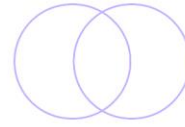


## مهلت ارسال پروپزال

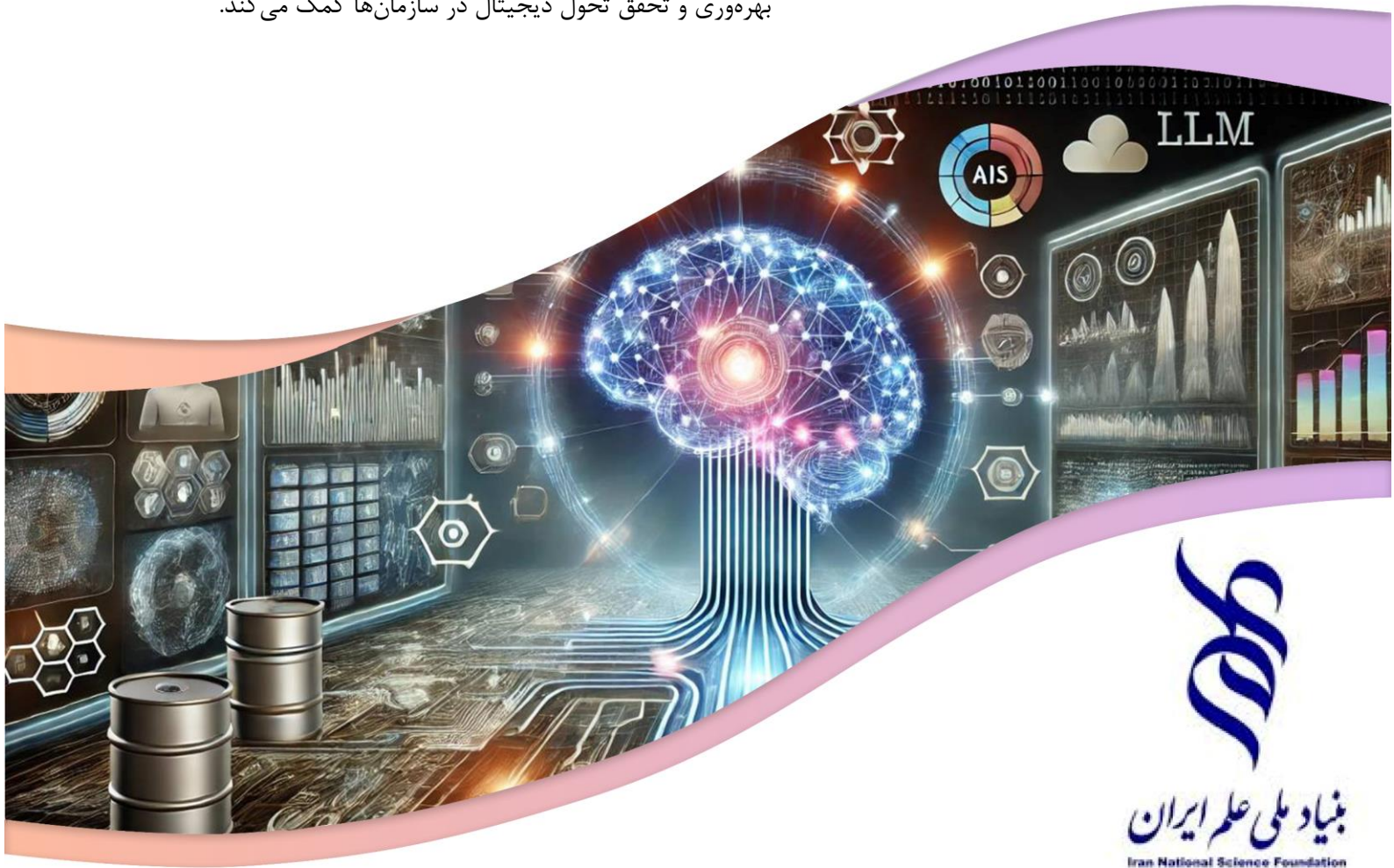


۱۴۰۳/۱۱/۲۹

## فراخوان

### ریزنظیم مدل زبانی بزرگ به منظور تشخیص و طراحی هوشمند نقشه یکپارچگی داده‌های پراکنده فرآیندهای مدیریتی و پشتیبانی سازمانی

این پژوهش بر توسعه یک مدل زبانی بزرگ (LLM) متمرکز است که بتواند داده‌های سازمانی را از منابع متنوع به‌طور خودکار جمع‌آوری، یکپارچه‌سازی و تحلیل کند. مدل پیشنهادی با ترسیم نقشه‌ای جامع از ارتباطات داده‌ها، از فرایندهای مدیریتی و داده‌های سازمانی به‌صورت یکپارچه پشتیبانی می‌کند. خروجی پژوهش شامل یک مدل زبانی بهینه‌شده و سیستم خودکار پردازش داده‌ها خواهد بود که به بهبود بهره‌وری و تحقق تحول دیجیتال در سازمان‌ها کمک می‌کند.



## پیشینه مسئله تحقیقاتی

در سال‌های اخیر، سازمان‌ها با انفجاری از داده‌ها روبرو شده‌اند که اغلب در سیستم‌های مختلف و به صورت پراکنده ذخیره شده‌اند. این پراکندگی داده‌ها چالش‌های مهمی را در زمینه یکپارچه‌سازی، تحلیل و مدیریت اطلاعات ایجاد کرده است. روش‌های سنتی ادغام داده‌ها، مانند فرآیندهای ETL و ساخت API های سفارشی، نه تنها زمان‌بر و پرهزینه هستند، بلکه در مواجهه با حجم زیاد و پیچیدگی داده‌های مدرن نیز ناکارآمد نشان داده‌اند. ظهور مدل‌های زبانی بزرگ (LLM) یک تحول اساسی در این زمینه ایجاد کرده است. این مدل‌ها با توانایی درک و پردازش حجم عظیمی از داده‌های متنی، امکان کشف الگوها و روابط پنهان در داده‌ها را فراهم می‌کنند و به ابزاری قدرتمند برای یکپارچه‌سازی و استخراج دانش از منابع مختلف تبدیل شده‌اند. تحقیقات اخیر بر توسعه معماری‌های جدید ترنسفورمر، آموزش کارآمد آنها با داده‌های سازمانی، و بهینه‌سازی این مدل‌ها برای کارهای خاص تمرکز داشته‌اند.

این پیشنهاد پژوهش بر کاربرد مدل‌های زبانی بزرگ در یکپارچه‌سازی داده‌های سازمانی، به‌ویژه داده‌های مدیریتی و پشتیبانی متمرکز است. این حوزه از داده‌ها اغلب کمتر مورد توجه قرار گرفته‌اند، در حالی که نقش کلیدی در مدیریت و تصمیم‌گیری‌های سازمانی دارند.

هدف نهایی این پژوهش، ایجاد یک پلتفرم جامع و یکپارچه است که سازمان‌ها بتوانند با استفاده از آن، انواع مختلف داده‌های خود را به طور موثر مدیریت، یکپارچه و تحلیل کرده و در نهایت تصمیم‌گیری‌های بهتری اتخاذ کنند. این امر نه تنها به بهبود بهره‌وری و کارایی سازمان‌ها کمک می‌کند، بلکه فرصت‌های جدیدی را برای نوآوری و رقابت‌پذیری در دنیای داده‌محور امروزی فراهم می‌آورد.

## مشروح مسئله تحقیقاتی

در عصر حاضر، سازمان‌ها با حجم فزاینده‌ای از داده‌ها مواجه هستند که در منابع مختلف و سیستم‌های گوناگون پراکنده شده‌اند. این داده‌ها، که شامل اطلاعات مربوط به فرآیندهای مدیریتی، پشتیبانی و عملیاتی می‌شوند، به دلیل تنوع قالب‌ها، ساختارها و منابع، به‌طور کارآمد قابل تحلیل و بهره‌برداری نیستند. این امر، چالش‌های متعددی را برای سازمان‌ها ایجاد می‌کند، از جمله:

- **عدم دید جامع از داده‌ها:** پراکندگی داده‌ها در سیستم‌های مختلف، منجر به عدم وجود دید یکپارچه از وضعیت سازمان می‌شود. این امر، شناسایی الگوها، روندها و روابط پنهان در داده‌ها را دشوار می‌سازد و مانع از تصمیم‌گیری‌های آگاهانه و مبتنی بر داده می‌شود.
- **اتلاف منابع:** تلاش برای یکپارچه‌سازی داده‌ها با روش‌های سنتی، مانند ادغام نقطه به نقطه و توسعه API ها، پرهزینه، زمان‌بر و پیچیده است. این روش‌ها، نیازمند نیروی انسانی متخصص و صرف زمان و منابع زیادی هستند و اغلب با مشکلات نگهداری و به‌روزرسانی مواجه می‌شوند.
- **کاهش بهره‌وری:** عدم وجود داده‌های یکپارچه و قابل دسترس، منجر به کاهش بهره‌وری و کارایی سازمان می‌شود. کارکنان، زمان زیادی را صرف جستجو و جمع‌آوری داده‌ها از منابع مختلف می‌کنند و در نتیجه، زمان کمتری برای انجام وظایف اصلی خود دارند.
- **عدم توانایی تحلیل داده‌ها:** داده‌های پراکنده و ناهمگون، امکان تحلیل جامع و عمیق داده‌ها را از سازمان سلب می‌کند. سازمان‌ها، نمی‌توانند با استفاده از این داده‌ها، بینش‌های ارزشمندی را برای بهبود فرآیندها، افزایش سودآوری و کاهش ریسک‌ها استخراج کنند.
- **چالش‌های تحول دیجیتال:** تحول دیجیتال، نیازمند داده‌های یکپارچه و قابل دسترس است. سازمان‌هایی که نتوانند داده‌های خود را به‌طور موثر مدیریت و تحلیل کنند، در پیاده‌سازی تحول دیجیتال با مشکلات جدی مواجه خواهند شد.
- **محدودیت در مقیاس‌پذیری:** روش‌های سنتی یکپارچه‌سازی داده‌ها، به دلیل پیچیدگی و محدودیت‌های فنی، قابلیت مقیاس‌پذیری ندارند. سازمان‌ها با افزایش حجم و تنوع داده‌ها، با مشکلات جدی در یکپارچه‌سازی داده‌ها روبرو می‌شوند.

با توجه به چالش‌های فوق، نیاز به رویکردی نوین برای یکپارچه‌سازی داده‌های سازمانی احساس می‌شود. مدل‌های زبانی بزرگ (LLM)، به عنوان یک فناوری نوظهور، پتانسیل بالایی برای حل این چالش‌ها دارند. این مدل‌ها قادرند به طور خودکار، نقشه‌ای جامع از ساختار داده‌ها در سازمان ایجاد و داده‌های سازمانی را یکپارچه‌سازی کنند.

## مسئله اصلی این پژوهش:

مسئله اصلی این پژوهش، توسعه یک مدل زبانی بزرگ است که بتواند با دقت و کارایی بالایی، داده‌های سازمانی را از منابع مختلف جمع‌آوری، یکپارچه و تحلیل کند. این مدل، باید قادر باشد تا:

- یکپارچه‌سازی داده‌های سازمانی: به‌طور خودکار، پایگاه‌های داده‌های سازمانی و ارتباط بین آن‌ها را شناسایی کند.
- ترسیم نقشه یکپارچگی داده‌ها: نقشه‌ای جامع از ارتباطات بین داده‌ها در سیستم‌های مختلف سازمان ایجاد کند.
- پشتیبانی از فرآیندهای مختلف: فرآیندهای مدیریتی، پشتیبانی و عملیاتی را به‌طور یکپارچه پشتیبانی کند.
- مقیاس‌پذیری: به‌راحتی با افزایش حجم و تنوع داده‌ها، سازگار شود.

## محدودیت‌های مسئله:

این پژوهش، بر روی داده‌های سطح بالای سازمانی (فرآیندهای مدیریتی) متمرکز است و داده‌های عملیاتی و تخصصی صنایع خرد را پوشش نمی‌دهد. علاوه بر این، توسعه کامل یک سیستم یکپارچه‌سازی داده، فراتر از محدوده این پژوهش است و این پژوهش، تمرکز اصلی خود را بر روی توسعه مدل زبانی بزرگ و ایجاد پایپ‌لاین داده‌ها قرار می‌دهد.

## چالش‌های کلیدی نیاز فناورانه

پروژه توسعه مدل زبانی بزرگ برای یکپارچه‌سازی داده‌های سازمانی، با چالش‌های فناورانه زیر روبرو است:

- توسعه مدل زبانی مناسب: انتخاب و توسعه یک مدل با دقت بالا برای داده‌های متنوع سازمانی، نیازمند دانش تخصصی، داده‌های آموزشی مناسب، ریزتنظیم دقیق و سازگاری با داده‌های مختلف است.
- پایپ‌لاین داده کارآمد: ایجاد یک سیستم خودکار و مقیاس‌پذیر برای جمع‌آوری، پردازش و ذخیره‌سازی داده‌ها از منابع مختلف، چالش‌های مهندسی داده پیچیده‌ای دارد.
- یکپارچه‌سازی با سیستم‌های موجود: سازگاری مدل زبانی بزرگ با سیستم‌های موجود، تضمین امنیت، و ایجاد رابط کاربری آسان، چالش‌های پیاده‌سازی هستند.
- اعتبارسنجی و کیفیت داده: اطمینان از دقت، جامعیت و به‌روز بودن داده‌ها، نیازمند روش‌های اعتبارسنجی قوی است.

## گام‌های تحقیقاتی و الزامات پژوهش

### فاز ۱: مطالعات پایه

این فاز به بررسی جامع ادبیات علمی و فنی موجود در زمینه‌های مدل‌های زبانی و یکپارچه سازی داده‌ها اختصاص دارد. هدف اصلی این فاز، شناسایی دقیق نقاط قوت و ضعف روش‌های موجود و درک عمیق از چالش‌ها و فرصت‌های پیش‌رو می‌باشد. نتایج این فاز به عنوان پایه و اساس برای تصمیم‌گیری‌های بعدی در طراحی و توسعه پلتفرم مورد نظر عمل خواهد کرد.

### فاز ۲: تحلیل و طراحی

در این فاز، پس از بررسی مطالعات پایه، نیازمندی‌های دقیق پروژه مشخص و اهداف قابل اندازه‌گیری تعریف می‌شوند. بر اساس این نیازمندی‌ها، معماری مناسب برای مدل LLM، پایپ لاین داده و سایر اجزای پلتفرم طراحی خواهد شد. این فاز شامل انتخاب تکنولوژی‌های مناسب، تعیین چارچوب‌های کلی توسعه و تعریف معیارهای موفقیت پروژه است.

### فاز ۳: جمع‌آوری داده

این فاز به جمع‌آوری داده‌های مورد نیاز برای آموزش و اعتبارسنجی LLM اختصاص دارد. داده‌های جمع‌آوری شده باید شامل داده‌های سازمانی از منابع مختلف و با فرمت‌های متفاوت باشند. پس از جمع‌آوری داده‌ها، مراحل پیش‌پردازش و آماده‌سازی داده‌ها برای آموزش مدل انجام می‌شود. این مراحل شامل پاکسازی داده‌ها، استانداردسازی فرمت‌ها، و تبدیل داده‌ها به فرمت مناسب برای آموزش مدل است.

### فاز ۴: توسعه مدل

در این فاز، یک مدل مناسب انتخاب شده و به منظور سازگاری با داده‌های سازمانی، مراحل ریز تنظیم و بهینه‌سازی بر روی آن انجام می‌شود. این فاز شامل آموزش مدل با استفاده از داده‌های آماده شده در فاز قبلی، تنظیم دقیق پارامترهای مدل، و بهبود عملکرد مدل در اهداف خاص مورد نظر پروژه است.

### فاز ۵: آزمایش و اعتبارسنجی

پس از توسعه مدل، عملکرد آن در محیط شبیه‌سازی شده و با استفاده از معیارهای دقیق و علمی ارزیابی می‌شود. هدف این فاز، تعیین میزان دقت و کارایی مدل در کارهای مختلف و همچنین شناسایی نقاط ضعف و محدودیت‌های آن است. بر اساس نتایج این فاز، مدل در صورت لزوم مورد بازبینی و بهینه‌سازی بیشتر قرار خواهد گرفت.

## فاز ۶: انتقال دانش

در این فاز، نتایج پژوهش و دانش فنی به دست آمده مستندسازی می شود. همچنین، یک برنامه آموزشی برای انتقال دانش به کاربران نهایی و تیم های پشتیبانی سازمان تدوین می گردد.

### « خروجی پژوهش:

- مدل زبانی بزرگ بهینه شده : مدل زبانی بزرگ آموزش دیده برای فهم ساختار و روابط داده های سازمانی .
- رابط کاربری مدل و انتقال داده از پایگاه داده

### « تسهیم مالکیت فکری

**مالکیت معنوی:** مجری در مالکیت معنوی ناشی از اجرای پژوهش سهیم خواهد بود و انتشار مقاله مشترک توسط مجری و متقاضی در ژورنال های داخلی و خارجی، ارائه مقاله در کنفرانس ها و سمینارها با موافقت و اشاره به نام همه دست اندرکاران مجاز خواهد بود.

**مالکیت مادی:** با توجه به مدل کسب و کار و اجرا و اثبات دستاوردهای حاصل از طرح توسط شرکت متقاضی، منافع مالی ناشی از توسعه این فناوری برای شرکت متقاضی خواهد شد اما مطابق تراضی بین شرکت متقاضی و مجری، قابل اشتراک بین آنها خواهد بود.

### « نحوه پذیرش

پذیرش طرح ها رقابتی است و از بین پروپوزال های دریافتی، موردی که شرایط زیر را داشته باشد، در اولویت خواهد بود:

۱. ترکیب متخصصین تیم پیشنهادی مرتبط باشد.
۲. افراد پیشنهاد شده، دارای سابقه پژوهشی و فنی در آن موضوع باشند.
۳. زمان بندی، هزینه و شرح خدمات، متناسب و مرتبط با پژوهش مورد تقاضا باشد. (در این بخش، مجری می تواند برآورد اولیه خود را اعلام کند اما بدیهی است جزئیات اجرایی در ابتدای امر مشخص نیست و مجری و کارفرما با علم به این موضوع وارد این توافق خواهند شد).
۴. پروپوزال، طبق فرمت پیشنهادی بنیاد، تهیه و از طریق سامانه کاپیر ارسال شده باشد.
۵. فونت حروف و اعداد فارسی B Nazanin و اندازه قلم ۱۳ و فونت حروف و اعداد انگلیسی Times New Roman و اندازه قلم ۱۱ باشد.

## هزینه‌های قابل قبول

- حق‌التحقیق نیروی انسانی
- تجهیزات و سخت‌افزار
- خدمات

## حوزه‌های اولویت‌دار

- تحقیق در عملیات (مدیریت و ریاضی)، علوم کامپیوتر، مهندسی کامپیوتر

## واجدین شرایط

پژوهشگر اصلی تیم لازم است عضو هیئت‌علمی فعال یکی از دانشگاه‌ها و مؤسسات آموزش عالی کشور باشد. تیم مورد نظر در دو حوزه هوش مصنوعی و ساختار پایگاه داده و به ویژه ساختار پایگاه داده سازمانی آشنایی داشته و در زمینه موضوعات سازمانی و مدیریتی کار کرده باشد. بنابراین واجدین شرایط برای انجام این طرح بهتر است از میان موارد زیر باشند: ۱- تحقیق در عملیات (ریاضی و مدیریت) ۲- علوم کامپیوتر ۳- مهندسی کامپیوتر

پس از دریافت پروپوزال از طریق سامانه، ارزیابی انجام گرفته و در صورت کسب امتیاز بالا، تیم برگزیده جهت مذاکره با بنیاد و شرکت متقاضی دعوت خواهد شد.

## تاریخ فراخوان

کلیه افراد واجد شرایط به مدت یک ماه از تاریخ انتشار فراخوان یعنی تا ۲۹ بهمن فرصت دارند که پروپوزال خود را از طریق سامانه کاپیر برای بنیاد ملی علم ایران ارسال نمایند.

## مبلغ حمایت

پژوهش پیشنهاد شده تا سقف ۸۰ درصد، حداکثر ۲/۵ میلیارد تومان، توسط بنیاد ملی علم ایران حمایت خواهد شد. بدیهی است که مابقی هزینه‌ها باید توسط شرکت متقاضی پژوهش تأمین شود.

## شیوه ثبت نام و ارسال درخواست

پژوهشگران گرامی جهت ثبت نام می توانند به سامانه کاپیر مراجعه و از طریق بخش متقاضیان/پژوهشگران اقدام نمایند. در صورتی که در این سامانه پروفایل مشخصات فردی ندارید، ابتدا ثبت نام نموده و سپس به وسیله نام کاربری (Email) و رمز عبور اعطا شده وارد سامانه شوید. پس از ورود در بخش ارسال طرح جدید می توانید از کار تابل پژوهش عمیق شرکت های دانش بنیان اقدام به ارسال طرح نمایید.

## مسئول پاسخگویی

در صورت داشتن هرگونه سوال می توانید با کارگروه دانش بنیان با ایمیل [hatamkhani.a@insf.org](mailto:hatamkhani.a@insf.org) و شماره تلفن ۰۲۱۸۲۱۶۱۳۵۰ (آقای حتم خانی) تماس بگیرید.





بنیاد ملی علم ایران

Iran National Science Foundation

آدرس:

تهران، خیابان کارگر شمالی، بالاتر از تقاطع جلال

آل احمد، کوچه پنجم، پلاک ۳۳

تلفن تماس: ۰۲۱۸۲۱۶۱۱۸۴

وبسایت:

[www.insf.org](http://www.insf.org)