



رویه استخراج متن از تصویر و PDF به کمک OCR

(محمد امین کرمعلی)

توجه: ZDR مخفف Zero Data Retention است که به مفهوم عدم ذخیره سازی اطلاعات کاربر بعد از انجام کامل پردازش اطلاعات اشاره دارد.

ABBYY OCR: لینک دانلود نرم افزار

- توضیح کامل و نحوه کار
 - o ABBYY FineReader OCR یک نرم افزار تشخیص کاراکترهای نوری است که اسناد اسکن شده و PDF ها را به متن قابل ویرایش تبدیل می کند. نحوه کار: فایل را آپلود یا اسکن کنید؛ AI ساختار را حفظ و استخراج می کند. تمرکز روی امنیت: سیاست های HIPAA و GDPR، رمزنگاری و عدم ذخیره داده های حساس. لینک دانلود حاوی کرک و همچنین نسخه کرک شده نرم افزار است.
- محدودیت ها
 - o ریسک نشت در نسخه های ابری وجود دارد اگر امنیت تنظیم نشود. نیاز به لایسنس برای ویژگی های پیشرفته. محدودیت در پردازش اسناد بسیار پیچیده. بدون سیاست های صریح ZDR، داده ها ممکن است موقتاً ذخیره شوند.
- پیشنهاد برای تسک ها
 - o برای اسناد حقوقی یا پزشکی، از نسخه محلی استفاده کنید تا نشت جلوگیری شود. در آرشیو سازی، رمزنگاری فعال کنید. برای استخراج داده های شرکتی، با GDPR ادغام کنید. مناسب برای محیط های با الزامات امنیتی بالا.
- نقاط قوت و ضعف
 - o نقاط قوت: تعهد قوی به امنیت AI، سیاست های حفظ حریم خصوصی جامع، رمزنگاری داخلی.
 - o نقاط ضعف: ریسک نشت در حالت ابری، وابستگی به تنظیمات کاربر، عدم گزارش ZDR پیش فرض.

توجه داشته باشید که ابزار زیر برای استفاده یک برنامه نویسی توصیه شده که بتوان با چند خط کد، تعداد انبوهی از تصاویر و فایل PDF را به متن تبدیل کرد. این امکان با دسترسی به API میسر می شود و نسخه تحت وب آن مدنظر نیست. در صورت تمایل می توانید از نسخه تحت وب آن استفاده بفرمایید.

Mistral OCR: لینک دسترسی

- توضیح کامل و نحوه کار
 - o Mistral OCR API یک ابزار هوش مصنوعی برای تشخیص کاراکترهای نوری است که اسناد، تصاویر و PDF ها را پردازش می کند. نحوه کار: داده را آپلود کنید؛ AI متن، جداول و عناصر را استخراج می کند با دقت ۹۷٪-۹۹٪. تمرکز روی امنیت: گزینه self-hosting برای جلوگیری از نشت داده ها و فعال سازی (ZDR) برای عدم ذخیره اطلاعات کاربر.
- محدودیت ها
 - o ریسک نشت اطلاعات وجود دارد، مانند پردازش IP بدون رضایت کامل، که منجر به شکایت به CNIL شده. بدون self-hosting، داده ها به سرورهای ابری ارسال می شود. محدودیت در زبان ها و حجم فایل ZDR. نیاز به درخواست دارد و ممکن است برای همه کاربران فعال نباشد.
- پیشنهاد برای تسک ها
 - o برای استخراج داده های حساس مانند اسناد مالی، از self-hosting استفاده کنید تا نشت جلوگیری شود. در تحلیل اسناد تحقیقاتی، ZDR را فعال کنید. برای RAG مدل ها، داده ها را محلی پردازش کنید. مناسب برای سازمان های با الزامات GDPR.



شرکت مهندسی
تهران فرآیند دنا
بخش مهندسی سواحل و بنادر

تبادل تجارب درون سازمانی

رویه استخراج متن از تصویر و PDF به کمک OCR

• نقاط قوت و ضعف

0 قوت‌ها: دقت بالا

0 ضعف‌ها: نگرانی‌های حفظ حریم خصوصی مانند بهره‌برداری از داده‌ها، وابستگی به درخواست برای ZDR، ریسک در حالت ابری.