

با حمایت صندوق نوآوری و شکوفایی  
و به سفارش یک شرکت دانش بنیان منتشر می شود:

## فراخوان

توسعه مدل تحلیل ریسک تقلب  
در بیمه شخص ثالث جرحی مبتنی  
بر هوش مصنوعی



مهلت ارسال پروپوزال ها:

۱۴۰۰/۰۸/۰۸

تصادفات ساختگی و به سبب آن پرداخت خسارات جعلی، سالانه حدود ۱۰ درصد خسارت پرداختی شرکت های بیمه را به خود اختصاص می دهد، بنابراین کشف این نوع تقلب ها حائز اهمیت است. کشف موارد تقلب در صنعت بیمه با چالش های زیادی همراه است. کمبود نیروی متخصص و هزینه بالای بازرسی از جمله این چالش ها است. برای حل چالش های ذکر شده می توان از الگوریتم های هوش مصنوعی برای کشف تقلب ها بهره گرفت. روش های یادگیری ماشینی با نظارت، با توجه به نیاز به دادگان برچسب گذاری شده، هزینه بر و زمان بر است. بنابراین در این پروژه تحقیقاتی انتظار می رود مدل تحلیل ریسک تقلب در رشته بیمه شخص ثالث جرحی مبتنی بر فناوری هوش مصنوعی و با استفاده از روش های بدون نظارت توسعه یابد. این مدل بایستی از دقت قابل قبول برخوردار باشد و تعداد مثبت های کاذب که موجب بی اعتمادی کاربران به سیستم می شود، کم باشد.

شرکت در این فراخوان تحقیقاتی و ارائه پروپوزال در قالب انفرادی، گروهی، شرکتی و سازمانی مجاز است.



پروپوزالی که بیشترین تناسب را با الزامات این نیاز تحقیقاتی داشته باشد انتخاب و به عنوان مجری به شرکت دانش بنیان متقاضی معرفی خواهد شد.



## بسمه تعالی

صندوق نوآوری و شکوفایی به منظور تقویت توان توسعه فناوری شرکت‌های دانش‌بنیان با رویکرد نوآوری باز و همکاری فناورانه، خدمت جدیدی را طراحی و عرضه کرده است که در قالب آن، نیازهای تحقیقاتی و فناورانه شرکت‌های دانش‌بنیان و متعاقباً، گروه‌های پژوهشی و فناور توانمند برای اجرای طرح‌های تحقیقاتی و توسعه فناوری‌های مورد نیاز این شرکت‌ها را شناسایی می‌نماید.

آنچه پیش رو دارید، نیاز تحقیقاتی/فناورانه یکی از شرکت‌های دانش‌بنیان متقاضی است که توسط صندوق نوآوری و شکوفایی شناسایی و در قالب فراخوان منتشر شده است. لطفاً به موارد زیر توجه فرمائید:

۱) شرکت در این فراخوان تحقیقاتی و ارائه پروپوزال در قالب انفرادی، گروهی، شرکتی یا سازمانی مجاز است. همه پژوهشگران، دانشجویان، دانش‌آموختگان و اعضای هیئت‌علمی دانشگاه‌ها و مراکز تحقیقاتی، شرکت‌های دانش‌بنیان و فناور و سایر علاقمندان می‌توانند با تدوین و ارسال پروپوزال در این فراخوان شرکت کنند.

۲) پروپوزال‌ها صرفاً باید در چارچوب تدوین شده صندوق نوآوری و شکوفایی و حداکثر تا تاریخ ۸ آبان ماه ۱۴۰۰ در قالب Word در سامانه غزال به آدرس <https://ghazal.inif.ir/grant> ارسال شوند. پروپوزال‌هایی که در چارچوبی غیراز آن، یا به روش‌های دیگر به دست صندوق برسند، وارد فرایند ارزیابی نخواهند شد.

۳) پس از اتمام مهلت ارسال پروپوزال‌ها، فرایند ارزیابی آن‌ها توسط صندوق نوآوری و شکوفایی آغاز خواهد شد. پروپوزالی که بیشترین تناسب را با الزامات این نیاز تحقیقاتی داشته باشد، انتخاب و به‌عنوان «مجری» برای مذاکرات تکمیلی به شرکت دانش‌بنیان متقاضی معرفی خواهد شد.

۴) در صورت توافق پروپوزال‌دهنده منتخب (مجری تحقیق) و شرکت دانش‌بنیان (متقاضی تحقیق)، قرارداد ۳ جانبه‌ای مابین «صندوق»، «متقاضی» و «مجری» منعقد خواهد شد. در قالب این قرارداد، صندوق نوآوری تا ۷۰ درصد هزینه اجرای طرح تحقیقاتی را به شکل بلاعوض به متقاضی خواهد پرداخت تا به‌طور مرحله‌ای و متناسب با پیشرفت اجرای طرح، در اختیار مجری قرار گیرد.

۵) گرچه در این فراخوان، گام‌های کلی برای اجرای تحقیق مورد نظر پیش‌بینی و معرفی شده است، اما پیشنهاددهندگان می‌توانند از هر روش یا فناوری دلخواه و در قالب یک برنامه تحقیقاتی متفاوت برای حل این مسئله تحقیقاتی و دستیابی به اهداف آن استفاده کنند.

۶) تدوین و ارسال پروپوزال در قالب این فراخوان، به‌منزله بهره‌مندی از حمایت‌های صندوق نوآوری و شکوفایی نخواهد بود و برای فرستنده حقی ایجاد نمی‌کند. صندوق نوآوری و شکوفایی خود را ملزم به رعایت محرمانگی دانسته و مفاد کلیه طرح‌های ارسالی محرمانه نزد صندوق باقی خواهد ماند.

۷) هرگونه سؤال یا ابهام در خصوص این فرایند را با شرکت بومرنگ به‌عنوان کارگزار صندوق در میان بگذارید (شماره تماس: ۸۸۳۹۸۵۴۳ و ۸۸۳۹۸۵۶۳-۰۲۱)

## درباره شرکت دانش بنیان متقاضی

این فراخوان به سفارش یک شرکت دانش بنیان نوپای نوع ۲ و فعال در زمینه تحلیل داده در صنعت بیمه تدوین شده است. شرکت متقاضی فعالیت خود را ابتدا در قالب یک تیم استارت‌آپی در سال ۱۳۹۷ آغاز نمود و در سال ۱۳۹۹ به ثبت رسید.

سامانه تحلیل ریسک تقلب، محصول اصلی شرکت است که با تحلیل داده‌های مربوط به پرونده‌های خسارت شرکت‌های بیمه، ریسک تقلب در آن‌ها را تخمین می‌زند. اولین نسخه این سامانه در یکی از شرکت‌های بیمه در ایران مستقر شده است و هم‌اکنون در حال بهبود و آماده‌سازی برای استفاده در سایر شرکت‌ها می‌باشد.

## ضرورت مسئله

طبق آمارهای جهانی، هر شرکت سالانه ۵ درصد سرمایه خود را به علت تقلب از دست می‌دهد که این آمار در شرکت‌های بیمه (در رشته‌های غیر از درمان) به ۱۰ درصد هم می‌رسد. در ایران نیز این میزان در رشته بیمه شخص ثالث و بدنه در حدود ۲۰ درصد اعلام شده است. افزایش سالانه نرخ دیه و انباشت سرمایه، شرکت‌های بیمه را به طعمه مناسبی برای فرصت‌طلبان و متخلفان تبدیل کرده است. از گذشته رویه‌های بازرسی، هنگام ادعای خسارت از سوی بیمه‌گذاران وجود داشته است و کارشناسان با بررسی اسناد و مدارک، صحت و سقم ادعا را مشخص می‌کردند. امروزه با افزایش تعداد بیمه‌گذاران و به سبب آن، افزایش پرونده‌های خسارت و شرکت‌های بیمه‌گر، فرایند بازرسی پیچیده‌تر شده و نیازمند بررسی و تحلیل گسترده‌تر در زمان کم است. همچنین انتقال تجربه از خبرگان صنعت به کارشناسان جدید و توزیع دانش این افراد در شعب شرکت بیمه در سراسر کشور با روش‌های سنتی و پیشین دشوار شده است. علاوه بر این متقلبان شیوه‌های خود را مدام به‌روز می‌کنند و نیاز است دانش متخصصان همپای آن‌ها به‌روز شود.

روش‌های سنتی هزینه زیادی را به شرکت‌های بیمه تحمیل می‌نماید و وابسته به تخصص بازرسی است که در برخی اوقات می‌تواند تحت تأثیر خطای انسانی قرار گیرد. به همین دلیل شرکت متقاضی بر آن شده است که در محصول دستیار بیمه خود، ماژولی با عنوان پایش ریسک تقلب ارائه کند که بدون نیاز به نیروی انسانی و با روشی هوشمند نظیر الگوریتم‌های یادگیری ماشین و هوش مصنوعی، ریسک تقلب پرونده‌های خسارت را با دقت قابل قبولی مشخص و در اختیار کارشناسان قرار دهد.

## مسئله اصلی تحقیق (نیاز تحقیقاتی):

مسئله این تحقیق عبارت است از  
«توسعه مدل تحلیل ریسک  
تقلب در بیمه شخص ثالث جرحی  
مبتنی بر هوش مصنوعی»

## مشروح مسئله تحقیقاتی

هدف ماژول تحلیل ریسک تقلب بیمه شخص ثالث، محاسبه امتیاز ریسک تقلب پرونده‌ها، ارائه هشدار<sup>۱</sup> و اعلام موارد مشکوک بر اساس داده‌های ثبت شده در سامانه‌های عملیاتی شرکت‌های بیمه از جمله اطلاعات حادثه، زیان دیده‌ها و غیره است. از جمله موارد مشکوک به تقلب در پرونده‌ها می‌توان به بروز حادثه در جاده‌های برون شهری و خلوت، تصادف در ساعات پایانی شب و رده سنی یکسان افراد زیان دیده (۳۵ الی ۴۰ سال) اشاره کرد.

از جمله فعالیت‌های انجام شده برای انجام تحلیل ریسک تقلب توسط شرکت متقاضی، می‌توان به موارد زیر اشاره کرد:

- تهیه دیتاست دارای برچسب از داده‌های واقعی شامل ۷۰ پرونده خسارت متقلبانه و ۹۶۹ پرونده خسارت سالم
- استفاده از روش‌های مبتنی بر قاعده و روش‌های تشخیص ناهنجاری آماری نظیر z-score
- استفاده از الگوریتم‌های نظارت شده نظیر SVM، Decision Tree، Random Forest و غیره
- استفاده از روش‌های تحلیل شبکه اجتماعی با هدف شناسایی باندهای تقلب

روش‌های مورد استفاده فعلی قادر به تشخیص موارد متقلبانه هستند، اما نیاز به افزایش نرخ تشخیص صحیح موارد متقلبانه و کاهش مثبت‌های کاذب<sup>۲</sup> وجود دارد. دقت روش‌های استفاده شده بر دیتاست نمونه حدود ۸۰ درصد است اما با توجه به تغییر سناریوهای تقلب، اتکا به روش‌های موجود و انتظار برای باقی ماندن در این دقت درست نیست. از طرفی جمع‌آوری برچسب داده بسیار گران و زمان‌بر است، بنابراین استفاده از رویکردها و روش‌های یادگیری بدون نظارت<sup>۳</sup> مورد توجه قرار گرفته است.

در حال حاضر علاوه بر دیتاست برچسب‌دار اشاره شده در بالا، یک دیتاست شامل ۱۰،۰۰۰ پرونده خسارت بدون برچسب برای تحلیل بدون نظارت در دسترس است. این دیتاست شامل کلیه اطلاعات مربوط به پرونده خسارت است که توسط کارشناسان در سامانه عملیاتی ثبت می‌شود (مانند مشخصات حادثه، مشخصات زیان دیدگان، مشخصات و تعهدات بیمه‌نامه و ...). تعداد کل ویژگی‌ها اعم از خام و محاسبه شده، ۱۴۴ ویژگی شامل ۲۹ ویژگی باینری (مانند آیا راننده‌ی

<sup>۱</sup> Red Flag

<sup>۲</sup> پرونده‌هایی که توسط مدل متقلبانه ارزیابی می‌شوند، در حالی که سالم هستند.

<sup>۳</sup> Unsupervised Learning

مقصر، گواهی معتبر داشته است؟)، ۵۵ ویژگی دسته‌ای (مانند گروه خودرو شامل سواری، موتورسیکلت، بارکش، کشاورزی و راه‌سازی)، ۲۹ ویژگی عددی (مانند تعداد سال‌های عدم خسارت خودرو) و ۳۱ ویژگی ترتیبی (روز حادثه در ماه یا هفته) است. لازم به توضیح است اسناد و مدارک مرتبط به صورت تصویر و متن موجود نیست. همچنین مقادیر مرتبط با ویژگی‌های دسته‌ای گاه تا بیش از ۱۰۰ مقدار، منحصربه‌فرد خواهد بود (به‌طور مثال رنگ خودرو). با توجه به گمنام‌سازی داده‌ها، امکان دسترسی به کد ملی و نام افراد وجود ندارد، اما اطلاعات دموگرافیک افراد شامل جنسیت، سن، ملیت (ایرانی و غیرایرانی) و وضعیت تأهل موجود می‌باشد و سابقه خسارت به صورت تعداد در مجموعه ویژگی‌ها قابل دسترسی است.

در این پروژه انتظار می‌رود مجری تحقیق، مدل تحلیل ریسک تقلب مبتنی بر هوش مصنوعی با استفاده از روش‌های یادگیری ماشین بدون نظارت ارائه نماید. به این صورت که مدل، امتیاز ریسک برای پرونده‌های ورودی را مشخص کند. به‌طور مثال عددی بین ۱ تا ۵ که ۵ بیشترین میزان ریسک است. همچنین هشدارهایی در مورد موارد مشکوک ارائه کند (به‌طور مثال، پرونده به علت ناهنجاری بودن تعداد خسارات زیان‌دیده‌ها مشکوک است)، منظور از هشدار، علائمی است که کاربر انسانی با دیدن آن متوجه شود که چرا این پرونده مشکوک تلقی شده است. ارائه هشدارها به این جهت حائز اهمیت است که کارشناس بیمه با کمک آن‌ها سرنخی برای پیگیری مسئله خواهد داشت. همچنین نیاز است دقت سامانه ارائه‌شده برای پرونده‌های ورودی جدید حداقل ۷۰ درصد و تعداد مثبت‌های کاذب حداکثر ۳۰ درصد باشد. لازم به توضیح است رویکرد مدل، تنها ارائه هشدار و غربالگری پرونده‌های مشکوک است و ارائه ادله و اثبات حقوقی تقلب، خارج از حوزه طرح محسوب می‌شود.

## گام‌های تحقیق:

- پیش‌پردازش و آماده‌سازی داده‌ها (مدیریت ویژگی‌ها) و ارائه گزارش کیفیت داده و استخراج ویژگی‌ها
- انتخاب مدل تحلیل ریسک تقلب (در این مرحله پایگاه داده‌گان برچسب دار شامل ۷۰ پرونده خسارت متقلبانه و ۹۶۹ پرونده خسارت سالم در اختیار تیم قرار می‌گیرد که برای توسعه و ارزیابی‌های اولیه می‌تواند مورد استفاده قرار گیرد)
- بررسی و اندازه‌گیری دقت مدل برای یک مجموعه بدون برچسب از داده‌ها (انتخاب این مجموعه با کمک تیم مجری، شرکت متقاضی و کارشناسان صنعت بیمه)، در این مرحله تنها موارد مشکوکی که توسط تیم ارائه می‌شوند مورد ارزیابی قرار می‌گیرند.
- انجام مراحل ۲ و ۳، حداکثر برای ۳۰۰ پرونده بدون برچسب طی حداکثر ۵ دوره
- ارزیابی خروجی مدل ارائه‌شده با خروجی سه مدل فعلی کارفرما (مبتنی بر قاعده، روش آماری، Random Forest)
- پیاده‌سازی مدل در قالب کد پایتون و ارائه مستند تفصیلی تحلیل و پیاده‌سازی مدل ارائه‌شده
- ارائه مستندات مربوط به طراحی و ساخت مدل مانند نحوه پیش‌پردازش، الگوریتم و پارامترهای مرتبط و غیره

## خروجی‌های مورد انتظار تحقیق

- ارائه مدل تحلیل ریسک تقلب پرونده‌های شخص ثالث با رویکرد بدون نظارت و روش‌های تشخیص ناهنجاری<sup>۴</sup> که امتیاز ریسک تقلب و هشدارهای مرتبط با موارد مشکوک به ازای هر پرونده را مشخص کند.

## الزامات تحقیق

- حداقل دقت مدل ارائه‌شده ۷۰ درصد برای برچسب تقلب به ازای پرونده‌های ورودی جدید
- کاهش تعداد مثبت‌های کاذب برای برچسب تقلب تا حداکثر ۳۰ درصد
- پیاده‌سازی مدل در قالب نرم‌افزار پایتون و ارائه برچسب و هشدارهای خروجی با دریافت ورودی (یک پرونده)



<sup>4</sup> Anomaly Detection

### تجهیزات و زیرساخت‌هایی که متقاضی تحقیق می‌تواند در اختیار مجری قرار دهد

- یک دیتاست شامل ۱۰,۰۰۰ پرونده خسارت جانی شخص ثالث بدون برچسب برای تحلیل بدون نظارت

### راهکارهای غیر جذاب:

- روش‌های مبتنی بر قاعده و فاقد هوشمندی مدنظر نیست. هدف، بهره‌گیری از روش‌های یادگیری ماشین بدون نظارت است.

### گلوگاه‌های احتمالی:

- ارزیابی خروجی مدل‌ها بر ورودی‌های جدید، به دلیل نیاز به تعامل با شرکت‌های بیمه و شعب، فرایندی زمان‌بر است. همچنین به دلیل مشغله زیاد کارکنان شعب و متخصصان حوزه، دسترسی مداوم و درخواست برچسب‌گذاری میسر نیست؛ بنابراین تعداد دفعات ارجاع جهت ارزیابی مدل محدود و زمان‌بر است.

### معیارهای ارزیابی و انتخاب مجری

- تحصیلات و سوابق تیم تحقیقاتی و تناسب آن با مسئله
- رویکرد فنی تیم تحقیقاتی به مسئله
- دسترسی به تجهیزات آزمایشگاهی و مواد اولیه و سایر الزامات اجرای تحقیق
- زمان و هزینه اجرای تحقیق

### تسهیم مالکیت فکری

- مالکیت معنوی: مجری در مالکیت معنوی ناشی از اجرای تحقیق سهیم خواهد بود و انتشار مقاله مشترک توسط مجری و متقاضی در ژورنال‌های داخلی و خارجی، ارائه مقاله در کنفرانس‌ها و سمینارها با موافقت و اشاره به نام همه دست‌اندرکاران مجاز خواهد بود.



- مالکیت منافع مادی: با توجه به مدل کسب و کار شرکت متقاضی، منافع مالی ناشی از توسعه این فناوری تماماً متعلق به شرکت متقاضی بوده و مجری صرفاً حق الزحمه اجرای پروژه تحقیقاتی را دریافت خواهد کرد.

### ارسال پروپوزال

پروپوزال‌ها صرفاً باید در چارچوب موردنظر صندوق نوآوری و شکوفایی، تدوین و حداکثر تا تاریخ ۸ آبان‌ماه ۱۴۰۰ در سامانه غزال به آدرس <https://ghazal.inif.ir/grant> ارسال شوند. پروپوزال‌هایی که در چارچوبی غیراز آن، یا به روش‌های دیگر به دست صندوق برسند، وارد فرایند ارزیابی نخواهند شد.



تهران، میدان ونک، خیابان ملاصدرا، خیابان پردیس، زاینده رود  
شرقی، شماره ۲۴، مجتمع شکوفایی شرکت های دانش بنیان  
کد پستی: ۱۹۹۱۹۱۳۱۱۱  
تلفن: ۰۲۱-۴۲۱۷۰۰۰۰  
پست الکترونیک: info@inif.ir



www.boomerangtt.com

telegram:boomerangtt

insta:boomerangtt.co

۰۲۱-۸۸۳۹۸۵۶۳-۸۸۳۹۸۵۴۳

آدرس: خیابان شریعتی، بالاتر از مطهری، کوچه بینا،  
پلاک ۸، طبقه دوم