

## پروپزال پروژه RAG پرونده‌های حقوقی

موضوع: پیاده‌سازی RAG حقوقی بر روی پرونده‌های قضایی ایران برای مشاوره در پرونده	
<p>کاربرد و اهمیت مساله:</p> <p>۱- ایجاد یک دیتاست تمیز از پرونده‌های قضایی موجود شامل نوع پرونده، شماره پرونده، عنوان پرونده، متن دادنامه</p> <p>۲- کمک به افراد درگیر پرونده‌های قضایی برای آشنایی با پرونده‌های مشابه خود</p> <p>۳- آگاهی بخشی و راهنمایی برای سوال کنندگان موضوعات قضایی پیرامون پرونده‌های قضایی</p>	
دادگان مورد استفاده	<p>۱- دادنامه‌های قضایی</p> <p>برای بدست آوردن منابع فوق می‌توان از استخراج اطلاعات از سایت‌های زیر استفاده کرد:</p> <p><a href="https://ara.iri.ac.ir">https://ara.iri.ac.ir</a></p>
ورودی و خروجی‌های مساله	<p>ورودی: یک پرسش در مورد پرونده یا مسئله قضایی</p> <p>مثال:</p> <p>در معامله یک ماشین بنده محکوم به فروش مال غیر و تحصیل از طریق مال نامشروع شده ام.</p> <p>خروجی: ابتدا پرونده‌های مشابه خروجی از RAG نمایش داده می‌شود و در نهایت برپایه آن پرونده با کمک LLM پاسخ سوال ارائه می‌گردد.</p>
داده ارزیابی	<p>۱- انتخاب چند پرونده به عنوان تست و استخراج مسائل از درون آن و مقایسه جواب LLM براساس خروجی موجود در پرونده</p>
معیار ارزیابی	<p>۱- استفاده از LLM برای شباهت‌یابی خروجی مدل و خروجی صحیح</p> <p>۲- معیار MRR برای مدل بازیابی (و دیگر معیارهای مدل های بازیابی برای سنجش RAG</p>
مدل پیشنهادی	<p>۱- ایجاد مدل‌های بازیابی برای ایجاد RAG بر روی داده‌های استخراج شده</p> <p>۲- کار بر روی پرامپت و استفاده از اصول Prompt Engineering برای بهبود نتایج.</p> <p>۳- استفاده از یک مدل زبانی بزرگ برای تولید متن پاسخ</p>
مدل‌های پایه برای مقایسه	<p>۱- استفاده از مدل زبانی به صورت Zero-Shot بدون استفاده از RAG و Prompt Engineering برای دریافت پاسخ سوال</p>
مساله تولید متن	پاسخ سوال پیرامون مشکل قضایی با استفاده از Rag و Prompt Engineering

استفاده از مدل های زبانی بزرگ	- استفاده از یک مدل زبانی بزرگ لایه باز مناسب برای زبان فارسی برای تولید متن های خروجی برپایه اطلاعات واكشی شده از پایگاه داده و پرامت ایجاد شده
دموی پیشنهادی	استفاده از UI بات های آماده در بستر ماتریکس یا HuggingFace
اکانت گیت هاب نماینده	<a href="#">HamedJahantigh-git</a>
اکانت هاگینگ فیس نماینده	hamedjahantigh