به صورت تصادفی انتخاب شده) را بگوید. سپس ویدئو ضبط‌شده به همراه یک تصویر مرجع از چهره شخص به سامانه احراز هویت برای تشخیص زنده بودن و تطبیق چهره ارسال می‌شود. این سرویس وظایف زیر را انجام می‌دهد:

- تشخیص زنده بودن تعاملی تشخیص گفتار در ویدئو: سامانه با دریافت ویدئو و پردازش صدای آن، به صحت گفتار کاربر امتیاز می‌دهد و زنده بودن ویدئو را تایید یا رد می‌کند.
- تشخیص زنده بودن غیرتعاملی در ویدئو
- تطبیق چهره بین برخی فریم‌های ویدئو و تصویر مرجع

برای استفاده از این سرویس، مراحل زیر باید طی شوند:

- دریافت الگو: یک جمله فارسی به همراه یک «کد درخواست» یکتا از سامانه دریافت می‌شود. این جمله از بین جملات ساده و روان در موضوعات مختلف از متن اخبار فارسی استخراج شده است.
- ضبط ویدئو: هنگام ضبط یک ویدئو، باید از کاربر خواسته شود که جمله دریافت‌شده را روخوانی کند.
- ارسال ویدئو و تصویر مرجع: ویدئوی ضبط‌شده و تصویر مرجع به همراه «کد درخواست» دریافتی از درخواست اول، برای تطبیق چهره و همچنین صحت‌سنجی تعاملی و غیرتعاملی به سامانه ارسال می‌شود. این فایل‌ها باید با در نظر گرفتن «ملاحظات و محدودیت‌های عمومی» بیان شده در ابتدای این فصل ارسال شود. در ادامه جزئیات خروجی این درخواست ارائه می‌شود.

نحوه فراخوانی سرویس قسمت «دریافت الگو» در دو جدول زیر ملاحظه می‌شود.

جدول ۳–۳۰ جزئیات فراخوانی سرویس دریافت الگوی تشخیص گفتار

عنوان	توضیح
نوع API	GET
URL	https://[ServerAddress]/api/verification/speech-liveness-pattern
header موردنیاز	ApiToken

جدول ۳–۳۱ قالب خروجی موفق سرویس دریافت الگوی تشخیص گفتار

نام فیلد	نوع فیلد	توضیح
requestId	string	کد درخواست: باید در مرحله بعدی به همراه ویدئو ارسال شود.
type	string	نوع الگو: دارای مقدار ثابت «SpeechRecognition»
value	string	جمله: یک جمله فارسی

همچنین نحوه استفاده از سرویس قسمت «ارسال ویدئو و تصویر مرجع» نیز در جدول‌های زیر مشاهده می‌شود.

جدول ۳–۳۲ جزئیات نحوه فراخوانی سرویس تطبیق چهره و تشخیص زنده بودن تعاملی تشخیص گفتار

عنوان	توضیح
نوع API	POST
URL	https://[ServerAddress]/api/verification/speech-liveness-by-image
header موردنیاز	ApiToken

جدول ۳–۳۳ رودی‌های سرویس تطبیق چهره و تشخیص زنده بودن تعاملی تشخیص گفتار (بدنه Form-Data)

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	سطح حساسیت (مقادیر مجاز در جدول ۳–۱)
RequestId	String	کد درخواست دریافتی از مرحله قبل
Image	File	تصویر مرجع برای تطبیق چهره
Video	File	ویدئو

خروجی موفقیت‌آمیز درخواست دوم قالبی مانند جدول زیر خواهد داشت. این خروجی شامل پاسخ تطبیق چهره و تشخیص زنده بودن تعاملی و غیرتعاملی به صورت جداگانه بوده و در کنار آن یک نتیجه کلی را نیز ارائه می‌کند. در صورت ناموفق بودن درخواست، خطایی با قالب جدول ۳–۳ خواهد داشت. خطاهای ممکن در استفاده از این سرویس در بخش‌های «خطاهای عمومی» (جدول ۳–۴۰)، «خطاهای تطبیق چهره و تشخیص زنده بودن غیرتعاملی» (جدول ۳–۴۱) و «خطاهای تشخیص زنده بودن تعاملی» (جدول ۳–۴۲) از پیوست آورده شده است.

جدول ۳–۳۴ خروجی موفق سرویس تطبیق چهره و تشخیص زنده بودن تعاملی تشخیص گفتار

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	سطح حساسیت (مقادیر مجاز در جدول ۳–۱)
activeLivenessStatus	String	نتیجه تشخیص زنده بودن تعاملی تشخیص گفتار (مقادیر مجاز در جدول ۳–۲)
passiveLivenessStatus	String	نتیجه تشخیص زنده بودن غیرتعاملی در ویدئو (مقادیر مجاز در جدول ۳–۲)
verificationStatus	String	نتیجه تطبیق چهره ویدئو و تصویر مرجع (مقادیر مجاز در جدول ۳–۲)
status	String	نتیجه کلی سرویس (مقادیر مجاز در جدول ۳–۲)

۳ -۲ -۸ تطبیق چهره و تشخیص زنده بودن تعاملی (پلک زدن)

در این سرویس، از کاربر خواسته می‌شود هنگام ضبط ویدئو در زمان‌های مشخصی (که به صورت تصادفی انتخاب شده) پلک بزند. سپس ویدئو ضبط‌شده به همراه یک تصویر مرجع از چهره شخص به سامانه احراز هویت برای تشخیص زنده بودن و تطبیق چهره ارسال می‌شود. این سرویس وظایف زیر را انجام می‌دهد:

- تشخیص زنده بودن تعاملی پلک زدن در ویدئو: سامانه با پردازش تصویر چهره کاربر و بررسی ناحیه چشم، به صحت زمان‌های پلک زدن امتیاز می‌دهد و زنده بودن ویدئو تایید یا رد می‌شود.
- تشخیص زنده بودن غیرتعاملی در ویدئو
- تطبیق چهره بین برخی فریم‌های ویدئو و تصویر مرجع

برای استفاده از این سرویس، مراحل زیر باید طی شوند:

- **دریافت الگو:** یک الگوی پلک زدن (زمان‌های پلک زدن) به همراه یک «کد درخواست» یکتا از سامانه دریافت می‌شود. این الگو که به صورت تصادفی انتخاب می‌شود، شامل چند عدد صحیح است که زمان‌هایی (به میلی‌ثانیه) که کاربر باید پس از آن زمان‌ها پلک بزند را مشخص می‌کند.
- **ضبط ویدئو:** هنگام ضبط یک ویدئو از کاربر، باید با ارائه رابط کاربری مناسب از کاربر خواسته شود که در زمان‌های مناسب پلک بزند. به عنوان مثال اگر الگو به صورت «۵۰۰۰ و ۱۱۵۰۰» باشد، کاربر لازم است در ویدئو دو بار، یکی پس از لحظه ۵ ثانیه (حداکثر تا دو ثانیه پس از آن) و دیگری پس از لحظه ۱۱/۵ ثانیه (حداکثر تا دو ثانیه پس از آن) پلک بزند.
- **ارسال ویدئو و تصویر مرجع:** ویدئوی ضبط‌شده و تصویر مرجع به همراه «کد درخواست» دریافتی از درخواست اول، برای تطبیق چهره و همچنین صحت‌سنجی تعاملی و غیرتعاملی به سامانه ارسال می‌شود. این فایل‌ها باید با در نظر گرفتن «ملاحظات و محدودیت‌های عمومی» بیان شده در ابتدای این فصل ارسال شود. در ادامه جزئیات خروجی این درخواست ارائه می‌شود.

نحوه فراخوانی سرویس قسمت «دریافت الگو» در دو جدول زیر ملاحظه می‌شود.

جدول ۳–۳۵ جزئیات فراخوانی سرویس دریافت الگوی پلک زدن

عنوان	توضیح
نوع API	GET
URL	https://[ServerAddress]/api/verification/active-liveness-pattern
header موردنیاز	ApiToken

جدول ۳–۳۶ قالب خروجی موفق سرویس دریافت الگوی پلک زدن

نام فیلد	نوع فیلد	توضیح
requestId	string	کد درخواست: باید در مرحله بعدی به همراه ویدئو ارسال شود.
type	string	نوع الگو: دارای مقدار ثابت «BlinkTimes»
value	string	شامل اعداد صحیح که با کاما جدا شده‌اند. نمایانگر لحظه‌های موردنظر به میلی‌ثانیه است. به‌عنوان مثال اعداد ۵۰۰۰ و ۱۱۵۰۰ به صورت "5000,11500" بازگردانده می‌شود.

همچنین نحوه استفاده از سرویس قسمت «ارسال ویدئو و تصویر مرجع» نیز در دو جدول زیر مشاهده می‌شود.

جدول ۳–۳۷ جزئیات نحوه فراخوانی سرویس تطبیق چهره و تشخیص زنده بودن تعاملی پلک زدن

عنوان	توضیح
نوع API	POST
URL	https://[ServerAddress]/api/verification/blink-liveness-by-image
header موردنیاز	ApiToken

جدول ۳–۳۸ ورودی‌های سرویس تطبیق چهره و تشخیص زنده بودن تعاملی پلک زدن (بدنه Form-Data)

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	سطح حساسیت (مقادیر مجاز در جدول ۳–۱)
RequestId	String	کد درخواست دریافتی از مرحله قبل
Image	File	تصویر مرجع برای تطبیق چهره
Video	File	ویدئو

خروجی موفقیت‌آمیز درخواست دوم قالبی مانند جدول زیر خواهد داشت. این خروجی شامل پاسخ تطبیق چهره و تشخیص زنده بودن تعاملی و غیرتعاملی به صورت جداگانه بوده و در کنار آن یک نتیجه کلی را نیز ارائه می‌کند. در صورت ناموفق بودن درخواست، خطایی با قالب جدول ۳–۳ خواهد داشت. خطاهای ممکن در استفاده از این سرویس در بخش‌های «خطاهای عمومی» (جدول ۳–۴۰)، «خطاهای تطبیق چهره و تشخیص زنده بودن غیرتعاملی» (جدول ۳–۴۱) و «خطاهای تشخیص زنده بودن تعاملی» (جدول ۳–۴۲) از پیوست آورده شده است.

جدول ۳–۳۹ خروجی موفق سرویس تطبیق چهره و تشخیص زنده بودن تعاملی پلک زدن

نام فیلد	نوع فیلد	توضیح
SensitivityType	String	سطح حساسیت (مقادیر مجاز در جدول ۳–۱)
activeLivenessStatus	String	نتیجه تشخیص زنده بودن تعاملی تشخیص گفتار (مقادیر مجاز در جدول ۳–۲)
passiveLivenessStatus	String	نتیجه تشخیص زنده بودن غیرتعاملی در ویدئو (مقادیر مجاز در جدول ۳–۲)
verificationStatus	String	نتیجه تطبیق چهره ویدئو و تصویر مرجع (مقادیر مجاز در جدول ۳–۲)
status	String	نتیجه کلی سرویس (مقادیر مجاز در جدول ۳–۲)

۳- ۳ کدهای خطا

در صورت بروز خطا در سیستم خطای مناسب آن خطا برگردانده می‌شود. از جمله موارد بروز خطا می توان به ارسال فایل با فرمت اشتباه و یا سایر موارد لیست شده در بخش های بعدی اشاره نمود. در صورت وقوع خطا در سیستم، ساختار خطای زیر با کد وضعیت (HTTP Status Code) مناسب برگردانده می‌شود.


```
{
 "__unauthorizedRequest": true,
 "__wrapped": true,
 "__traceId": "",
 "error": {
  "errorCode": "UNSUPPORTED_IMAGE_FORMAT",
  "message": "Invalid image file, supported formats are JPG,JPEG,PNG,BMP,and TIF.",
  "details": "",
  "source": ""
 }
}
```

- **خطاهای عمومی:** این خطاها به واسطه اضافه کردن یک تصویر یا یک ویدئو ممکن است رخ دهد (جدول ۳–۴۰).

جدول ۳–۴۰ کد خطاها و جزئیات خطاهای عمومی

کد خطا (ErrorCode)	جزئیات (Message)	کد وضعیت	توضیح
INVALID_REQUEST_BODY	Invalid request body, missing: 'image'	400	درخواست نامعتبر است
UNSUPPORTED_IMAGE_FORMAT	Invalid image file, supported formats are JPG, JPEG, PNG, BMP, and TIF.	400	فرمت تصویر پشتیبانی نمی‌شود
UNSUPPORTED_VIDEO_FORMAT	Invalid video file, supported video formats are WebM, MP4, MOV, and AVI.	400	فرمت ویدئو پشتیبانی نمی‌شود
TOO_SMALL_IMAGE_DIMENTIONS	The minimum image size is 100 pixels for both height and width.	400	ابعاد تصویر بسیار کوچک است
TOO_LARGE_IMAGE_DIMENTIONS	The maximum image size is 7000 pixels for both height and width.	400	ابعاد تصویر بسیار بزرگ است
TOO_SMALL_VIDEO_DIMENTIONS	The minimum video size is 300 pixels for both height and width.	400	ابعاد ویدئو بسیار کوچک است
TOO_LARGE_VIDEO_DIMENTIONS	The maximum video size is 2000 pixels for both height and width.	400	ابعاد ویدئو بسیار بزرگ است
TOO_SHORT_VIDEO_LENGTH	The minimum video length is 1 second(s).	400	ویدئو بسیار کوتاه است
TOO_LONG_VIDEO_LENGTH	The maximum video length is 30 seconds.	400	ویدئو بسیار طولانی است
SERVICE_UNAVAILABLE	The server is temporarily unable to service your request due to maintenance downtime or capacity problems. Please try again later.	503	سرویس موقتاً در دسترس نیست
INTERNAL_SERVER_ERROR	The server encountered an internal error and was unable to complete your request. Either the server is overloaded or there is an error in the application.	500	خطایی در سرور رخ داده است

- **خطاهای تطبیق چهره و تشخیص زنده بودن غیرتعاملی:** در صورتی که یکی از وظایف سرویس انجام عملیات تطبیق چهره یا بررسی تشخیص زنده بودن غیرتعاملی باشد، ممکن است یکی از خطاهای جدول ۳–۴۱ رخ دهد.

جدول ۳–۴۱ کد خطاهای تطبیق چهره و تشخیص زنده بودن غیرتعاملی

کد خطا (ErrorCode)	جزئیات (Message)	کد وضعیت	توضیح
TOO_DARK_VIEW	Image is too dark, try again with proper lightning.	400	نور تصویر بسیار کم است
TOO_LIGHT_VIEW	Image is too light, try again with proper lightning.	400	نور تصویر بسیار زیاد است
LOW_QUALITY	Image quality is poor.	400	کیفیت تصویر نامطلوب است
UNACCEPTABLE_BACKLIGHT	Remove backlight and try again.	400	نور پس‌زمینه زیاد است
NO_FACE_DETECTED	Can not detect any face; make sure your face is clearly visible in the image.	400	چهره در تصویر شناسایی نشد
	Can not detect face; make sure your face is clearly visible in every single frame of the the video.	400	چهره در تمام طول ویدئو شناسایی نشد
MULTIPLE_FACE_DETECTED	More than one face detected; record the video again with a plain background.	400	بیش از یک چهره در تصویر شناسایی شد
	More than one face detected; record the video again with a plain background.	400	بیش از یک چهره در ویدئو شناسایی شد
SMALL_FACE_SIZE	The minimum face size is 150 pixels.	400	اندازه چهره در تصویر کوچک است
SMALL_INTERPUPILLARY_DISTANCE	The minimum interpupillary distance is 80 pixels.	400	فاصله بین مرکز دو چشم در تصویر کم است
CROPPED_FACE	Face should appear in center of the frame.	400	چهره در مرکز تصویر قرار نگرفته است

- **خطاهای تشخیص زنده بودن تعاملی:** در صورتی که یکی از وظایف سرویس، تشخیص زنده بودن تعاملی باشد، ممکن است یکی از خطاهای جدول ۳–۴۲ رخ دهد.

جدول ۳–۴۲ کد خطاهای تشخیص زنده بودن تعاملی

کد خطا (ErrorCode)	جزئیات (Message)	کد وضعیت	توضیح
INVALID_ACTIVE_LIVENESS_TOKEN	Pattern token is not valid.	400	کد درخواست؛ نامعتبر است
EXPIRED_ACTIVE_LIVENESS_TOKEN	Your pattern token is expired, you can use your pattern id for 3 minutes.	400	کد درخواست؛ منقضی شده است
USED_ACTIVE_LIVENESS_TOKEN	Your pattern token have used once, please try getting an other pattern id and perform active liveness again.	400	کد درخواست قبلا استفاده شده است

کد خطا (ErrorCode)	جزئیات (Message)	کد وضعیت	توضیح
INVALID_ACTIVE_LIVENESS_OPERATION	The video length is not compatible with pattern token.	400	طول ویدئو مطابق با شرایط اجرای درخواست نیست
	You used this pattern token too early.	400	درخواست؛ سریعتر از حد انتظار ارسال شد

۳- ۴ استعلام از سامانه‌های دیگر

استعلام از سامانه‌های دیگر یک بخش جداناپذیر از یک سامانه احراز هویت غیرحضوری است، استعلام‌هایی مانند دریافت تصویر چهره مرجع برای مقایسه، تطبیق شماره ملی با شماره تلفن همراه، تطبیق اطلاعات بانکی با شماره ملی و ... . همانگونه که بیان شد، در سرویس‌های تطبیق چهره و یا سایر سرویس‌های بر پایه‌ی تطبیق با چهره، یکی از تصاویر ورودی به عنوان عکس چهره مرجع در نظر گرفته می‌شود. با توجه به اینکه سامانه‌های مرجع مانند ثبت احوال تصاویر ثبت شده همه ایرانیان را دارد، می‌توان آن را در سامانه‌های احراز هویت به عنوان مرجع دریافت تصویر مرجع چهره افراد استفاده کرد و با استعلام از آن، تصویر مرجع را دریافت کرد. در سامانه این پروژه امکان استعلام‌های بیرونی از سایر سامانه‌های مرتبط پیش بینی شده است و فراخور نیاز می‌توان از آن استفاده کرد. در سامانه فرماندهی انتظامی کل، می‌توان از سامانه‌های داخلی این فرماندهی مانند گواهینامه، گذرنامه و ... بدین منظور استفاده کرد.

از آنجا که انجام استعلام و سناریوی بررسی استعلام‌ها در سطح منطقی پروژه اتفاق می‌افتد، با بکارگیری سرویس‌های هوش مصنوعی این پروژه در قالب فراخوانی API، از نظر معماری فنی بهتر آن است انجام و بررسی استعلام در سطح سامانه بهره‌بردار صورت پذیرد.

فصل ۴ نصب سرویس‌ها در سرور کارفرما

۴- ۱ نصب و راه‌اندازی سرویس‌ها

همان‌طور که در فصل سه اشاره گردید، ساختار پروژه دارای سه لایه هوش مصنوعی، مدیریت و دمو می‌باشد. برای اجرای هر کدام از لایه‌ها در سرور کارفرما لازم است که مجموعه‌ای از عملیات‌ها اجرا شود. معماری سیستم بر اساس سرویس‌های مستقل و بر بستر داکر طراحی شده است و برای نصب بر روی سرور، به سرورهای لینوکسی نیاز است. به این ترتیب هر سرویس به صورت ایمیج‌های داکر پیاده‌سازی شده‌اند که از مزایای آن می‌توان به افزایش توزیع‌پذیری، استقلال سرویس‌ها نسبت به یکدیگر و محیط دیپلوی اشاره نمود. در حال حاظر سیستم عامل سرور اصلی توزیع Centos بوده که آخرین نسخه داکر بر روی آن نصب شده است.

لایه هوش مصنوعی شامل مجموعه از فایل‌های داکر می‌باشد که به ترتیب با دستورات زیر اجرا می‌شوند. تمام فایل‌های لازم و ایمیج‌های مورد نیاز در قالب یک پوشه در سرور قرار داده شده‌اند و برای نصب بر روی سرورهای جدید نیاز است تا یک نسخه از آن بر روی سرور جدید بارگذاری شود. این پوشه شامل سه فایل زیر است:

- فایل load-images.sh
- فایل docker-compose.yml
- فایل‌هایی با پسوند sepidsystem-dide-services با حجم حداکثر ۵۱۲ مگابایت (ایمیج‌های ذخیره‌شده داکر)

به دلیل اهمیت لایه هوش مصنوعی، روال فعال‌سازی برای آن در نظر گرفته شده است. این روال به ازای نصب بر روی هر سرور باید انجام و لایه هوش مصنوعی بر روی آن سرور فعال شود. به این منظور با استفاده از دستورات زیر یک کلید با فرمت -xxx-xxx

xxx ایجاد می‌شود که باید در اختیار تیم پشتیبانی مجری قرار گرفته تا با توجه به آن یک کد فعال‌سازی تولید شود. کد فعال‌سازی باید در فایل sepidsystem-dide.lic در روت سرور ذخیره شود.

```
chmod +x load-images.sh
/.load-images.sh
```

با انجام روال فوق، فعال‌سازی سیستم تکمیل می‌شود و فقط نیاز به اجرای برنامه می‌باشد که با دستور زیر برنامه اجرا می‌شود.

```
docker-compose up -d
```

لایه مدیریت ترکیبی از چند داکر ایمیج می‌باشد که تمام ایمیج‌ها در پوشه app قرارداده شده‌اند. این لایه نیاز به فعال‌سازی ندارد و صرفا با اجرای دستور زیر فعال می‌شود.

```
docker-compose up -d
```

۴- ۲ گزارش انجام کار

تا کنون سه ورژن مختلف از پروژه طی پنج جلسه برای ناجا نصب شده است. در ادامه به تشریح گزارشی از کارهای انجام شده در هر جلسه پرداخته شده است.

۴ -۲ -۱ اولین جلسه (۲۶ – ۴ - ۱۴۰۰)

برای اولین بار سرور در اختیار مجری قرار گرفت تا برنامه‌های مورد نیاز همچون داکر نصب و ایمیج‌های مورد نیاز دانلود شوند. در ابتدای حضور سرور از سمت ناجا آماده نبود و بخشی از زمان در انتظار آماده‌سازی آن شد. در ادامه به دلیل کندی زیاد در سرعت اینترنت، زمان زیادی برای دانلود برنامه های مورد نیاز شد و در نهایت نسخه اول با سرویس‌های اولیه تطبیق چهره و زنده بودن در اختیار همکاران ناجا قرار گرفت. محل انجام کار ساختمان حکمت است.

۴ -۲ -۲ دومین جلسه (۱۱ - ۷ - ۱۴۰۰)

هدف از این جلسه نصب نسخه جدید و اضافه نمودن سایت دمو بوده است. در این جلسه به دلیل مشکل در اینترنت ناجا نصب به صورت کامل انجام نشد و تصمیم بر این شد تا همه برنامه‌های مورد نیاز به صورت آفلاین و از قبل تنظیم شده آماده شوند. در طی این جلسه اینترنت ناجا برای دریافت برخی از پکیج‌های رایج جهت نصب کاملا قطع بود. محل کار همان ساختمان حکمت بود.

۴ -۲ -۳ سومین جلسه (۲۶ – ۷ - ۱۴۰۰)

به دلیل مشکلات پیش آمده در جلسات قبل و قطعی اینترنت، تمام نیازمندی ها به صورت آفلاین تهیه و بعد از تست توسط تیم فنی مجری در محل توسعه، برای نصب به محل ناجا آورده شد و نصب به درستی انجام گردید. وب‌سایت دمو جهت نمایش سرویی‌های نصب شده در این جلسه به همکاران ناجا تحویل داده شد (مکان: ساختمان حکمت).

۴ -۲ -۴ چهارمین جلسه (۲۸ – ۷ - ۱۴۰۰)

به دلیل محدودیت‌های شبکه و بسته بودن پورت ۱۲۰۰۰ سرور نصب، همکاران ناجا از مرکز ناجا واقع در میدان عطار به وب‌سایت دمو دسترسی نداشتند که با حضور در محل و ارتباط با تیم شبکه ناجا مشکل برطرف گردید (مکان: ناجا – میدان عطار).

۴ -۲ -۵ پنجمین جلسه (۲۴ – ۰۸ - ۱۴۰۰)

با توجه به نیازمندی‌های ناجا، برخی سرویس‌ها و امکانات به وب‌سایت دمو افزوده گردید (مانند دریافت ویدئو) و نصب به صورت کامل انجام شد (مکان: ساختمان حکمت).