



الف	معرفی طرح	
۱	عنوان طرح پیشنهادی	دانايار، مشاور سازمانی هوشمند و باسابقه‌ی مبتنی بر هوش مصنوعی مولد. danayarai.ir
2	محور طرح پیشنهادی	<ul style="list-style-type: none"> <li>✓ چت ربات‌های هوشمند</li> <li>✓ مشاور هوشمند</li> <li>✓ عامل‌های هوشمند</li> </ul>
3	بیان مساله و راه حل پیشنهادی	<p>روزانه حجم بسیار بالایی از داده در کشور تولید می‌شود. تسلط داشتن و تحلیل کردن این حجم عظیم داده‌ها از یک سو و ذهن محدود انسان برای پردازش آن از سویی دیگر، وجود سامانه‌ای امن را که بر تمام این داده‌ها اشراف داشته باشد و به سازمان‌ها و دولت برای تصمیم‌گیری‌های بهینه کمک کند، ضروری می‌کند. با این حال، هیچ سازمانی سامانه‌ای را ندارد که توانایی پاسخگویی سریع (در کسری از ثانیه) به زبان انسانی و با ارجاع دقیق به داده‌های سازمان را داشته باشد.</p> <p>یک سازمان با اخذ تصمیمات و سیاست‌گذاری‌های جدید بدون تحلیل و در نظر گرفتن اطلاعات قبلی، می‌تواند متحمل هزینه‌های گزافی شود. از طرفی با توجه به حجم عظیم داده‌ها حتی باهوش‌ترین انسان‌ها هم امکان تحلیل و به خاطر سپاری این میزان داده را ندارند. همچنین، این نیاز مبرم به بازیابی اطلاعات و استفاده صحیح از آن‌ها برای مدیریت و پیشبرد تصمیمات سازمان در آینده، نیاز به هزینه‌های اضافی و وقت بیشتری دارد. در حال حاضر سیستمی که پاسخگویی این نیاز به بازیابی صحیح و دقیق این حجم داده و اطلاعات و پاسخگویی به مدیران ارشد سازمانی باشد در سازمان‌ها وجود ندارد. بیشتر امور مربوط به بازیابی اسناد به صورت دستی توسط کارمندان نسبتاً قدیمی سازمان‌ها صورت می‌گیرد. دسترسی به سیستم پاسخگویی امن، صحیح و دقیق که به تمام داده‌ها مسلط است و قادر به تعامل به زبان انسانی است می‌تواند منجر به ایجاد تحول در تصمیمات سازمان شود و باعث شود سازمان متحمل هزینه‌های گزاف ناشی از تصمیمات ناصحیح نشود.</p> <p>«دانايار» یک عامل هوش مصنوعی پیشرفته است که با بهره‌گیری از تکنولوژی‌های پردازش زبان طبیعی و الگوریتم‌های پیشرفته، با تشخیص و استخراج اطلاعات از انواع مختلف اسناد سازمانی مانند فرم‌ها، رسیدها، صورتحساب‌ها، جداول و ... نقش یک مشاور آگاه و باسابقه را که بر تاریخچه سازمان مسلط است، بازی می‌کند. دانايار به واسطه قابلیت‌های پیشرفته تشخیص و استخراج اطلاعات از متن، تصاویر و جداول و به طور کلی داده‌های یک سازمان، یک مشاور هوشمند است که به پرسش‌های مدیران در مورد تاریخچه سازمان با زبان طبیعی و استفاده از نمودارهایی که به صورت هوشمند ترسیم می‌کند، پاسخ می‌دهد و آن‌ها را در تصمیم‌گیری یاری می‌کند. در کنار توانایی‌های پیشرفته در تحلیل و درک داده‌ها، امنیت اطلاعات در دانايار به عنوان یکی از اصول بنیادی طراحی لحاظ شده است. این عامل هوشمند با بهره‌گیری از استانداردهای امنیتی پیشرفته، از داده‌های حساس سازمان در مقابل تهدیدات مختلف محافظت می‌کند. تمامی فرآیندهای پردازش و ذخیره‌سازی اطلاعات در دانايار با استفاده از مکانیزم‌های دسترسی کنترل‌شده انجام می‌شود، به طوری که تنها افراد مجاز می‌توانند به اطلاعات حساس دسترسی داشته باشند. همچنین، دانايار به طور مداوم تهدیدات امنیتی را شناسایی و مدیریت می‌کند تا از هرگونه نفوذ و دسترسی غیرمجاز جلوگیری شود.</p> <p>دانايار به عنوان یک فناوری چند رشته‌ای، قادر است در زمینه‌های مختلفی متخصص باشد. در مقابل، یک انسان معمولاً نمی‌تواند در بیش از یک رشته تخصص داشته باشد و به طور همزمان در تمامی زمینه‌های سیاست‌گذاری و اجرایی موجود تخصص داشته باشد. دانايار از هفت ماژول اصلی تشکیل شده است که عبارت است از:</p> <ul style="list-style-type: none"> <li>• پردازش داده‌های بدون ساختار</li> </ul>



		<ul style="list-style-type: none"> <li>• غنی سازی داده ها</li> <li>• دریافت داده</li> <li>• اتصال گر</li> <li>• کنترل دسترسی</li> <li>• پرس و جوی تعاملی</li> <li>• مصور سازی</li> </ul> <p>جزئیات این ماژول ها در پروپوزال این محصول که به پیوست ارسال شده است، توضیح داده شده است.</p> <p>جدول ۱ زمان بندی توسعه و اجرای پروژه دانایار را نشان می دهد. همان طور که مشاهده می شود، توسعه محصول دانایار طی ده فاز (۱) مطالعات پژوهشی پیرامون طرح، (۲) فهم کسب و کار، (۳) آماده سازی محیط توسعه، (۴) پردازش داده های بدون ساختار، (۵) غنی سازی داده ها، (۶) ذخیره سازی داده ها، (۷) طراحی و پیاده سازی اتصال گر، (۸) طراحی و پیاده سازی ماژول Query، (۹) طراحی و پیاده سازی ماژول Visual Generator و (۱۰) طراحی و پیاده سازی ماژول Access Control انجام خواهد شد. بر اساس فازهای اصلی توسعه و اجرای محصول مطابق جدول ۱ و پیچیدگی زمانی قابلیت های قابل توسعه، یک برآورد زمانی از هزینه توسعه و اجرای محصول صورت گرفته که جزئیات آن در جدول ۲ ارائه شده است.</p>
4	<p>منافع و جنبه های کاربر پذیری</p>	<p>برنامه کاربردی مبتنی بر هوش مصنوعی مولد که:</p> <ul style="list-style-type: none"> <li>• توانایی پردازش انواع مختلف داده (عکس، ویدیو، متن و...) را دارد.</li> <li>• امکان اتصال به انواع مختلف پایگاه داده های سازمان با کمک اتصال گر ها و با استفاده از تکنولوژی مدل های زبانی بزرگ را دارد.</li> <li>• قادر به ارائه داشبوردهای تحلیلی با پرس و جو به زبان طبیعی است.</li> <li>• قادر به جستجو روی محتوای اسناد کاغذی با تکنولوژی OCR است.</li> <li>• امکان پرس و جو را به شکل گفتاری (علاوه بر نوشتاری) و تعامل دوطرفه را (درست شبیه تعامل با یک انسان) فراهم می کند.</li> <li>• قابلیت ارجاع به اسناد و داده های سازمان را دارد.</li> <li>• با درک عمیق داده های سازمان، قادر است تحلیل ارائه دهد.</li> <li>• قابلیت های آن همگام با پیشرفت هوش مصنوعی، بروزرسانی می شود.</li> </ul>
5	<p>تعیین الگوریتم های مورد استفاده</p>	<p>تکنولوژی های استفاده شده در این طرح:</p> <ul style="list-style-type: none"> <li>• پردازش زبان طبیعی</li> <li>• هوش مصنوعی مولد</li> <li>• مدل های زبانی بزرگ</li> <li>• عامل های هوشمند</li> </ul>



					<ul style="list-style-type: none"> <li>• بازیابی اطلاعات</li> <li>• موتور جستجو</li> <li>• گراف دانش</li> <li>• پردازش تصویر (تکنولوژی OCR)</li> <li>• مهندسی prompt</li> </ul>
	6	نرم افزارهای مورد نیاز	<ul style="list-style-type: none"> <li>✓ Python</li> <li>✓ HTML</li> <li>✓ CSS</li> <li>✓ JavaScript</li> <li>✓ Langgraph</li> <li>✓ Langchain</li> <li>✓ Chainlit</li> </ul>		
	7	داده های مورد نیاز برای اجرای تحلیل	<p>دانايار يك مشاور سازمانی است که قادر به تعامل با افراد بر اساس داده هایی است که بر اساس آن ها آموزش دیده است. بنابراین اگر از دانايار سوالی پرسیده شود که داده های مربوط به آن سوال توسط دانايار از قبل مشاهده نشده باشد، دانايار قادر به پاسخگویی نخواهد بود. لیست زیر شامل داده هایی است که می توان به دانايار داد، بدیهی است که به تمام داده های زیر برای تعامل با دانايار نیاز نیست و دانايار قادر است تنها با بخشی از این داده ها نیز به سوالات کاربران راجع به آن داده ها پاسخ دهد:</p> <ul style="list-style-type: none"> <li>• فایل pdf و word</li> <li>• فایل ارائه مانند powerpoint</li> <li>• فایل excel</li> <li>• فایل های اسکن شده</li> <li>• انواع فرمت عکس</li> <li>• فایل صوتی ضبط شده جلسات</li> <li>• پایگاه داده رابطه ای</li> <li>• پایگاه داده غیر رابطه ای</li> <li>• اتوماسیون سازمانی</li> <li>• هر جا که داده با API بتوان دریافت کرد.</li> </ul>		
□ تحقیقاتی	□ افزایش سودآوری	□ افزایش سهم بازار	□ افزایش کیفیت یا ظرفیت	✓ مدیریت و کنترل	□ سایر
	8	زمینه کاربرد	<p>در هر سازمانی، وجود یک مشاور سازمانی بی طرف و متخصص از اهمیت بالایی برخوردار است. مشاوران می توانند با تحلیل دقیق ساختار، فرآیندها و فرهنگ سازمانی، نقاط قوت و ضعف را شناسایی کرده و راهکارهایی عملی برای بهبود عملکرد ارائه دهند. از آنجا که مشاوران از بیرون به سازمان نگاه می کنند، بدون سوگیری های درونی و تعصبات شخصی، دیدگاهی عینی و استراتژیک دارند. این دیدگاه بی طرفانه به مدیریت کمک می کند تا تصمیماتی مبتنی بر واقعیت بگیرند و استراتژی هایی مناسب برای رشد و توسعه سازمان تدوین کنند. همچنین، مشاوران می توانند با انتقال دانش و تجربه خود، به افزایش بهره وری و ایجاد نوآوری در سازمان کمک کنند و از چالش های احتمالی پیشگیری کنند.</p> <p>دانايار، به عنوان یک مشاور سازمانی مبتنی هوش مصنوعی مولد که بر تمامی داده های سازمان مسلط است، می تواند با ارائه تحلیل های دقیق و به موقع، نقش مهمی در کاهش هزینه ها ایفا کند. به دلیل سرعت پردازش بالا و عدم فراموشی، دانايار می تواند روندهای هزینه بر را شناسایی کرده و راهکارهای بهینه برای کاهش آن ها ارائه دهد. همچنین، با تحلیل</p>		



		<p>جامع داده‌ها، از دوباره‌کاری‌ها و هدررفت منابع جلوگیری می‌کند و فرصت‌های بهبود را سریع‌تر از انسان‌ها شناسایی می‌نماید. دانایار با توانایی ارائه پیش‌بینی‌ها و پیشنهادهای هوشمندانه، می‌تواند مدیریت ریسک‌ها را بهبود ببخشد و از صرف هزینه‌های غیرضروری در آینده جلوگیری کند. این سیستم مشاور هوشمند همچنین به تصمیم‌گیری‌های استراتژیک و مبتنی بر داده کمک می‌کند و باعث افزایش بهره‌وری سازمان می‌شود.</p> <p>این ویژگی‌ها دانایار را به یک مشاور هوشمند تبدیل می‌کند که می‌تواند در تمامی زمینه‌های سازمان، از تحلیل داده‌ها تا ارائه راهکارهای استراتژیک، به مدیران کمک کند.</p>
9	نمونه کارهای مشابه	<p>در سال‌های اخیر، توسعه سیستم‌های هوش مصنوعی که به‌طور خاص برای کمک به سازمان‌ها در تحلیل و مدیریت اسناد و داده‌های سازمانی طراحی شده‌اند، به طور قابل توجهی افزایش یافته است. این سیستم‌ها، به سازمان‌ها امکان می‌دهند که از طریق تجزیه و تحلیل داده‌ها و ارائه مشاوره‌های مبتنی بر داده، تصمیم‌گیری‌های بهینه‌تری انجام دهند. برخی از این سیستم‌ها عبارت‌اند از:</p> <p><b>Microsoft 365 Copilot:</b> یک ابزار هوش مصنوعی مبتنی بر پردازش زبان طبیعی است که به‌طور یکپارچه با میکروسافت آفیس ادغام شده و به سازمان‌ها کمک می‌کند تا از داده‌ها و اسناد خود بهره‌برداری بهینه‌تری داشته باشند. Copilot می‌تواند به‌طور خودکار اسناد را بررسی کرده، اطلاعات کلیدی را استخراج کند، خلاصه‌سازی کند و به پرسش‌های کاربران پاسخ دهد. این ابزار به مدیران و کارمندان سازمان‌ها کمک می‌کند تا به سرعت به اطلاعات مورد نیاز دسترسی پیدا کنند و تصمیمات مبتنی بر داده بگیرند.</p> <p><b>IBM Watson Discovery:</b> یک پلتفرم هوش مصنوعی پیشرفته برای جستجو و تجزیه و تحلیل داده‌های سازمانی است. این سیستم به‌طور خاص برای سازمان‌ها طراحی شده است و از تکنیک‌های پردازش زبان طبیعی و یادگیری ماشین برای استخراج اطلاعات ارزشمند از اسناد غیرساختاریافته (مانند ایمیل‌ها، گزارش‌ها، و متون) استفاده می‌کند. این سیستم همچنین می‌تواند داده‌های سازمانی را به صورت خودکار دسته‌بندی کند، الگوهای مهم را شناسایی کند، و به پرسش‌های مدیران سازمان به زبان طبیعی پاسخ دهد.</p> <p><b>Google Cloud AutoML Document AI:</b> یکی از سرویس‌های ارائه‌شده توسط Google Cloud است که به سازمان‌ها امکان می‌دهد مدل‌های هوش مصنوعی سفارشی‌سازی‌شده برای تجزیه و تحلیل و پردازش اسناد سازمانی ایجاد کنند. این سرویس قادر است اطلاعات را از انواع مختلف اسناد (مانند فاکتورها، قراردادهای، و ایمیل‌ها) استخراج کند و داده‌های ساختاریافته تولید کند. این سیستم به‌طور ویژه برای سازمان‌هایی طراحی شده است که نیاز به پردازش حجم بالایی از اسناد دارند و به داده‌های دقیق و تحلیلی برای تصمیم‌گیری‌های تجاری نیاز دارند.</p>
10	سایر موارد	



کاربرگ معرفی طرح  
رویداد دیتاتون بانک سپه

جدول 1. زمان‌بندی توسعه و اجرای محصول دانایار

شماره فاز	شرح فعالیت	ماه تقویمی																	
		۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸
۱	مطالعات پژوهشی پیرامون طرح																		
	مطالعات پژوهشی مدل‌های زبانی بزرگ (LLM)																		
	مطالعه تکنیک‌های تزریق، ذخیره‌سازی و پرس‌وجوهای اسناد																		
	مطالعه تکنیک‌های NLP و بینایی ماشین																		
	ارزیابی تکنیک‌ها و انتخاب مدل‌های برتر																		
۲	فهم کسب‌وکار																		
	شناخت اسناد سازمانی																		
	شناخت سلسله مراتب سازمانی																		
	شناخت داده‌های بدون ساختار و داده‌های ساختارمند سازمان																		
۳	آماده‌سازی محیط توسعه																		
	تامین و پیکربندی سخت‌افزار																		
	پیکربندی محیط توسعه، CI/CD و تست																		
	استقرار مدل‌های زبانی بزرگ (LLM) متن باز																		
	آماده‌سازی داده‌ها																		
۴	پردازش داده‌های بدون ساختار																		
	پیاده‌سازی OCR																		
	تبدیل اسناد کاغذی به نسخه الکترونیکی																		
۵	غنی‌سازی داده‌ها																		
	دسته‌بندی و تگ‌گذاری اسناد																		
	استخراج موجودیت‌های نام‌دار و کلمات کلیدی																		
	استخراج خلاصه اسناد																		
۶	همبسته‌سازی اسناد با مدل گراف‌ی																		
	ذخیره‌سازی (Ingestion) داده‌ها																		
	پیاده‌سازی پایگاه داده برداری																		
	پیاده‌سازی پایگاه داده گراف‌ی																		
۷	پیکربندی موتور جستجوی تمام متن																		
	طراحی و پیاده‌سازی اتصال گر																		
	طراحی و پیاده‌سازی اتصال گر برای SQL																		
	طراحی و پیاده‌سازی اتصال گر برای No-SQL																		
۸	طراحی و پیاده‌سازی اتصال گر برای سامانه‌های مختلف																		
	طراحی و پیاده‌سازی ماژول Query																		
	پیاده‌سازی ماژول پرس‌وجوی متنی مبتنی بر LLM																		
۹	پیاده‌سازی ماژول پرس‌وجوی صوتی مبتنی بر LLM																		
	طراحی و پیاده‌سازی ماژول Visual Generator																		
	پیاده‌سازی ماژول Summarizer																		
	پیاده‌سازی ماژول Goal Explorer																		
۱۰	پیاده‌سازی ماژول Viz Generator																		
	طراحی و پیاده‌سازی ماژول Access Control																		
	شناسایی و استخراج سیاست‌های کنترل دسترسی سازمان																		
	اعمال سیاست‌های کنترل دسترسی به ماژول‌های مختلف																		



جدول 2. تخمین هزینه توسعه و اجرای پروژه دانایار

ردیف	نام فاز	گروه‌های کاری درگیر	هزینه زمانی <sup>1</sup> (نفر ماه)
1	فاز اول: مطالعات پژوهشی پیرامون طرح	۲ نفر متخصص هوش مصنوعی ۲ نفر متخصص بینایی ماشین ۲ نفر متخصص NLP ۱ نفر مدیر پروژه	۵۰
2	فاز دوم: فهم کسب‌وکار	۲ نفر متخصص هوش مصنوعی ۲ نفر متخصص NLP ۴ نفر مشاور سازمان ۱ نفر مدیر پروژه	۴۹
3	فاز سوم: آماده‌سازی محیط توسعه	۲ نفر متخصص هوش مصنوعی ۲ نفر متخصص NLP ۲ نفر Backend Developer ۲ نفر DevOps ۱ نفر مدیر پروژه	۴۹
4	فاز چهارم: پردازش داده‌های بدون ساختار	۲ نفر متخصص هوش مصنوعی ۵ نفر متخصص بینایی ماشین ۲ نفر متخصص NLP ۲ نفر Frontend Developer ۲ نفر Backend Developer ۲ نفر DevOps ۱ نفر مدیر پروژه	۱۴۴
5	فاز پنجم: غنی‌سازی داده‌ها	۲ نفر متخصص هوش مصنوعی ۲ نفر متخصص NLP ۲ نفر Frontend Developer ۲ نفر Backend Developer ۲ نفر DevOps ۱ نفر مدیر پروژه	۵۹
6	فاز ششم: ذخیره‌سازی (Ingestion) داده‌ها	۲ نفر متخصص هوش مصنوعی ۲ نفر متخصص NLP ۲ نفر Frontend Developer ۲ نفر Backend Developer ۲ نفر DevOps ۱ نفر مدیر پروژه	۴۰
7	فاز هفتم: طراحی و پیاده‌سازی اتصال‌گر	۲ نفر متخصص هوش مصنوعی ۲ نفر متخصص NLP	۴۰

<sup>1</sup> این هزینه زمانی همراه با سربار محاسبه شده است.



کاربرگ معرفی طرح  
رویداد دیتاتون بانک سپه

	۲ نفر Frontend Developer ۲ نفر Backend Developer ۲ نفر DevOps ۱ نفر مدیر پروژه		
۵۹	۲ نفر متخصص هوش مصنوعی ۲ نفر متخصص NLP ۲ نفر Frontend Developer ۲ نفر Backend Developer ۲ نفر DevOps ۱ نفر مدیر پروژه	۸ فاز هشتم: طراحی و پیاده‌سازی ماژول Query	
۷۰	۲ نفر متخصص هوش مصنوعی ۲ نفر متخصص NLP ۲ نفر Frontend Developer ۲ نفر Backend Developer ۲ نفر DevOps ۲ نفر گرافیکست ۱ نفر مدیر پروژه	۹ فاز نهم: طراحی و پیاده‌سازی ماژول Visualization	
۴۹	۲ نفر متخصص هوش مصنوعی ۲ نفر متخصص NLP ۲ نفر Frontend Developer ۲ نفر Backend Developer ۲ نفر DevOps ۲ نفر متخصص امنیت ۱ نفر مدیر پروژه	۱۰ فاز دهم: طراحی و پیاده‌سازی ماژول Access Control	
۶۰۹		مجموع	



## فرم (2)

### سایر ملاحظات:

- لطفاً در صورتی که پاسخ به برخی از بندها مقدور نمی باشد آن را خالی بگذارید؛ هر چند تکمیل بودن کاربرگ در نظر هیئت داوران تأثیرگذار خواهد بود.
- فایل خام این کاربرگ بر روی سایت شرکت نوآوران بوم گستر امید (امیدبوم) به نشانی <https://omidinofactory.ir/datathon> قرار دارد. شرکت کنندگان بایستی پس از بارگیری فایل فرم و تکمیل آن ، فایل را در همان سامانه بارگذاری نمایند.
- مهلت ارسال کاربرگ های تکمیل شده 30 آبان ماه 1403 می باشد.

- در صورت وجود مشکل با شماره پشتیبانی سایت شماره 09120913915 تماس حاصل فرمایید و جهت اخذ اطلاعات تکمیلی به کانال روبیکایی با نام



مراجعه فرمایند.

"دیتاتون بانک سپه" با کد QR

اعضای تیم							
نام	نام خانوادگی	سمت سازمانی	محل خدمت	آخرین مدرک تحصیلی	نقش در تیم	شماره تلفن همراه	پست الکترونیکی
۱	رسول	رمضانیان	مدیر	امیدتک	دکتر	راهبر واحد تحقیق و توسعه	ramezani@sharif.ir
۲	سمیه	سلطانی	کارشناس	امیدتک	دکتر	محقق هوش مصنوعی و امنیت	09154176096
۳	محمد رضا	فاتحی نیا	کارشناس	امیدتک	کارشناسی	توسعه دهنده هوش مصنوعی	09156297713
۴	وجیهه	میشری	کارشناس	امیدتک	کارشناسی ارشد	توسعه دهنده هوش مصنوعی	09102381376
۵	یکتا	نصیرپور	کارشناس	امیدتک	کارشناسی ارشد	توسعه دهنده هوش مصنوعی	09357671334
۶	نوید	نیسی	کارشناس	امیدتک	کارشناسی	توسعه دهنده رابط کاربری	09011230864