

انسان، وب یا هوش مصنوعی؟ مقایسه قابلیت اعتماد و پایداری سه رویکرد بر آورد هزینه پروژه‌های ساختمانی

شایان فلاحي^۱، شایان حجت پناه^۲، سحر علینژاد مجیدی^{۳*}

۱- کارشناسی ارشد، گروه مدیریت پروژه و ساخت، دانشگاه شهید بهشتی، تهران، ایران.

sh.fallahi@mail.sbu.ac.ir

۲- دانشجوی دکتری، گروه مهندسی عمران، دانشکده مهندسی و مدیریت ساخت، دانشگاه تهران، تهران، ایران.

shayan.hojatpanah@ut.ac.ir

۳- استادیار، گروه مهندسی معماری، دانشگاه آزاد اسلامی واحد شاهرود، سمنان، ایران. (نویسنده مسئول)

s.alinejad@iau-shahrood.ac.ir

تاریخ پذیرش: [۱۴۰۵/۳/۱۳]

تاریخ دریافت: [۱۴۰۴/۱۲/۲۵]

چکیده

برآورد دقیق هزینه، از پیش‌نیازهای اساسی برنامه‌ریزی و مدیریت موفق پروژه‌های ساختمانی است. توسعه مدل‌های هوش مصنوعی مولد و سامانه‌های برآورد آنلاین، ضرورت ارزیابی عملکرد این ابزارها در مقایسه با قضاوت خبرگان انسانی را افزایش داده است. با وجود گسترش این فناوری‌ها، تاکنون مقایسه هم‌زمان این سه رویکرد بر مبنای هزینه واقعی یک پروژه ساختمانی گزارش نشده است. این پژوهش با هدف مقایسه عملکرد این سه رویکرد در برآورد هزینه پروژه ساختمانی «آتشگاه» در شهر سمنان انجام شد. داده‌ها از برآوردهای ۲۹ مهندس خبره، هشت سامانه برآورد هزینه و پاسخ‌های پنج مدل هوش مصنوعی مولد ChatGPT، Gemini، Claude، DeepSeek و Grok در ۱۰ تکرار با پرامپت یکسان گردآوری شد. روایی و پایایی ابزار پژوهش با شاخص‌های CVI، CVR و آلفای کرونباخ تأیید شد. عملکرد رویکردها با استفاده از هزینه واقعی پروژه و شاخص‌های خطای مطلق و درصد خطای مطلق میانگین و ضریب تغییرات ارزیابی و تفاوت گروه‌ها با آزمون‌های کروسکال-والیس و من-ویتنی تحلیل شد. نتایج نشان داد خبرگان انسانی با وجود کمترین خطا (۸۰٪/۰)، پراکندگی قابل‌توجهی (۶۴٪/۳۴) داشتند. سامانه‌های آنلاین بیشترین خطا (۳۴٪/۲۰) و کمترین ثبات را نشان دادند. در میان مدل‌های هوش مصنوعی، Grok بهترین عملکرد و Claude ضعیف‌ترین عملکرد را داشت. آزمون کروسکال-والیس تفاوت معناداری بین سه رویکرد نشان داد ($p < ۰/۰۰۱$)، همچنین تحلیل کیفی پاسخ‌ها نشان داد مدل‌های مختلف در میزان توجه به شرایط محلی، کیفیت مصالح و عوامل مؤثر بر هزینه تفاوت قابل ملاحظه‌ای دارند. بر اساس یافته‌ها، رویکرد ترکیبی «انسان-در-چهارچوب» می‌تواند دقت و ثبات برآورد هزینه را بهبود بخشد.

واژگان کلیدی: برآورد هزینه ساخت، خبرگان انسانی، سامانه‌های آنلاین، هوش مصنوعی مولد، قابلیت اعتماد.

۱-مقدمه

پروژه‌های بزرگ ساختمانی معمولاً با هزینه‌های هنگفت و ریسک‌های مالی قابل توجه همراه هستند و هرگونه خطا در برآورد هزینه می‌تواند منجر به شکست پروژه و خسارت‌های جبران‌ناپذیر شود. تخمین دقیق و قابل تفسیر هزینه ساخت همچنان یک چالش اساسی است که به دلیل وابستگی‌های غیرخطی پیچیده، توزیع‌های ناهمگن داده‌ها و عدم قطعیت ذاتی در پارامترهای پروژه است (Chen et al., 2026). برآورد هزینه فرآیندی نظام‌مند برای پیش‌بینی کل هزینه‌های مورد نیاز جهت تکمیل یک پروژه در محدوده تعریف‌شده آن است (Ghimire et al., 2026). پیش‌بینی هزینه پروژه‌های ساختمانی شامل یک سیستم چندبعدی و پویا است و هر دسته‌بندی اصلی را می‌توان به عوامل جزئی‌تر متعددی تقسیم کرد. همچنین، روابط بین متغیرها یک شبکه پیچیده از غیرخطی بودن و تعاملات را نشان می‌دهند که به‌طور جدی بر دقت پیش‌بینی تأثیر می‌گذارند (Liu et al., 2026). برآوردگران همواره از ترکیبی از منطق، عقل سلیم، مهارت، تجربه و قضاوت برای تولید یک برآورد نهایی دقیق استفاده می‌کنند (Rush & Roy, 2001). وجود ذهنیت در فرآیند تدوین برآوردهای هزینه امری اجتناب‌ناپذیر است و استفاده از قضاوت خبرگی چه در مدل‌های پیچیده هزینه و چه در صفحات گسترده ساده، ضرورت دارد (Rush & Roy, 2001). از آنجا که حدود ۷۰ تا ۸۰ درصد از هزینه یک محصول در مرحله مفهوم‌سازی تعیین می‌شود، تصمیم‌گیری اشتباه در این مرحله می‌تواند بسیار پرهزینه باشد (Rush & Roy, 2001)؛ بنابراین، اصلاح خطا در مراحل اولیه بسیار ضروری است و در صورت بروز خطا، اصلاح آن در مراحل بعدی بسیار پرهزینه خواهد بود (Rush & Roy, 2001). سه دسته اصلی تخمین در مدیریت ساخت شامل تخمین مرتبه بزرگ، تخمین مفهومی و تخمین جزئیات است (Yu et al., 2006). تخمین مفهومی در مراحل اولیه پروژه که اطلاعات طراحی محدود است، بیشترین کاربرد را دارد (Ghimire et al., 2026). نوسانات قیمت مصالح ساختمانی به‌عنوان یکی از عوامل اصلی نا اطمینانی در برآورد هزینه، ریسک‌های قابل توجهی را برای بودجه‌بندی و تحویل پروژه ایجاد می‌کند (Lyu et al., 2025). پیچیدگی ناشی از تنوع و حجم عظیم داده‌های درگیر در پروژه‌های ساختمانی، به تکنیک‌های پیش‌بینی پیشرفته نیاز دارد و مدل‌های سنتی اغلب در مواجهه با چنین مجموعه داده‌های پیچیده‌ای برای دستیابی به دقت مطلوب با مشکل مواجه می‌شوند (Shi & Shide, 2025).

روش‌های سنتی برآورد هزینه شامل تخمین مرتبه بزرگ، تخمین مفهومی و تخمین جزئیات است که هر یک بر اساس میزان اطلاعات موجود و دقت مورد نیاز انتخاب می‌شوند (Tochaiwat & Budda, 2025). تغییرات طراحی در پروژه‌های ساختمانی یکی از عوامل اصلی هزینه‌های اضافی است و مطالعات نشان داده است که این تغییرات می‌توانند هزینه‌ها را بین ۵ تا ۲۰ درصد از بودجه کل ساخت افزایش دهند (Ahn et al., 2024). در سال‌های اخیر، روش‌های سنتی با روش‌های نوین مبتنی بر هوش مصنوعی و سامانه‌های برخط تکمیل شده‌اند (Sadikoglu & Demirkesen, 2025) و در سه رویکرد اصلی قابل دسته‌بندی هستند: خبرگان انسانی، سامانه‌های آنلاین و مدل‌های هوش مصنوعی مولد. روش سنتی قضاوت خبرگی علیرغم کاربرد گسترده، با چالش‌هایی نظیر ذهنی بودن، ریسک خطای بالا، وابستگی شدید به سطح تجربه و عدم شفافیت در فرآیند استدلال مواجه است (Rush & Roy, 2001). سامانه‌های آنلاین برآورد هزینه که مزایایی نظیر سرعت و دسترسی آسان دارند، از محدودیت‌هایی نظیر عدم به‌روزرسانی مناسب و عدم توجه به شرایط محلی رنج می‌برند (He et al., 2024). در مقابل، مدل‌های بزرگ زبانی با قابلیت درک زبان طبیعی و استدلال چندمرحله‌ای، پتانسیل بالایی برای خودکارسازی فرآیند برآورد هزینه نشان داده‌اند (Abdelsalam et al., 2026). این مدل‌ها می‌توانند ورودی‌های کاربر را تفسیر کرده و خروجی‌های منسجم و مرتبط با زمینه تولید کنند (Ghimire et al., 2026). با این حال، این مدل‌ها با چالش پدیده توهم مواجه هستند که می‌تواند منجر به تولید پاسخ‌های به ظاهر منطقی اما نادرست شود (Abdelsalam et al., 2026).

علیرغم پیشرفت‌های اخیر در حوزه هوش مصنوعی و سامانه‌های برخط، شکاف قابل توجهی در ادبیات پژوهشی مدیریت ساخت وجود دارد. ابزارهای برنامه‌ریزی موجود، عمدتاً بر زمان‌بندی و تخصیص منابع متمرکز هستند و محدودیت‌های مالی و جریان نقدی را به‌طور کامل در نظر نمی‌گیرند (Liu et al., 2026). تا کنون مطالعه جامعی که به صورت هم‌زمان به مقایسه خبرگان انسانی،

سامانه‌های آنلاین و مدل‌های مختلف هوش مصنوعی مولد در یک پروژه واقعی ساختمانی پردازد، انجام نشده است (Jafary et al., 2026). پژوهش‌های موجود عمدتاً بر روی یکی از این رویکردها متمرکز بوده و از ارائه یک چارچوب مقایسه‌ای کامل ناتوان مانده‌اند (Liu et al., 2026). یکی از چالش‌های اساسی در پروژه‌های ساختمانی، وجود تبانی در مناقصات است که منجر به تعیین قیمت‌های غیرواقعی و خسارت به کارفرمایان می‌شود (Kim et al., 2024). به طور خاص، موضوع پایداری پاسخ‌های مدل‌های هوش مصنوعی در تکرارهای متعدد با پرامپت ثابت (تکرارپذیری) کمتر مورد توجه پژوهشگران قرار گرفته است (Saeidlou & Ghadiminia, 2024). همچنین، نگرش و میزان اعتماد مهندسان شاغل در صنعت ساخت نسبت به استفاده از این ابزارهای نوین به ندرت بررسی شده است (Ghimire et al., 2026).

بر اساس شکاف‌های شناسایی شده، این پژوهش چهار هدف اصلی را دنبال می‌کند. اولین هدف، ارزیابی تفاوت میانگین برآوردهای سه رویکرد خبرگان انسانی، سامانه‌های آنلاین و مدل‌های هوش مصنوعی مولد در پروژه آتشگاه است. دومین هدف، سنجش دقت هر رویکرد بر اساس شاخص‌های خطای مطلق و درصد خطای مطلق میانگین نسبت به هزینه واقعی پروژه و همچنین اندازه‌گیری پایداری آن‌ها از طریق ضریب تغییرات است. سومین هدف، سنجش تکرارپذیری پاسخ‌های پنج مدل هوش مصنوعی در ۱۰ بار تکرار با یک پرامپت ثابت است. چهارمین و آخرین هدف، بررسی نگرش مهندسان نسبت به این ابزارها و تعیین میزان اعتماد آن‌ها به خروجی سامانه‌های آنلاین و هوش مصنوعی مولد است؛ چراکه پذیرش عملی این فناوری‌ها، علاوه بر عملکرد فنی، به عوامل انسانی و ادراکی نیز وابسته است.

۲- مرور مبانی نظری و پیشینه

۲-۱- روش‌های سنتی برآورد هزینه

روش‌های سنتی برآورد هزینه ساخت، به طور کلی در سه دسته اصلی قضاوت خبرگی^۱، تخمین پارامتریک^۲ و تخمین قیاسی یا تمثیلی^۳ طبقه‌بندی می‌شوند (Rush & Roy, 2001). هر یک از این روش‌ها بر اساس میزان اطلاعات موجود در دسترس، زمان مورد نیاز برای برآورد و سطح دقت مطلوب، در مراحل مختلف چرخه حیات پروژه مورد استفاده قرار می‌گیرند (Rush & Roy, 2001) (Kim et al., 2004).

قضاوت خبرگی، رایج‌ترین و در عین حال بحث‌برانگیزترین روش برآورد هزینه است. این روش بر این اصل استوار است که برآوردگر با تکیه بر دانش، تجربه و درک شهودی خود از پروژه‌های مشابه در گذشته و با مراجعه به داده‌های هزینه پروژه‌های تاریخی و اعمال تعدیلات افزایشی یا کاهش‌ی بر اساس تفاوت‌های پروژه جدید، هزینه نهایی را پیش‌بینی می‌کند. با وجود آنکه قضاوت خبرگی به تنهایی یک تکنیک برآورد هزینه به شمار نمی‌رود و فاقد ساختار ریاضی مشخصی است، اما به عنوان یک ابزار ضروری و مکمل برای تولید برآوردهای معنادار، به طور گسترده در صنعت مورد استفاده قرار می‌گیرد و در بسیاری از موارد، تنها روش ممکن در مراحل بسیار اولیه پروژه است (Rush & Roy, 2001).

با این حال، این روش به شدت مستعد سوگیری‌های شناختی و خطاهای انسانی است و حتی سه خبره با اطلاعات شروع یکسان، برآوردهای متفاوتی ارائه خواهند داد. از مهم‌ترین محدودیت‌های آن می‌توان به ذهنی بودن فرآیند، ریسک خطای بالا، عدم ثبات در نتایج در تکرارهای مختلف، نبود فرآیندی ساختاریافته و شفاف و وابستگی شدید به سطح تجربه، تخصص و حتی روحیه برآوردگر

1 Expert Judgment

2 Parametric Estimating

3 Analogous Estimating

در لحظه انجام برآورد اشاره کرد (Rush & Roy, 2001). این محدودیت‌ها باعث شده است که پژوهشگران همواره به دنبال روش‌های مکمل و ساختاریافته‌تری برای کاهش اثرات منفی قضاوت خبرگی باشند (Kim et al., 2004).

در مقابل، تخمین پارامتریک که از دهه ۱۹۷۰ در صنعت ساخت رواج یافته است، بر پایه روابط ریاضی و آماری بین پارامترهای تأثیرگذار (مانند مساحت زیربنا، تعداد طبقات، کاربری ساختمان، نوع اسکلت و ...) با هزینه نهایی پروژه استوار است (Kim et al., 2004). در این روش، با استفاده از داده‌های تاریخی و تکنیک‌های رگرسیونی، یک تابع ریاضی برازش داده می‌شود که بتواند هزینه را بر اساس پارامترهای ورودی پیش‌بینی کند. هرچند این روش دارای پایه ریاضی تعریف‌شده‌ای است و نسبت به قضاوت خبرگی عینی‌تر عمل می‌کند، اما روش‌هایی مانند رگرسیون چندمتغیره خطی برای توصیف روابط غیرخطی، چندبعدی و پیچیده حاکم بر هزینه‌های ساخت در دنیای واقعی مناسب نیستند و می‌توانند منجر به برآوردهای نادرست شوند. به همین دلیل، استفاده از این روش نیازمند داده‌های تاریخی فراوان و باکیفیت و همچنین درک عمیق از روابط بین متغیرها است (Kim et al., 2004).

تخمین قیاسی (یا تمثیلی) نیز روشی است که در آن برآوردگر با شناسایی پروژه‌های مشابه تاریخی از نظر ابعاد، کاربری، کیفیت و موقعیت جغرافیایی و تطبیق پارامترهای آن‌ها با پروژه جدید، سعی در پیش‌بینی هزینه دارد. این روش اغلب در مراحل اولیه پروژه و زمانی که اطلاعات طراحی محدود است، استفاده می‌شود و دقت آن به شدت به شباهت پروژه‌های مبنا با پروژه جدید بستگی دارد (Rush & Roy, 2001). نوسانات شدید قیمت مصالح ساختمانی در سال‌های اخیر، به عنوان یکی از عوامل اصلی نا اطمینانی در برآورد هزینه، ریسک‌های قابل توجهی را بر بودجه‌بندی و تحویل پروژه‌ها تحمیل کرده و اعتماد به روش‌های سنتی را که اغلب مبتنی بر داده‌های تاریخی با قیمت‌های ثابت هستند، با چالش جدی مواجه ساخته است (Lyu et al., 2025).

۲-۲- سامانه‌های آنلاین برآورد هزینه

سامانه‌های آنلاین برآورد هزینه با هدف اصلی ارائه دسترسی سریع، آسان و در عین حال با هزینه کمتر (نسبت به استخدام مشاوران اختصاصی) به داده‌های هزینه به‌روز و استاندارد، توسعه یافته‌اند (He et al., 2024). این سامانه‌ها معمولاً مبتنی بر پایگاه‌های داده هزینه مرکزی و معتبر مانند RSMeans، Craftsman و Sage طراحی می‌شوند که به‌عنوان مرجعی برای برآوردگران در سراسر جهان شناخته می‌شوند (Ghimire et al., 2026). در صنعت معماری، مهندسی و ساخت^۱، مدل‌سازی اطلاعات ساختمان^۲ به‌عنوان یک فناوری تحول‌آفرین و انقلابی ظاهر شده است که امکان استخراج خودکار مقادیر مصالح و اجزای ساختمان را مستقیماً از مدل سه‌بعدی فراهم می‌کند (Ahn et al., 2024). این قابلیت، پتانسیل بالایی برای کاهش خطاهای انسانی و افزایش سرعت فرآیند برآورد دارد (Ahn et al., 2024).

با این حال، شکاف‌های قابل توجهی در نحوه اتصال یکپارچه داده‌های مدل‌سازی اطلاعات ساختمان به فرآیندهای برآورد هزینه وجود دارد (Ahn et al., 2024). به‌طور خاص، فقدان استانداردسازی در ساختار داده‌ها و همچنین نبود اتوماسیون کامل در این فرآیند، منجر به تکرار تلاش‌ها، حذفیات سهوی، تطابق نداشتن واحدهای اندازه‌گیری و در نهایت، کاهش دقت و افزایش زمان برآورد می‌شود (Ghimire et al., 2026). این مسائل نشان می‌دهد که صرفاً وجود داده‌های باکیفیت در مدل BIM برای یک برآورد دقیق کافی نیست و نیاز به یکپارچگی هوشمندانه‌تری بین این دو حوزه احساس می‌شود (Ahn et al., 2024) (Ghimire et al., 2026).

سامانه‌های آنلاین علیرغم مزایایی نظیر سرعت بالا، هزینه کمتر و دسترسی آسان، از محدودیت‌های جدی و ساختاری رنج می‌برند. از جمله این محدودیت‌ها می‌توان به عدم به‌روزرسانی به موقع داده‌ها متناسب با نوسانات روزانه و هفتگی بازار، عدم توجه کافی به شرایط محلی و اقلیمی هر منطقه و نادیده گرفتن نوسانات لحظه‌ای و بعضاً غیرمنتظره قیمت مصالح اشاره کرد (He et al., 2024). برای مثال، نوسانات قیمت مصالح ناشی از عواملی مانند تورم، نرخ بهره، تحریم‌ها و عدم تعادل عرضه و تقاضا، پیش‌بینی هزینه را با

1 Architecture Engineering Construction (AEC)

2 Building Information Modeling (BIM)

چالش‌های جدی مواجه می‌کند و این سامانه‌ها به ندرت این متغیرهای پویا و پیچیده را در محاسبات خود لحاظ می‌کنند (Lyu et al., 2025). قیمت‌های واحد در بازار متغیر هستند و دانش تخمینی که در گذشته آموخته شده و مبتنی بر قیمت‌های ثابت است، ممکن است برای پروژه‌های آینده به راحتی قابل استفاده نباشد (Yu et al., 2006). شاخص‌های هزینه ساخت، به عنوان معیارهای وزنی تجمیعی که تغییرات هزینه مواد، نیروی کار و تجهیزات را در طول زمان ردیابی می‌کنند، می‌توانند تا حدودی این نوسانات را منعکس کنند، اما به ندرت در سامانه‌های آنلاین به صورت پویا و بلادرنگ مورد استفاده قرار می‌گیرند (Jafary et al., 2025).

۲-۳- هوش مصنوعی مولد در ساخت و ساز

در سال‌های اخیر، پژوهش‌های متعددی به بررسی عوامل کلیدی مؤثر بر دقت تخمین هزینه پروژه‌ها پرداخته‌اند. به عنوان مثال، پژوهشگران با استفاده از تکنیک‌های تحلیل شبکه‌ای^۱ و تحلیل سلسله مراتبی^۲ به این نتیجه رسیده‌اند که معیارهایی مانند «اجرای استاندارد برای تعیین اطلاعات هزینه» با ضریب اهمیت ۲۶.۹۸ درصد و «روش استفاده شده در تهیه تخمین» با ضریب اهمیت ۲۲.۲۵ درصد، بیشترین تأثیر را در دقت تخمین اولیه هزینه دارند (Al-Azzawi, 2016). این یافته‌ها به وضوح نشان می‌دهد که انتخاب روش برآورد و ساختار فرآیند، به اندازه خود داده‌های هزینه، در دقت نهایی تأثیرگذار است (Al-Azzawi, 2016). استفاده از تکنیک‌های پردازش زبان طبیعی^۳ و یادگیری ماشین برای تحلیل داده‌های متنی و عددی تغییرات طراحی، می‌تواند دقت پیش‌بینی هزینه را به طور قابل توجهی افزایش دهد (Ahn et al., 2024).

از سوی دیگر، پژوهش‌های اخیر به طور فزاینده‌ای نشان می‌دهند که روش‌های یادگیری ماشین و هوش مصنوعی به دلیل توانایی بالای خود در شناسایی الگوهای پیچیده و تولید پیش‌بینی‌های دقیق، به عنوان ابزارهای کارآمد و قدرتمندی در برآورد هزینه پروژه‌های ساختمانی ظهور کرده‌اند (Shi & Shide, 2025). مدل‌های یادگیری عمیق با قابلیت‌های قدرتمند پردازش داده و انعطاف‌پذیری خود، به عنوان فناوری‌های حیاتی برای بهبود دقت پیش‌بینی‌های هزینه ظاهر شده‌اند (Shi & Shide, 2025). در این میان، هوش مصنوعی مولد^۴ و به طور خاص مدل‌های بزرگ زبانی^۵ توجه بسیاری از پژوهشگران حوزه مدیریت ساخت را به خود جلب کرده است (Abdelsalam et al., 2026). این مدل‌ها با قابلیت استثنایی خود در درک زبان طبیعی، استدلال چندمرحله‌ای، تولید کد و تحلیل متون تخصصی، قادر به اجرای وظایف پیچیده و چندوجهی هستند که پیش از این تنها از عهده خبرگان انسانی بر می‌آمد (Ahn et al., 2024).

مدل‌های مولد هوش مصنوعی بر اساس مکانیسم‌های تولید و ساختار داخلی خود، به دسته‌های مختلفی از جمله شبکه‌های تخاصمی مولد^۶، خود رمز گذارهای متغیر^۷ و مدل‌های انتشار^۸ تقسیم می‌شوند که هر یک کاربردهای خاص خود را در حوزه‌های مختلف از جمله تولید تصویر، متن و صوت دارند (Ghimire et al., 2026). پژوهش‌های متعددی، پتانسیل بالای این مدل‌ها را در مراحل مختلف چرخه حیات پروژه‌های ساخت، از جمله ارزیابی امکان‌سنجی اولیه، طراحی مفهومی، تدارکات و مناقصه، مرحله اجرا و نظارت و همچنین بهره‌برداری و نگهداری، به اثبات رسانده‌اند (Ghimire et al., 2026). به عنوان مثال، مدل ChatGPT برای برنامه‌ریزی و

1 ANP

2 AHP

3 NLP

4 Generative AI

5 LLMs

6 GANs

7 VAEs

8 Diffusion Models

زمان‌بندی پروژه‌های ساختمانی و مدل نمایش‌های رمزگذار دو جهته از ترانسفورماتورها^۱ برای شناسایی خودکار ریسک‌های ایمنی در محیط‌های کارگاهی مورد استفاده قرار گرفته است (Ghimire et al., 2026).

رویکرد مهندسی پرامپت^۲ به‌عنوان یک روش کلیدی و حیاتی برای هدایت مدل‌های بزرگ زبانی به سمت تولید خروجی‌های قابل اعتماد و دقیق در حوزه برآورد هزینه شناخته می‌شود (Ghimire et al., 2026). استفاده از روش زنجیره تفکر^۳ که مسائل پیچیده را به گام‌های میانی و صریح تجزیه می‌کند، دقت مدل‌ها را در وظایف استدلال چندمرحله‌ای به‌طور قابل توجهی افزایش می‌دهد (Ghimire et al., 2026). در یک مطالعه تجربی، روش زنجیره تفکر ماژولار توانست نمره اطمینان انسانی را از ۶۴ درصد به ۸۴ درصد افزایش دهد که نشان‌دهنده تأثیر شگرف این روش در بهبود عملکرد مدل‌ها است (Ghimire et al., 2026).

علاوه بر این، مدل‌های بزرگ زبانی از طریق پروتکل استاندارد MCP^۴ با نرم‌افزارهای مدل‌سازی اطلاعات ساختمان ادغام شده و فرآیند برآورد هزینه را به‌طور کامل خودکار می‌کنند (Ahn et al., 2024). یک مطالعه موردی بر روی سیستم الکتریکی یک ساختمان تجاری نشان داد که این رویکرد یکپارچه می‌تواند زمان برآورد را از ۳ ساعت (در حالت سنتی) به کمتر از ۴۵ ثانیه کاهش دهد که معادل ۹۸.۶ درصد صرفه‌جویی در زمان است (Ahn et al., 2024). با وجود این موفقیت‌های چشم‌گیر، مدل‌های بزرگ زبانی با چالش‌های اساسی پدیده توهم^۵ مواجه هستند که در آن مدل، پاسخ‌هایی را تولید می‌کند که به ظاهر منطقی و ساختاریافته هستند اما از نظر محتوایی کاملاً نادرست و بی‌اساس می‌باشند (Abdelsalam et al., 2026). همچنین، مطالعات نشان داده‌اند که مدل‌های کد باز مانند LLaMA و Gemma عملکرد پایین‌تری نسبت به مدل‌های اختصاصی و تجاری مانند GPT-4 در وظایف تخصصی و پیچیده برآورد هزینه دارند که می‌تواند ناشی از حجم داده‌های آموزش و بهینه‌سازی کمتر آن‌ها باشد (Wu et al., 2025).

درنهایت مرور ادبیات نشان می‌دهد که روش‌های سنتی برآورد هزینه (قضاوت خبرگی، تخمین پارامتریک و قیاسی) به دلیل ذهنی بودن، نبود شفافیت و عدم توانایی در مدل‌سازی روابط غیرخطی، با چالش‌های جدی مواجه هستند (Rush & Roy, 2001) (Lyu et al., 2025). در مقابل، سامانه‌های آنلاین اگرچه دسترسی سریعی فراهم می‌کنند، اما به دلیل عدم به‌روزرسانی و نادیده گرفتن شرایط محلی، از قابلیت اعتماد پایینی برخوردارند (He et al., 2024). در سال‌های اخیر، مدل‌های هوش مصنوعی مولد با قابلیت‌های پردازش زبان طبیعی و استدلال چندمرحله‌ای، پتانسیل بالایی در خودکارسازی برآورد هزینه نشان داده‌اند (Ghimire et al., 2026). با این حال، پژوهش‌های موجود عمدتاً بر روی یکی از این رویکردها متمرکز بوده و مطالعه‌ای که به صورت هم‌زمان به مقایسه سیستماتیک هر سه رویکرد در یک پروژه واقعی پرداخته باشد، انجام نشده است (Jafary et al., 2026). همچنین، موضوع تکرارپذیری مدل‌های هوش مصنوعی و نگرش مهندسان نسبت به این ابزارها کمتر مورد توجه قرار گرفته است (Ghimire et al., 2026) (Saeidlou & Ghadiminia, 2024). پژوهش حاضر با هدف پر کردن این شکاف‌ها، به مقایسه تطبیقی این سه رویکرد در پروژه آتشگاه می‌پردازد.

۳- روش‌شناسی

۳-۱- نوع تحقیق

پژوهش حاضر از نظر هدف، کاربردی است، زیرا نتایج آن می‌تواند به طور مستقیم در بهبود فرآیند برآورد هزینه پروژه‌های ساختمانی مورد استفاده قرار گیرد. از نظر روش، توصیفی-تحلیلی و پیمایشی-مقایسه‌ای است، چراکه به توصیف وضعیت موجود برآورد هزینه

1 BERT

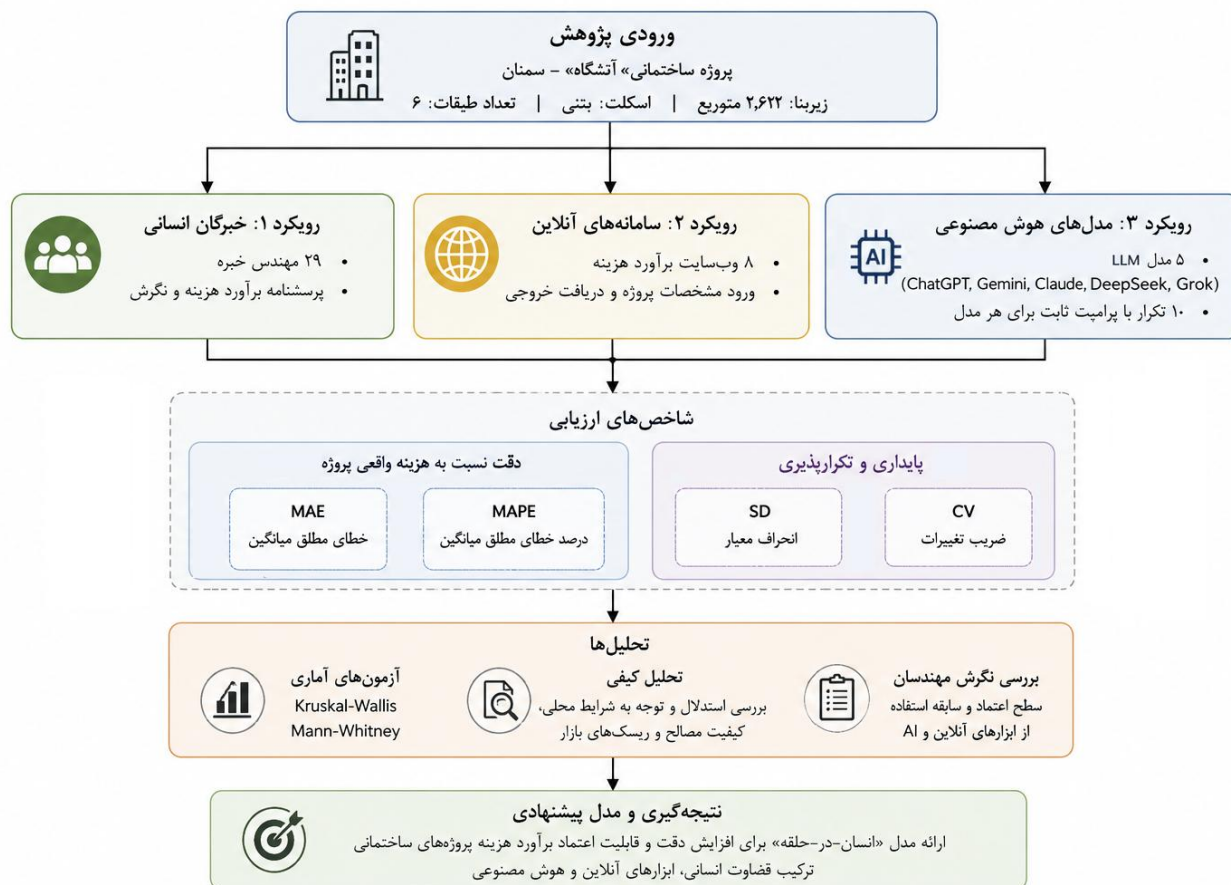
2 Prompt Engineering

3 Chain-of-Thought - CoT

4 Model Context Protocol

5 Hallucination

در سه رویکرد مختلف پرداخته و سپس به مقایسه سیستماتیک آن‌ها می‌پردازد. این پژوهش با رویکرد مطالعه موردی^۱ بر روی پروژه ساختمانی «آتشگاه» در شهر سمنان انجام شده است. انتخاب روش مطالعه موردی به دلیل قابلیت بررسی عمیق و دقیق پدیده‌های پیچیده در بستر واقعی پروژه‌های ساختمانی و همچنین امکان کنترل متغیرهای مداخله‌گر است. به منظور ارائه تصویری جامع از مراحل اجرای پژوهش، چارچوب مفهومی و فرآیند تحقیق در شکل (۱) ارائه شده است. همان‌گونه که مشاهده می‌شود، پژوهش از مشخصات پروژه واقعی آغاز شده و با استفاده از سه رویکرد شامل خبرگان انسانی، سامانه‌های آنلاین و مدل‌های هوش مصنوعی، برآورد هزینه انجام شده است. سپس خروجی‌های حاصل با استفاده از شاخص‌های دقت و پایداری، آزمون‌های آماری و تحلیل کیفی مورد ارزیابی قرار گرفته و در نهایت، بر اساس یافته‌ها، رویکرد انسان در حلقه به‌عنوان چارچوب پیشنهادی پژوهش ارائه شده است.



شکل ۱: چارچوب مفهومی و فرآیند اجرای پژوهش

۳-۲- معرفی پروژه مورد مطالعه

پروژه «آتشگاه» یک ساختمان ۶ طبقه با کاربری مسکونی-تجاری است که در بلوار آیت‌الله عالمی، واقع در جنوب شهر سمنان، قرار دارد. این پروژه بر اساس قراردادی دوعاملی به یک شرکت پیمانکاری ساختمانی واگذار شده است تا طراحی و ساخت

آن انجام شود. مبلغ کل قرارداد این پروژه ۸۱.۵۰۰.۰۰۰.۰۰۰ تومان (معادل هشتاد و یک میلیارد و پانصد میلیون تومان) با زیربنای ۲,۶۲۲ متر مربع است. مدت زمان قرارداد دو سال بوده و تا تاریخ ۳۰/۱۰/۱۴۰۶ پیش‌بینی شده است.

ساختمان دارای ۶ طبقه است که شامل یک طبقه زیرزمین به‌عنوان پارکینگ، طبقه همکف با کاربری تجاری و طبقات فوقانی با کاربری مسکونی است. در طبقات مسکونی، عقب‌نشینی پلکانی در هر طبقه اجرا شده است که بخشی از بام طبقه پایین به‌عنوان تراس برای طبقه بالا در نظر گرفته شده است. اسکلت ساختمان از نوع بتنی با سقف تیرچه‌بلوک بوده و از مصالح درجه یک ایرانی در ساخت آن استفاده شده است. در زمان انجام پژوهش (اردیبهشت ۱۴۰۵)، عملیات بتن‌ریزی اسکلت به پایان رسیده و تیغه‌چینی طبقه چهارم در حال انجام بود. مشخصات کلی پروژه، نمای رندر شده و وضعیت فعلی پروژه در جدول (۱) خلاصه شده است.

جدول ۱: مشخصات کلی پروژه مورد مطالعه



رندر پروژه



وضعیت فعلی پروژه

مشخصه	توضیحات
مبلغ قرارداد	۸۱/۵ میلیارد تومان
زیربنای کل	۲,۶۲۲ متر مربع
هزینه هر مترمربع	۳۱/۰۸ میلیون تومان
تعداد طبقات	۶ طبقه
کاربری	ترکیبی (۴ واحد تجاری در همکف + واحدهای مسکونی)
نوع اسکلت	بتنی با سقف تیرچه‌بلوک
کیفیت مصالح	درجه یک ایرانی
نما	آجر و سنگ
تأسیسات	کامل (پکیج، اسپلیت، کولر آبی)
اسانسور	۶ نفره
پنجره‌ها	دو جداره
شهر	سمنان
نوع قرارداد	دو عاملی (کارفرما - پیمانکار)
زمان برآورد	اردیبهشت ۱۴۰۵

۳-۳- جامعه و نمونه آماری

جامعه آماری این پژوهش شامل کلیه مهندسان و خبرگان فعال در حوزه ساخت‌وساز در شهر سمنان است. نمونه آماری با استفاده از روش نمونه‌گیری هدفمند^۱ انتخاب شد. در این روش که در پژوهش‌های کیفی و آمیخته رایج است، نمونه‌ها بر اساس معیارهای از پیش تعیین شده که با هدف پژوهش هماهنگ هستند، انتخاب می‌شوند. برخلاف نمونه‌گیری تصادفی که هدف آن تعمیم‌پذیری آماری است، در نمونه‌گیری هدفمند، محقق به‌صورت آگاهانه افرادی را انتخاب می‌کند که دسترسی به اطلاعات غنی و مرتبط با موضوع پژوهش را فراهم می‌کنند. این رویکرد برای مطالعه موردی عمیق و مقایسه‌ای که نیازمند قضاوت تخصصی درباره هزینه پروژه است، بسیار مناسب است.

معیارهای انتخاب خبرگان به‌گونه‌ای تعیین شد که افرادی با آگاهی و تجربه مستقیم از شرایط ساخت‌وساز در شهر سمنان وارد نمونه شوند. بر این اساس، حداقل یکی از معیارهای زیر برای ورود به نمونه کافی در نظر گرفته شد: (۱) دارا بودن مدرک کارشناسی یا

^۱ Purposeful Sampling

بالتر در رشته‌های مهندسی عمران، معماری یا تأسیسات؛ (۲) دارا بودن پروانه اشتغال نظام‌مهندسی؛ و (۳) آشنایی با شرایط بومی، قیمت‌های مصالح و شیوه‌های اجرایی رایج در استان سمنان (که از طریق تجربه کاری در این استان حاصل می‌شود). این معیارها با هدف دستیابی به خبرگانی طراحی شده‌اند که شناخت کافی از بازار محلی و پروژه‌های مشابه در سمنان داشته باشند تا برآوردهای آن‌ها مبتنی بر واقعیت‌های منطقه باشد. ضمناً همه مشارکت‌کنندگان به‌صورت داوطلبانه و با اعلام تمایل قبلی در پژوهش شرکت کردند. بر اساس این معیارها، ۲۹ مهندس با توزیع مشخص در زمینه‌های تحصیلات، سابقه کاری و نقش شغلی به‌عنوان نمونه آماری انتخاب شدند. برای تأیید کفایت حجم نمونه، از روش اشباع نظری^۱ استفاده شد. به این صورت که پس از جمع‌آوری هر پرسشنامه، پاسخ‌ها بررسی می‌شدند و پس از ۲۵ پرسشنامه، الگوی پاسخ‌ها به نقطه‌ای از تکرار رسید که داده‌های جدید اطلاعات تازه‌ای به پژوهش اضافه نمی‌کردند. برای اطمینان بیشتر، نمونه تا ۲۹ نفر افزایش یافت. به‌عنوان یک مطالعه موردی با رویکرد مقایسه‌ای، این حجم نمونه قابل قبول است، چراکه در پژوهش‌های مشابه در حوزه مدیریت ساخت، نمونه‌هایی با حجم ۲۰ تا ۵۰ نفر برای تحلیل‌های آماری توصیفی و استنباطی پایه مورد استفاده قرار گرفته‌اند. جدول (۲) توزیع جمعیت‌شناختی نمونه را نشان می‌دهد.

جدول ۲: توزیع جمعیت‌شناختی نمونه آماری

تحصیلات	درصد	تعداد	نقش	درصد	تعداد	سابقه کاری	درصد	تعداد
دیپلم	۶۲٪/۲۱	۸	پیمانکار	۲۴٪/۱۷	۵	کمتر از ۵ سال	۹٪/۶	۲
کارشناسی	۲۲٪/۱۶	۶	ناظر	۰۳٪/۳۱	۹	۵-۱۰ سال	۵۱٪/۶۵	۱۹
کارشناسی ارشد	۶۲٪/۲۱	۸	مجری	۵۹٪/۲۷	۸	۱۱-۲۰ سال	۱۴٪/۲۴	۷
دکتری	۱۰٪/۸	۳	مدیر پروژه	۱۴٪/۲۴	۷	بیشتر از ۲۰ سال	۴۵٪/۳	۱
	۲۲٪/۱۶	۶	سرپرست کارگاه					
	۲۲٪/۱۶	۶	سایر					

۳-۴- ابزار جمع‌آوری داده

ابزارهای جمع‌آوری داده در این پژوهش شامل سه بخش اصلی است که هر یک به یکی از رویکردهای سه‌گانه برآورد هزینه (خبرگان انسانی، سامانه‌های آنلاین و مدل‌های هوش مصنوعی) اختصاص دارد.

بخش اول: پرسشنامه مهندسان (رویکرد خبرگان انسانی)

پرسشنامه‌ای با هدف بررسی سطح اطمینان و نگرش مهندسان نسبت به سه رویکرد برآورد هزینه طراحی و توزیع شد. این پرسشنامه شامل سؤالات زیر بود:

- ✓ برآورد هزینه هر متر مربع توسط مهندس
- ✓ سطح اطمینان به برآورد خود (از ۱۰)
- ✓ میزان اعتماد به سامانه‌های آنلاین (از ۱۰)
- ✓ میزان اعتماد به هوش مصنوعی (از ۱۰)
- ✓ میزان اعتماد به تعرفه نظام مهندسی (از ۱۰)
- ✓ سابقه استفاده از سامانه‌های آنلاین (بله/خیر)
- ✓ سابقه استفاده از هوش مصنوعی (بله/خیر)

✓ دلایل بی‌اعتمادی (سؤال باز)

برای بررسی روایی محتوایی^۱، پرسشنامه در اختیار ۵ نفر از اساتید و خبرگان حوزه مدیریت ساخت قرار گرفت و اصلاحات لازم اعمال شد. نسبت روایی محتوا (CVR) و شاخص روایی محتوا (CVI) محاسبه گردید که به ترتیب برابر با ۰/۸۲ و ۰/۸۶ به دست آمد و نشان‌دهنده روایی مناسب پرسشنامه است.

برای بررسی پایایی^۲، نسخه آزمایشی پرسشنامه به صورت آزمایشی^۳ در اختیار ۱۰ نفر از مهندسان قرار گرفت و هماهنگی درونی گویه‌های اصلی پرسشنامه با استفاده از ضریب آلفای کرونباخ ارزیابی شد. مقدار آلفای کرونباخ برابر با ۰/۸۴ به دست آمد که نشان‌دهنده پایایی مطلوب پرسشنامه است.

بخش دوم: سامانه‌های آنلاین برآورد هزینه

هشت وب‌سایت فعال در حوزه برآورد هزینه ساخت مورد بررسی قرار گرفتند. معیار انتخاب این وب‌سایت‌ها، قابلیت ارائه برآورد آنلاین و در دسترس بودن برای کاربران عمومی بود. برای هر سایت، مشخصات پروژه «آتشگاه» (جدول ۱) بر اساس فیلدهای ورودی آن سایت، وارد و خروجی (هزینه هر متر مربع و هزینه کل) ثبت شد. جدول (۳) نتایج این بررسی را نشان می‌دهد.

جدول ۳: نتایج برآورد وب‌سایت‌های آنلاین

ردیف	نام سایت	فیلدهای ورودی سایت
۱	www.namaplan.com	نوع ساختمان، نوع سازه، امکانات اضافی، محل پروژه، شیب زمین
۲	www.dezharco.com	متراژ هر طبقه، تعداد طبقات، نوع اسکلت، کیفیت مصالح، مشمول بودن یا نبودن تأسیسات مکانیکی و برقی
۳	www.bkesh.ir	نوع پروژه، نوع سازه، سطح کیفیت اجرا، زیربنای کل، تعداد طبقات، نوع شهر
۴	www.ahan100.com	منطقه، نوع کاربری، متراژ هر طبقه، تعداد طبقات، پیلوت پارکینگ بلی یا خیر، تعداد طبقات زیرزمین، نوع سازه، کیفیت ساخت، ارتفاع هر طبقه، آسانسور، نوع نما، هوشمند سازی، امکانات اضافی
۵	www.karkoshte.com	تعداد طبقات، نوع سازه، نیاز به سازه نگهدارنده، کاربری، کیفیت ساخت
۶	www.pardissazeh.com	نوع ساختمان، کیفیت ساخت، استان، زیربنا، تعداد طبقات پارکینگ، دهانه زمین، تعداد نما، نوع سقف، مترال نما، تعداد آسانسور، کیفیت تأسیسات، شیب زمین، متراژ محوطه‌سازی، امکانات اضافی
۷	www.peymanu.com	متراژ، نوع سازه، سطح کیفیت، موقعیت جغرافیایی، مرحله ساخت
۸	www.khanehtarh.ir	نوع ساختمان، کیفیت ساخت، استان، زیربنا، تعداد طبقات مسکونی، تعداد طبقات پارکینگ، دهانه زمین، تعداد نما، مترال نما، نوع سقف، تعداد آسانسور، کیفیت تأسیسات، نوع سازه، شیب زمین، امکانات اضافی

بررسی فیلدهای ورودی هشت وب‌سایت نشان می‌دهد که این سامانه‌ها از نظر تعداد، نوع و جامعیت متغیرها تفاوت‌های چشم‌گیری دارند. این تفاوت‌ها را می‌توان در قالب یک طیف از رویکرد ساده (اکتفا به متغیرهای پایه مانند زیربنا و تعداد طبقات) تا رویکرد جامع (درخواست متغیرهای کیفی، محلی و پیشرفته) دسته‌بندی کرد. سامانه‌هایی که تنها به متغیرهای کلی اکتفا کرده‌اند، هزینه را به صورت خطی و مبتنی بر روابط سطح بالا تخمین می‌زنند و به طور طبیعی از توجه به عواملی مانند کیفیت مصالح، نوع نما، شرایط زمین، تأسیسات و امکانات اضافی غافل می‌شوند. در مقابل، سامانه‌های جامع‌تر با درخواست جزئیات اجرایی و محلی، پتانسیل بیشتری برای ارائه برآوردهای واقع‌بینانه و متناسب با پروژه‌های خاص دارند. این تنوع در رویکردها نشان می‌دهد که وب‌سایت‌های برآورد هزینه، از یک ابزار ساده و عمومی تا یک سامانه تخصصی و چندبعدی متغیر هستند و انتخاب هر یک توسط کاربر، باید بر

1 Content Validity

2 Reliability

3 Pilot Study

اساس سطح اطلاعات موجود و دقت مورد نیاز انجام شود. تأثیر این تفاوت‌ها بر دقت نهایی برآوردها، در بخش یافته‌ها به صورت کمی مورد تحلیل قرار خواهد گرفت.

بخش سوم: مدل‌های هوش مصنوعی مولد

پنج مدل هوش مصنوعی مولد شامل ChatGPT (نسخه 5.5)، Gemini (نسخه 3.5 Flash)، Claude (نسخه 4 Sonnet)، DeepSeek (نسخه V4) و Grok (نسخه 4) مورد آزمایش قرار گرفتند. تمامی آزمایش‌ها در اردیبهشت ۱۴۰۵، با حساب‌های رایگان و در یک گفت‌وگوی جدید (بدون سابقه) انجام شد و حالت جستجوی اینترنتی غیرفعال بود. معیارهای انتخاب این پنج مدل به شرح زیر است:

- ✓ تنوع در معماری و ساختار: مدل‌های انتخاب‌شده شامل طیف گسترده‌ای از معماری‌ها هستند، از مدل‌های اختصاصی تا مدل‌های کد باز بررسی گردید. این تنوع امکان مقایسه عملکرد مدل‌ها با ساختارهای مختلف را فراهم می‌کند.
- ✓ دسترسی و کاربرد گسترده: تمامی مدل‌های انتخاب‌شده در دسترس عموم هستند و مهندسان به راحتی می‌توانند از آن‌ها استفاده کنند. این ویژگی، نتایج پژوهش را برای جامعه مهندسی کاربردی‌تر می‌سازد.
- ✓ پوشش جغرافیایی و زبانی: مدل‌های انتخاب‌شده شامل مدل‌های آمریکایی و یک مدل چینی هستند. این تنوع جغرافیایی، امکان بررسی تأثیر داده‌های آموزشی و سوگیری‌های منطقه‌ای را بر روی خروجی‌ها فراهم می‌کند.
- ✓ پذیرش اجتماعی و فراوانی استفاده: بر اساس نتایج پرسشنامه این پژوهش، مدل‌های ChatGPT و Gemini بیشترین فراوانی استفاده را در بین مهندسان داشته‌اند؛ بنابراین، بررسی دقیق‌تر این مدل‌ها می‌تواند به درک بهتر نیازهای کاربران واقعی کمک کند.

فرآیند آزمایش با این مدل‌ها به شرح زیر بود: مشخصات پروژه آتشگاه (جدول ۱) با یک پرامپت ثابت به هر مدل ارائه شد. این پرامپت از مدل می‌خواست که با فرض نقش مشاور برآورد هزینه، صرفاً هزینه ساخت (بدون زمین و سود فروش) را به قیمت روز اردیبهشت ۱۴۰۵ برآورد کرده و خروجی را در قالب هزینه هر متر مربع، هزینه کل و بازه اطمینان ارائه دهد. هر مدل ۱۰ بار به صورت مجزا (با شروع یک گفت‌وگوی جدید در هر تکرار) اجرا گردید که در مجموع ۵۰ پاسخ هوش مصنوعی (۵ مدل × ۱۰ تکرار) جمع‌آوری شد.

۳-۵- فرآیند آزمایش

فرآیند آزمایش در سه مرحله انجام شد:

مرحله اول: برآورد خبرگان انسانی

۲۹ مهندس با توزیع تحصیلی و سابقه کاری مشخص (بخش ۳-۳)، برآورد خود را از هزینه هر متر مربع و هزینه کل پروژه ارائه دادند. همچنین سطح اطمینان خود را از برآوردشان اعلام کردند.

مرحله دوم: برآورد سامانه‌های آنلاین

هشت وب‌سایت برآورد هزینه با وارد کردن مشخصات پروژه آتشگاه (با توجه به فیلدهای ورودی هر سایت)، هزینه هر متر مربع و هزینه کل را محاسبه کردند.

مرحله سوم: برآورد مدل‌های هوش مصنوعی

پنج مدل هوش مصنوعی با پرامپت ثابت شامل مشخصات کامل پروژه، در ۱۰ تکرار مجزا، هزینه هر متر مربع، هزینه کل، بازه حداقل و حداکثر و توضیحات تکمیلی را تولید کردند.

۴- یافته‌ها

۴-۱- نتایج خبرگان

برآوردهای ۲۹ مهندس خبره نشان‌دهنده پراکندگی قابل‌توجهی در دیدگاه‌ها نسبت به هزینه پروژه آتشگاه است. میانگین برآورد خبرگان (۳۰/۸۳ میلیون تومان) به هزینه واقعی پروژه (۳۱/۰۸ میلیون تومان) بسیار نزدیک است و اختلاف فقط ۲۵۰ میلیون تومان (معادل ۰/۸۰ درصد) را نشان می‌دهد. با این حال، دامنه گسترده برآوردها از ۱۰ تا ۶۵ میلیون تومان و میانه ۳۰ میلیون تومان (نزدیک به میانگین) حاکی از تقارن نسبی داده‌ها، اما پراکندگی بالای نظرات خبرگان است. ضریب تغییرات محاسبه‌شده (۳۴/۶۴٪) بیانگر تنوع بالای قضاوت‌های کارشناسی است که می‌تواند ناشی از تفاوت در سطح تجربه، تخصص و نیز درک متفاوت از نوسانات بازار و شرایط اجرایی پروژه باشد. انحراف معیار ۱۰/۶۸ میلیون تومانی نیز نشان می‌دهد که حدود دو سوم برآوردها در بازه ۲۰ تا ۴۱ میلیون تومان قرار دارند که خود بیانگر عدم اجماع قابل‌توجه در بین خبرگان است. سطح اطمینان مهندسان به برآورد خود نسبتاً بالاست (۷/۰۴ از ۱۰) که نشان از اعتماد آن‌ها به قضاوت حرفه‌ای‌شان دارد، با وجود پراکندگی قابل‌توجه در خود برآوردها.

جدول ۴: آمار توصیفی برآورد خبرگان انسانی

شاخص	مقدار
میانگین هزینه هر متر مربع	تومان ۳۰۸۳۰۰۰۰
انحراف معیار	تومان ۱۰۶۸۰۰۰۰
(CV) ضریب تغییرات	۳۴/۶۴٪
میانه	تومان ۳۰۰۰۰۰۰۰
حداقل برآورد	تومان ۱۰۰۰۰۰۰۰
حداکثر برآورد	تومان ۶۵۰۰۰۰۰۰
میانگین سطح اطمینان (از ۱۰)	۷/۰۴

۴-۲- نگرش و اعتماد مهندسان به ابزارهای برآورد

مهندسان اعتماد پائینی به هر سه رویکرد غیرانسانی دارند، میانگین اعتماد در هیچ رویکردی از ۵ از ۱۰ فراتر نمی‌رود. بیشترین میزان اعتماد مربوط به تعرفه نظام مهندسی (۴/۸۹ از ۱۰) و کمترین آن مربوط به هوش مصنوعی (۴/۶۷ از ۱۰) است. انحراف معیار نسبتاً بالای هوش مصنوعی نشان‌دهنده تنوع قابل‌توجه در دیدگاه مهندسان نسبت به این فناوری است، در حالی که انحراف معیار پایین‌تر سایت‌های آنلاین بیانگر اجماع نسبی در بی‌اعتمادی به این ابزارهاست. این بی‌اعتمادی با نرخ پایین استفاده عملی از این ابزارها همخوانی دارد: تنها ۲۰/۶۹ درصد از مهندسان سابقه استفاده از سایت‌های آنلاین و ۳۵/۷۱ درصد سابقه استفاده از هوش مصنوعی را داشته‌اند. جالب توجه است که با وجود عملکرد قابل‌قبول مدل‌هایی مانند Grok در برآورد هزینه، اعتماد مهندسان به هوش مصنوعی همچنان در پایین‌ترین سطح قرار دارد که نشان‌دهنده شکاف عمیق بین عملکرد فنی و پذیرش عملی است. تحلیل پاسخ‌های باز نشان می‌دهد که مهم‌ترین دلایل این بی‌اعتمادی، به‌روز نبودن داده‌ها، نادیده گرفتن نوسانات بازار و عدم توجه به شرایط محلی است.

جدول ۵: میزان اعتماد مهندسان به رویکردهای برآورد هزینه

رویکرد	میانگین اعتماد (از ۱۰)	انحراف معیار
تعرفه نظام مهندسی	۴/۸۹	۲/۵۵
سایت‌های آنلاین	۴/۶۸	۲/۱۳
هوش مصنوعی	۴/۶۷	۲/۷۳

تحلیل پاسخ‌های باز پرسشنامه، دلایل اصلی این بی‌اعتمادی را آشکار می‌سازد. مهم‌ترین دلایل بی‌اعتمادی به سایت‌های آنلاین عبارت است از: ثابت نبودن قیمت‌ها و عقب ماندن سامانه‌ها از نوسانات بازار، وجود اشتباهات در محاسبات و چند نرخ بودن ارقام در بازار، نرخ تورم بالا و غیرقابل پیش‌بینی ایران، جامع نبودن و عدم دقت در لحاظ کردن جزئیات اجرایی پروژه و مشکلات دسترسی به اینترنت. در مورد هوش مصنوعی نیز مهندسان عواملی چون پشتیبان‌گیری از داده‌ها و منابع غیردقیق، عدم اعتماد عمومی به تکنولوژی‌های جدید در صنعت سنتی ساخت، عدم درک کافی از شرایط خاص شهر سمنان و تفاوت‌های منطقه‌ای و ناتوانی در لحاظ کردن نوسانات لحظه‌ای بازار را به‌عنوان دلایل اصلی بی‌اعتمادی خود مطرح کرده‌اند. این یافته‌ها نشان می‌دهد که شکاف اعتماد، ریشه در مشکلات ساختاری (مانند نبود داده‌های به‌روز و محلی) و مشکلات ادراکی (مانند بی‌اعتمادی به فناوری جدید) دارد و صرفاً بهبود فنی ابزارها بدون توجه به این عوامل، نمی‌تواند پذیرش آن‌ها را افزایش دهد.

۳-۴- نتایج وب‌سایت‌ها

تمامی برآوردهای وب‌سایت‌ها در یک بازه زمانی مشخص (اردیبهشت ۱۴۰۵) و به‌صورت هم‌زمان انجام شد تا از تأثیر نوسانات قیمت‌های روز بر روی خروجی‌ها جلوگیری شود. برای هر سایت، بر اساس فیلدهای ورودی آن، مشخصات پروژه وارد و خروجی ثبت گردید. قیمت‌های اعلام‌شده در این سامانه‌ها عموماً به قیمت روز اشاره دارند. در برآوردهای این سامانه‌ها، هزینه‌های جانبی مانند عوارض شهرداری، طراحی، نظارت و سود پیمانکار به‌طور یکسان لحاظ نشده است که می‌تواند یکی از دلایل پراکندگی خروجی‌ها باشد. جدول (۶) نتایج کامل این برآوردها را به همراه آمار توصیفی آن‌ها ارائه می‌دهد.

جدول ۶: نتایج برآورد وب‌سایت‌های آنلاین

ردیف	نام سایت	هزینه هر متر مربع	هزینه کل پروژه
۱	www.namaplan.com	۵۳۴۸۶۹۹۰	۱۴۰۲۴۲۸۸۹۰۸۰
۲	www.dezharco.com	۴۲۰۰۰۰۰	۱۰۹۰۱۲۰۰۰۰۰۰
۳	www.bkesh.ir	۴۶۲۰۰۰۰	۱۲۱۱۳۶۴۰۰۰۰۰
۴	www.ahan100.com	۳۰۵۵۷۲۰۸	۸۰۱۲۱۰۰۰۰۰۰
۵	www.karkoshte.com	۳۱۰۵۰۰۰	۸۱۴۱۳۱۰۰۰۰۰
۶	www.pardissazeh.com	۲۵۰۰۰۰۰	۵۴۵۴۴۰۰۰۰۰۰
۷	www.peymanu.com	معادل ^۱ ۲۹۴۰۰۰۰۰	۷۷۰۸۶۸۰۰۰۰۰
۸	www.khanehtarh.ir	معادل ۵۰۴۲۱۰۰۰	۱۳۲۲۰۳۸۶۲۰۰۰

بر اساس نتایج جدول (۶)، برآوردهای وب‌سایت‌ها از ۲۵ تا ۵۳/۵ میلیون تومان به ازای هر متر مربع متغیر است که اختلافی بیش از ۱۵۷ درصد را نشان می‌دهد. میانگین برآورد این رویکرد (۳۷/۴ میلیون تومان) حدود ۲۰/۳۴ درصد بالاتر از هزینه واقعی پروژه (۳۱/۰۸ میلیون تومان) و همچنین حدود ۲۱/۷ درصد بالاتر از میانگین برآورد خبرگان انسانی (۳۰/۸۳ میلیون تومان) قرار دارد. این فاصله قابل توجه نشان‌دهنده تمایل سامانه‌های آنلاین به برآوردهای بالاتر از واقعیت است. ضریب تغییرات ۲۴/۲۹ درصد (جدول ۷) حاکی از پراکندگی قابل توجه در بین برآوردهای سامانه‌های آنلاین است. این پراکندگی گسترده را می‌توان به عوامل متعددی نسبت داد: تفاوت در فیلدهای ورودی، الگوریتم‌های محاسباتی متفاوت و نرخ به‌روزرسانی قیمت‌ها. با توجه به پراکندگی بالا و عدم

۱ این دو سایت قیمت گذاری سال ۱۴۰۴ را داشتند با توجه به نرخ تورم سالانه بخش ساخت‌وساز در اردیبهشت ۱۴۰۵ معادل (۴۷٪) در ۱/۴۷ ضرب شد.

همگرایی بین برآوردها، می‌توان نتیجه گرفت که سامانه‌های آنلاین برآورد هزینه در وضعیت فعلی، قابلیت اعتماد کافی برای کاربرد عملی در پروژه‌های واقعی را ندارند و استفاده از آن‌ها باید با احتیاط فراوان و همراه با تفسیر نتایج توسط یک متخصص صورت گیرد.

جدول ۷: شاخص‌های آماری برآورد وب‌سایت‌های آنلاین

شاخص	مقدار
میانگین هزینه هر متر مربع	۳۷,۳۹۹,۵۲۴ تومان
انحراف معیار	۹,۰۸۶,۰۰۰ تومان
ضریب تغییرات (CV)	۲۴.۲۹٪
حداقل هزینه هر متر مربع	۲۵,۰۰۰,۰۰۰ تومان (pardissazeh)
حداکثر هزینه هر متر مربع	۵۳,۴۸۶,۹۹۰ تومان (namaplan)
میانگین هزینه کل	۳۶۰.۹۹۰.۰۰۰.۰۰۰ تومان
حداقل هزینه کل	۵۴,۵۴۴,۰۰۰,۰۰۰ تومان
حداکثر هزینه کل	۱۴۰,۲۴۲,۸۸۹,۰۸۰ تومان

۴-۴- نتایج مدل‌های هوش مصنوعی

پنج مدل هوش مصنوعی در ۱۰ تکرار با پرامپت ثابت آزمایش شدند. نتایج نشان می‌دهد که مدل‌ها از نظر دقت نسبت به هزینه واقعی و پایداری پاسخ‌ها تفاوت‌های چشم‌گیری دارند. Grok، با خطای ۲/۳۲ درصد نسبت به هزینه واقعی پروژه و بالاترین پایداری (۵.۳۲٪) بهترین عملکرد را در پروژه مورد مطالعه داشته است. Gemini با خطای ۱۲/۴۸ درصد و پایداری متوسط، در رتبه دوم قرار دارد. ChatGPT و DeepSeek با خطای حدود ۲۲ تا ۲۴ درصد، عملکرد متوسطی داشته‌اند، عملکرد متوسطی داشته‌اند که DeepSeek با ضریب تغییرات ۳۴/۸۸ درصد از پایداری کمتری برخوردار است. در مقابل، Claude با خطای ۶۳.۷۷٪ و پایداری (۷۱.۲۳٪) ضعیف‌ترین عملکرد را نشان داده است. این تفاوت‌ها می‌تواند ناشی از معماری متفاوت مدل‌ها، حجم و کیفیت داده‌های آموزش و همچنین رویکرد طراحی آن‌ها باشد.

جدول ۸: نتایج برآورد مدل‌های هوش مصنوعی

مدل	میانگین هزینه هر متر مربع (تومان)	میانگین هزینه کل (تومان)	انحراف معیار (SD)	ضریب تغییرات (CV%)	حداقل (میلیارد)	حداکثر (میلیارد)
Grok	۳۱۸۰۰۰۰۰	۸۳,۴۰۰,۰۰۰,۰۰۰	۱,۶۹۰,۰۰۰	۵/۳۵	۷۱	۹۹
ChatGPT	۲۴۳۵۰۰۰۰	۶۳,۸۰۰,۰۰۰,۰۰۰	۳,۴۳۰,۰۰۰	۱۴/۰۹	۴۸/۵	۹۵
Gemini	۲۷۲۰۰۰۰۰	۷۱,۳۰۰,۰۰۰,۰۰۰	۵,۱۸۰,۰۰۰	۱۹/۴	۳۹/۵	۹۴/۳
DeepSeek	۲۴۰۲۰۰۰۰	۶۳,۰۰۰,۰۰۰,۰۰۰	۸,۲۰۰,۰۰۰	۳۴/۱۴	۲۳/۴	۱۰۵
Claude	۱۱۲۶۰۰۰۰	۲۹,۵۰۰,۰۰۰,۰۰۰	۸,۰۲۰,۰۰۰	۷۱/۲۳	۹/۵	۷۵

علاوه بر تحلیل کمی، محتوای پاسخ‌های ارائه‌شده توسط هر مدل نیز مورد بررسی قرار گرفت. هدف این تحلیل، ارزیابی کیفیت استدلال هر مدل در زمینه‌هایی مانند توجه به شرایط محلی، کیفیت مصالح و ریسک‌های بازار بود. نتایج نشان داد که مدل‌های Gemini و Grok در پاسخ‌های خود به عواملی مانند «نوسانات قیمت مصالح در ایران» و «تفاوت هزینه‌ها در شهرستان‌ها» اشاره کرده‌اند که نشان‌دهنده درک نسبی از شرایط بومی است. در مقابل، مدل Claude در ۳ تکرار از ۱۰ تکرار از ارائه برآورد عددی خودداری کرده و پاسخ‌هایی با محتوای عمومی و غیرتخصصی ارائه داد. مدل‌های ChatGPT و DeepSeek عمدتاً بر اساس

پارامترهای ورودی (مانند زیربنا و تعداد طبقات) به محاسبه پرداخته‌اند، اما به عواملی مانند نوسانات ارز و تورم اشاره‌ای نداشتند. این یافته نشان می‌دهد که مدل‌های هوش مصنوعی در صورت ارائه پرامپت دقیق، توانایی نسبی در شناسایی عوامل مؤثر بر هزینه دارند، اما همچنان برای درک عمیق‌تر از شرایط محلی نیاز به بهبود دارند. معماری مدل‌ها و داده‌های آموزش‌دیده آن‌ها تأثیر مستقیمی بر کیفیت استدلال و نوع نگرش به عوامل مؤثر بر هزینه دارد.

۴-۵- مقایسه نهایی سه رویکرد

برای ارزیابی دقیق عملکرد هر یک از سه رویکرد برآورد هزینه، دو معیار اصلی مورد بررسی قرار گرفت: دقت و پایداری. مبنای سنجش دقت، هزینه واقعی پروژه است که بر اساس مبلغ نهایی قرارداد (۸۱,۵۰۰,۰۰۰,۰۰۰ تومان) و زیربنای کل (۲,۶۲۲ متر مربع) معادل ۳۱,۰۸۰,۰۰۰ تومان به ازای هر متر مربع محاسبه شده است. این مقدار به عنوان معیار مرجع برای تعیین خطای هر روش به کار رفته است. شاخص‌های مورد استفاده شامل خطای مطلق^۱ و درصد خطای مطلق میانگین^۲ برای سنجش دقت، و ضریب تغییرات برای سنجش پایداری هستند. جدول (۹) خلاصه مقایسه‌ای سه رویکرد را بر اساس این شاخص‌ها نمایش می‌دهد.

جدول ۹: مقایسه نهایی سه رویکرد برآورد هزینه بر اساس دقت و پایداری

گروه	میانگین هزینه هر متر مربع (تومان)	خطای مطلق (تومان $\frac{m^2}{m^2}$)	درصد خطای مطلق میانگین (%)	انحراف معیار	ضریب تغییرات
هزینه واقعی (مرجع)	۳۱۰۸۰۰۰۰	-	-	-	-
مهندسان (خبرگان)	۳۰۸۳۰۰۰۰	۳۶۰۰۰۰	۱.۱۶٪	۱۱,۱۶۰,۰۰۰	۳۶.۳۳٪
وب سایت‌ها	۳۷۴۰۰۰۰۰	۶۳۲۰۰۰۰	۲۰.۳۴٪	۹,۰۸۶,۰۰۰	۲۴.۲۹٪
Chat GPT	۲۴۳۵۰۰۰۰	۷۰۵۵۰۰۰	۲۲.۷۰٪	۳,۰۷۰,۰۰۰	۱۲.۸٪
Gemini	۲۷۲۰۰۰۰۰	۳۸۸۰۰۰۰	۱۲.۴۸٪	۵,۱۸۰,۰۰۰	۱۹.۴٪
Claude	۱۱۲۶۰۰۰۰	۱۹۸۲۰۰۰۰	۶۳.۷۷٪	۸,۰۲۰,۰۰۰	۷۱.۲۳٪
DeepSeek	۲۴۰۲۰۰۰۰	۷۵۶۰۰۰۰	۲۴.۳۳٪	۸,۲۰۰,۰۰۰	۳۴.۸۸٪
Grok	۳۱۸۰۰۰۰۰	۷۲۰۰۰۰	۲.۳۲٪	۱,۶۹۰,۰۰۰	۵.۳۲٪

برای تعیین این‌که تفاوت‌های مشاهده‌شده میان برآوردهای سه رویکرد از نظر آماری معنی‌دار است یا خیر، از آزمون ناپارامتریک کروسکال-والیس استفاده شد. از آنجا که داده‌ها در هر سه گروه از توزیع نرمال پیروی نمی‌کنند و حجم نمونه‌ها نابرابر است، این آزمون نسبت به آزمون پارامتریک ANOVA ارجحیت دارد. جدول (۱۰) نتایج این آزمون را نشان می‌دهد.

جدول ۱۰: نتایج آزمون کروسکال-والیس

مقدار	شاخص
۲۴/۶	آماره آزمون (H)
۲	درجه آزادی (df)
(p<۰/۰۰۱)	سطح معنی‌داری (p-value)
تفاوت معنی‌دار است	نتیجه

نتایج جدول (۱۰) نشان می‌دهد که تفاوت میان سه رویکرد از نظر آماری معنی‌دار است (p<۰/۰۰۱)، به عبارت دیگر، حداقل یکی از رویکردها با دیگران تفاوت دارد و این تفاوت تصادفی نیست.

^۱ Mean Absolute Error (MAE)

^۲ Mean Absolute Percentage Error (MAPE)

برای مشخص کردن این که کدام رویکردها با یکدیگر تفاوت دارند، آزمون تعقیبی من-ویتنی با اصلاح بونفرونی انجام شد. سطح معنی داری اصلاح شده برابر با ۰/۰۱۷ در نظر گرفته شد. جدول (۱۱) نتایج این آزمون را نشان می دهد.

جدول ۱۱: نتایج آزمون من-ویتنی (با اصلاح بونفرونی)

مقایسه	p-value (تصحیح شده)	نتیجه
خبرگان - وبسایتها	۰/۰۷۵	معنی دار نیست
خبرگان - هوش مصنوعی	۰/۱۰۵	معنی دار نیست
وبسایتها - هوش مصنوعی	($p < ۰/۰۰۱$)	معنی دار

نتایج جدول (۱۱) نشان می دهد که تنها تفاوت بین وبسایتها و مدل های هوش مصنوعی از نظر آماری معنی دار است. این یافته نشان می دهد که وبسایتها تمایل به برآوردهای بالاتر و مدل های هوش مصنوعی تمایل به برآوردهای پایین تر دارند و این تفاوت تصادفی نیست. در مقابل، اختلاف برآوردهای خبرگان با دو رویکرد دیگر از نظر آماری قابل تأیید نیست.

۵- بحث و نتیجه گیری

هدف این پژوهش، مقایسه سه رویکرد برآورد هزینه (خبرگان انسانی، سامانه های آنلاین و مدل های هوش مصنوعی مولد) در یک پروژه واقعی ساختمانی بود. یافته ها نشان داد که هر یک از این رویکردها نقاط قوت و ضعف مشخصی دارند. در مقایسه کمی، مدل های هوش مصنوعی مولد، به ویژه Grok، با خطای ۲/۳۲ درصد نسبت به هزینه واقعی پروژه و ضریب تغییرات ۵/۳۲ درصد، بهترین عملکرد را از نظر ترکیب دقت و پایداری در پروژه مورد مطالعه ارائه دادند. این یافته با پژوهش های پیشین که نشان می دهند مدل های بزرگ زبانی با رویکرد مهندسی پرامپت مناسب می توانند در وظایف تخصصی حوزه ساخت موفق عمل کنند، همسو است. در مقابل، مدل Claude با خطای ۶۳/۷۷ درصد و ضریب تغییرات ۷۱/۲۳ درصد، عملکردی نامناسب و ناپایدار نشان داد و در ۳ تکرار از ۱۰ تکرار از ارائه برآورد عددی خودداری کرد. این موضوع نشان می دهد که همه مدل های هوش مصنوعی برای کاربرد برآورد هزینه در پروژه های ساختمانی مناسب نیستند. سامانه های آنلاین نیز با خطای ۲۰/۳۴ درصد نسبت به هزینه واقعی و ضریب تغییرات ۲۴.۲۹ درصد در میانه این طیف قرار گرفتند، اما پراکندگی زیاد در بین خودشان (۱۵۷ درصد اختلاف بین کمترین و بیشترین برآورد) نشان دهنده عدم انسجام و قابلیت اعتماد پایین این رویکرد است. آزمون کروسکال-والیس نیز تفاوت معنی دار میان سه رویکرد را تأیید کرد ($p < ۰/۰۰۱$)، هر چند آزمون تعقیبی من-ویتنی نشان داد که تنها تفاوت بین وبسایتها و مدل های هوش مصنوعی از نظر آماری معنی دار است.

نکته قابل تأمل این است که خبرگان انسانی با وجود سطح اطمینان نسبتاً بالا (۷/۰۴ از ۱۰)، از پراکندگی قابل توجهی در برآوردها رنج می برند ($CV = ۳۴.۶۴\%$) که ریشه در تفاوت سطح تجربه، تخصص و درک متفاوت از نوسانات بازار دارد. با این حال، اجماع جمعی مهندسان (میانگین ۳۰/۸۳ میلیون تومان) به هزینه واقعی پروژه (۳۱/۰۸ میلیون تومان) بسیار نزدیک است و اختلاف فقط ۰/۸۰ درصد را نشان می دهد که بیانگر ارزش مرجعیت قضاوت جمعی خبرگان است. در مقابل، میانگین برآورد وبسایتها (۳۷/۴ میلیون تومان) حدود ۲۰/۳۴ درصد بالاتر از هزینه واقعی است که نشان دهنده تمایل این سامانه ها به برآوردهای بالاتر از واقعیت است.

بررسی تکرارپذیری مدل های هوش مصنوعی در تکرارهای متعدد با پرامپت ثابت نشان داد که این ویژگی در مدل های مختلف به طور قابل توجهی متفاوت است و ارتباط مستقیمی با معماری مدل و فرآیند آموزش آنها دارد. مدل Grok با بالاترین تکرارپذیری و مدل Claude با پایین ترین تکرارپذیری، دو سر طیف را نشان دادند. این یافته، شکاف پژوهشی اشاره شده در مقدمه را پر می کند و لزوم ارزیابی پایداری مدل ها پیش از استفاده عملیاتی در پروژه های واقعی را آشکار می سازد. تحلیل کیفی پاسخ ها نیز نشان داد که مدل های

Gemini و DeepSeek در توضیحات خود به عواملی مانند «نوسانات قیمت مصالح در ایران» و «تفاوت هزینه‌ها در شهرستان‌ها» اشاره کرده‌اند که نشان‌دهنده درک نسبی از شرایط بومی است، در حالی که مدل‌های ChatGPT و Grok توضیحات استاندارد و تکراری ارائه داده‌اند.

یکی دیگر از یافته‌های کلیدی، شکاف قابل توجه بین عملکرد واقعی مدل‌های هوش مصنوعی و اعتماد مهندسان به آن‌ها است. با وجود عملکرد نسبتاً مناسب مدل‌هایی مانند Gemini و Grok، میانگین اعتماد مهندسان به هوش مصنوعی تنها ۴/۶۷ از ۱۰ بود و تنها ۳۵/۷۱ درصد از آن‌ها سابقه استفاده از این ابزارها را داشتند. این یافته با پژوهش‌های حوزه پذیرش فناوری در صنعت ساخت همسو است که نشان می‌دهد عوامل انسانی و ادراکی، اغلب از موانع جدی‌تری نسبت به محدودیت‌های فنی برای پذیرش فناوری‌های جدید هستند. دلایل بی‌اعتمادی مهندسان (همانند «پشتیبان‌گیری از داده‌های غیردقیق»، «عدم درک شرایط محلی سمنان» و «ماهیت جعبه سیاه مدل‌ها») نشان می‌دهد که تنها بهبود عملکرد فنی کافی نیست، بلکه باید ابزارهایی توسعه یابند که خروجی مدل را به همراه توضیحات قابل فهم برای کاربر ارائه دهند تا شفافیت و قابلیت اعتماد درک‌شده افزایش یابد. در مورد سامانه‌های آنلاین نیز الگوی مشابهی مشاهده می‌شود: اعتماد پایین (۴/۶۸ از ۱۰) با عملکرد ضعیف (خطای ۲۰/۳۴ درصد و پراکندگی بالا) همخوانی دارد.

پژوهش حاضر با محدودیت‌هایی از جمله حجم نمونه خبرگان (۲۹ نفر) و تمرکز بر یک پروژه در یک شهر مواجه بود که تعمیم‌پذیری نتایج را محدود می‌کند. همچنین، مدل‌های هوش مصنوعی در بازه زمانی مشخصی (اردیبهشت ۱۴۰۵) آزمایش شده‌اند و ممکن است با به‌روزرسانی مدل‌ها، نتایج تغییر کند. با وجود این محدودیت‌ها، یافته‌های این پژوهش نشان می‌دهد که رویکرد ترکیبی «انسان-در-چهارچوب» می‌تواند به‌عنوان مؤثرترین راهکار برای برآورد هزینه در پروژه‌های ساختمانی مورد استفاده قرار گیرد. در این رویکرد، مدل‌های هوش مصنوعی پایدار (مانند Grok) برآورد اولیه را با سرعت بالا تولید می‌کنند و خبرگان انسانی با تکیه بر تجربه و دانش زمینه‌ای خود، آن را تأیید، اصلاح و تکمیل می‌نمایند.

بر اساس یافته‌ها، به مهندسان و مدیران پروژه توصیه می‌شود که پیش از استفاده از هر مدل هوش مصنوعی، پایداری آن را در تکرارهای متعدد با پرامپت‌های ثابت ارزیابی کرده و مدل‌هایی با ضریب تغییرات کمتر از ۱۵ درصد را برای کاربردهای عملی انتخاب نمایند. همچنین، سازمان نظام مهندسی ساختمان می‌تواند با تدوین دستورالعمل‌های مدون و برگزاری دوره‌های آموزشی تخصصی، به افزایش اعتماد و پذیرش این فناوری در جامعه مهندسی کمک کند. پژوهش‌های آینده می‌توانند با تکرار این روش در پروژه‌های متعدد در شهرهای مختلف و با حجم نمونه بزرگ‌تر، به تعمیم‌پذیری نتایج کمک کنند. بررسی تأثیر پرامپت‌های مختلف بر عملکرد مدل‌های هوش مصنوعی، توسعه ابزارهای توضیح‌پذیر^۱ و ادغام ویژگی‌های با ابعاد بالاتر مانند شاخص‌های کلان اقتصادی و نوسانات قیمت مصالح منطقه‌ای، از مسیرهای ارزشمند برای پژوهش‌های آتی است. در نهایت، با وجود پیشرفت‌های سریع هوش مصنوعی، قضاوت مهندسی همچنان نقش محوری و غیرقابل جایگزینی در فرآیند برآورد هزینه دارد و فناوری‌های نوین باید به‌عنوان ابزارهای مکمل در خدمت خبرگان انسانی قرار گیرند، نه جایگزین آن‌ها.

۶- تعارض منافع

نویسندگان اعلام می‌کنند که هیچ‌گونه تعارض منافی در ارتباط با این پژوهش وجود ندارد.

۷- مشارکت نویسندگان

شایان فلاحي: گردآوری داده‌ها، نگارش مقاله، بازنگری علمی

شایان حجت پناه: طراحی پژوهش، تحلیل داده‌ها، بازنگری علمی
 سحر علینژاد مجیدی: بازنگری علمی، تأیید نسخه نهایی

۸- بیانیه استفاده از هوش مصنوعی

در تهیه و نگارش این مقاله، از ابزارهای هوش مصنوعی مولد (شامل ChatGPT و DeepSeek) برای ویرایش نگارشی، بهبود ساختار متن، ترجمه و خلاصه‌سازی استفاده شده است. شایان ذکر است که این پژوهش خود به‌عنوان یک مطالعه تطبیقی، به‌طور نظام‌مند به بررسی و سنجش عملکرد مدل‌های هوش مصنوعی (شامل ChatGPT، Gemini، Claude، DeepSeek و Grok) در برآورد هزینه یک پروژه واقعی ساختمانی پرداخته است و داده‌های خام تولیدشده توسط این مدل‌ها به‌عنوان یکی از سه رویکرد اصلی پژوهش (در کنار خبرگان انسانی و سامانه‌های آنلاین) جمع‌آوری، تحلیل و در مقاله گزارش شده است؛ بنابراین، استفاده از هوش مصنوعی در این مقاله در دو سطح انجام شده است: (۱) به‌عنوان موضوع پژوهش در فرآیند جمع‌آوری داده‌ها و (۲) به‌عنوان ابزار نگارش در ویرایش و بهبود متن. کلیه تحلیل‌های علمی، تفسیر نتایج، نتیجه‌گیری‌ها و مسئولیت کامل محتوای مقاله بر عهده نویسندگان بوده و نسخه نهایی مقاله توسط آنان به‌طور کامل بازبینی و تأیید شده است.

۹- منابع

- 1- Rush, C., & Roy, R. (2001). Expert judgement in cost estimating: Modelling the reasoning process. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 215(9), 1289–1299. <https://doi.org/10.1177/1063293X0100900404>
- 2- Kim, G. H., An, S. H., & Kang, K. I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment*, 39(10), 1235–1242. <https://doi.org/10.1016/j.buildenv.2004.02.013>
- 3- Lyu, B., Yin, Q., Tommelein, I. D., Liu, H., Ranka, K., Yeluripati, K., & Shi, J. (2025). A granular framework for construction material price forecasting: Econometric and machine-learning approaches. *arXiv*. <https://doi.org/10.48550/arXiv.2512.09360>
- 4- He, C., Yu, B., Liu, M., Guo, L., Tian, L., & Huang, J. (2024). Utilizing large language models to illustrate constraints for construction planning. *Buildings*, 14(8), Article 2511. <https://doi.org/10.3390/buildings14082511>
- 5- Abdelsalam, M., Ashmawi, A., & Nguyen, P. H. D. (2026). AI-driven automation of construction cost estimation: Integrating BIM with large language models. *Buildings*, 16(3), Article 485. <https://doi.org/10.3390/buildings16030485>
- 6- Ahn, I.-S., Kim, J.-J., & Lee, J.-S. (2024). Development of a cost prediction model for design changes: Case of Korean apartment housing projects. *Sustainability*, 16(11), Article 4322. <https://doi.org/10.3390/su16114322>
- 7- Ghimire, P., Kim, K., Stentz, T., & Roy, T. (2026). Modular chain-of-thought (CoT) for LLM-based conceptual construction cost estimation. *Buildings*, 16(2), Article 396. <https://doi.org/10.3390/buildings16020396>
- 8- Tochaiwat, K., & Budda, A. (2025). Construction contract price prediction model for government buildings using a deep learning technique: A study from Thailand. *Buildings*, 16(3), Article 651. <https://doi.org/10.3390/buildings16030651>
- 9- Jafary, P., Shojaei, D., Rajabifard, A., & Ngo, T. (2026). Artificial intelligence (AI) and machine learning in building information modeling (BIM)-based construction cost estimation: A systematic review. *Geo-spatial Information Science*. Advance online publication. <https://doi.org/10.1080/10095020.2026.2651578>
- 10- Kim, K. J., Kim, D. S., Kim, K., Kim, E. W., & Shin, Y. J. (2024). Estimating hypothetical competitive bid price on bid-rigged projects based on the weights of major work groups. *KSCE Journal of Civil Engineering*. Advance online publication. <https://doi.org/10.1007/s12205-024-1891-9>
- 11- Sadikoglu, E., & Demirkesen, S. (2025). Review of machine learning and artificial intelligence use for cost estimation in construction projects. *Proceedings of the European Conference on Computing in Construction*. <https://doi.org/10.35490/EC3.2025.363>
- 12- Wu, Y., Wang, L., & Liu, R. (2025). CEQuest: Benchmarking large language models for construction estimation. *Proceedings of the ACM International Conference*. <https://doi.org/10.48550/arXiv.2508.16081>

- 13- Xiong, R., Wang, Y., Gunhan, S., Zhu, Y., & Berryman, C. (2025). Can AI master construction management (CM)? Benchmarking state-of-the-art large language models on CM certification exams. Proceedings of the ASCE Construction Research Congress. <https://doi.org/10.48550/arXiv.2504.08779>
- 14- Shi, T., & Shide, K. (2025). A comparative analysis of LSTM, GRU, and Transformer models for construction cost prediction with multidimensional feature integration. Journal of Asian Architecture and Building Engineering, 25(1), 634–649. <https://doi.org/10.1080/13467581.2025.2455034>
- 15- Chen, L., Lim, Z. M., Lim, W. H., Tiang, S. S., Sharma, A., Syamsunur, D., Alharbi, A. H., Eid, M. E., & El-Kenawy, E.-S. M. (2026). Search-guided regression ensembles for accurate, interpretable, and uncertainty-aware construction cost estimation. Scientific Reports, 16, Article 51706. <https://doi.org/10.1038/s41598-026-51706-5>
- 16- Liu, C. P., Sun, X. G., & Guan, J. H. (2026). DSCostPred: A double-stacking model for construction cost prediction. Scientific Reports, 16, Article 3365. <https://doi.org/10.1038/s41598-025-33305-y>
- 17- Jafary, P., Shojaei, D., Rajabifard, A., & Ngo, T. (2025). AI-augmented construction cost estimation: An ensemble natural language processing (NLP) model to align quantity take-offs with cost indexes. International Journal of Construction Management, 26(8), 1508–1526. <https://doi.org/10.1080/15623599.2025.2558070>
- 18- Saeidlou, S., & Ghadiminia, N. (2024). A construction cost estimation framework using DNN and validation unit. Building Research & Information, 52(1–2), 38–48. <https://doi.org/10.1080/09613218.2023.2196388>
- 19- Yu, W., Lai, C., & Lee, W. (2006). A WICE approach to real-time construction cost estimation. Automation in Construction, 15(1), 12–19. <https://doi.org/10.1016/j.autcon.2005.01.005>
- 20- Al-Azzawi, A. A. (2016). Exploring the factors affecting the elemental cost estimation with relationship analysis using AHP. Journal of Engineering, 22(8), 1–13. <https://doi.org/10.31026/j.eng.2016.08.13>



Human, Web, or AI? A Comparative Study of Reliability and Stability in Three Construction Cost Estimation Approaches

Shayan Fallahi¹, Shayan Hojatpanah², Sahar Alinejad Majidi^{3*}

1-M.Sc., Department of Project and Construction Management, Shahid Beheshti University, Tehran, Iran.

sh.fallahi@mail.sbu.ac.ir

2-Ph.D. Candidate, Department of Civil Engineering, Faculty of Engineering and Construction Management, University of Tehran, Tehran, Iran.

shayan.hojatpanah@ut.ac.ir

3-Assistant Professor, Department of Architectural Engineering, Islamic Azad University, Shahrood Branch, Semnan, Iran. (Corresponding Author)

s.alinejad@iau-shahrood.ac.ir

Extended Abstract

Aims: Accurate cost estimation is one of the fundamental prerequisites for successful planning and management of construction projects. The emergence of generative artificial intelligence (AI) models and online cost estimation systems has heightened the need to evaluate the performance of these tools in comparison with human expert judgment. Despite the proliferation of these technologies, no comprehensive study has simultaneously compared human experts, online systems, and generative AI models based on actual project costs. Existing research has largely focused on individual approaches in isolation, failing to provide an integrated comparative framework to guide practitioners. Moreover, the repeatability of AI-generated responses across multiple iterations with a fixed prompt has received limited attention, and the attitudes and trust levels of practicing engineers toward these novel tools remain underexplored. This research aims to address these gaps by systematically comparing the three aforementioned approaches in the context of the "Atashgah" building project in Semnan, Iran. Specifically, the study pursues four main objectives: (1) to evaluate differences in mean cost estimates among human experts, online systems, and generative AI models; (2) to assess the accuracy of each approach using MAE and MAPE relative to actual project cost, and measure their stability through coefficient of variation (CV); (3) to examine the repeatability of five AI models across 10 iterations with a fixed prompt; and (4) to investigate engineers' attitudes toward these tools and determine their level of trust in the outputs of online systems and generative AI.

Materials & Methods: This applied descriptive–analytical study adopted a survey-based comparative case study design. The case project was the *Atashgah* building, a six-story mixed-use (residential–commercial) structure located in Semnan, Iran, with a total floor area of 2,622 m². The building consists of a reinforced concrete frame with block-joint slabs and includes first-grade Iranian construction materials, a brick-and-stone façade, complete mechanical and electrical installations, a six-person elevator, and double-glazed windows. The actual construction cost (81.5 billion IRR), equivalent to 31.08 million IRR/m², served as the reference value for evaluating estimation accuracy. Data were collected through three parallel sources. First, a questionnaire was completed by 29 construction professionals representing different educational levels (65.51% bachelor's, 24.14% master's, 6.90% diploma, and 3.45% doctoral degrees), work experience (17.24% <5 years, 31.03% 5–10 years, 27.59% 11–20 years, and 24.14% >20 years), and professional roles (contractors, supervisors, site managers, and project managers). Participants estimated the project cost per square meter and reported their confidence in their own estimates (1–10), trust in online estimation systems (1–10), trust in AI-based estimations (1–10), trust in the official engineering tariff system (1–10), previous experience using online platforms and AI, and reasons for distrust through open-ended questions. The questionnaire demonstrated satisfactory validity (CVR = 0.82, CVI = 0.86) and reliability (Cronbach's α = 0.84). Second, eight active Iranian online cost estimation platforms (namaplan.com, dezharco.com, bkesh.ir, ahan100.com, karkoshte.com, pardissazeh.com, peymanu.com, and khanehtarh.ir) were evaluated. The same project specifications were entered into each platform according to its input requirements, and the estimated cost per square meter and total project cost were recorded. Because *peymanu.com* and *khanehtarh.ir* generated estimates based on a standard area of 100 m² using 2024 prices, their outputs were scaled to the project's actual area (2,622 m²) and updated to 2025 values using the official annual construction inflation rate of 47%. Third, five widely used generative AI models—ChatGPT, Gemini, Claude, DeepSeek, and Grok—were evaluated. These models were selected to represent different AI architectures (proprietary and open-source), geographic origins (U.S. and Chinese models), accessibility, and frequency of use reported by the surveyed engineers. All models received an identical prompt describing the complete project specifications, including building type, location, structural system, floor area,

material quality, façade, elevator, building services, and window specifications. Each model generated 10 independent estimates using a new conversation for every iteration, resulting in 50 AI-generated responses. Web search functionality was disabled, and all evaluations were conducted using free-tier accounts. Estimation performance was assessed using the mean, standard deviation (SD), coefficient of variation ($CV = SD/Mean \times 100$), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) relative to the actual project cost. Differences among estimation methods were examined using the Kruskal–Wallis test followed by Mann–Whitney U tests with Bonferroni correction. Engineers' confidence and trust ratings were summarized using descriptive statistics, while qualitative content analysis of the open-ended responses was conducted to evaluate the extent to which AI models considered local construction conditions, material quality, and market-related risks.

Findings: The results revealed that none of the three approaches individually possessed sufficient reliability for standalone application in real projects. Human expert estimates showed a mean of 30.83 million IRR/m², remarkably close to the actual cost of 31.08 million IRR/m² (MAPE = 0.80%), but exhibited considerable dispersion ($CV = 34.64\%$), with estimates ranging from 10 to 65 million IRR/m² and a median of 30 million IRR/m². Despite this variability, engineers reported relatively high confidence in their own estimates (mean = 7.04 out of 10). Online systems exhibited the lowest reliability, with estimates ranging from 25 to 53.49 million IRR/m²—a 157% difference between the lowest and highest estimates. The mean estimate was 37.40 million IRR/m², with a standard deviation of 9.09 million and $CV = 24.29\%$. This represented a 20.34% error relative to actual project cost, indicating a systematic tendency toward overestimation. Among the AI models, Grok demonstrated the best performance, with a mean estimate of 31.80 million IRR/m² (MAPE = 2.32%, $CV = 5.32\%$), and a narrow range of 71–99 billion IRR total cost. Gemini ranked second with MAPE = 12.48% and $CV = 19.4\%$. ChatGPT and DeepSeek showed moderate performance with MAPEs of 21.66% and 22.72%, and CVs of 14.09% and 34.14%, respectively. Claude performed poorly, with MAPE = 63.77%, $CV = 71.23\%$, and high instability across iterations (range: 9.5–75 billion IRR); it also refused to provide numerical estimates in 3 out of 10 iterations. Engineers expressed low trust in non-human approaches, with mean trust scores of 4.89 for the official tariff system, 4.68 for online sites, and 4.67 for AI—all below 5 out of 10. Only 35.71% had used AI before, and merely 20.69% had used online sites. The Kruskal–Wallis test confirmed significant differences among the three approaches ($H = 24.6$, $df = 2$, $p < 0.001$). However, post-hoc Mann–Whitney tests with Bonferroni correction revealed that only the difference between online systems and AI models was statistically significant ($p < 0.001$). Qualitative analysis of AI responses indicated that Gemini and DeepSeek showed greater attention to local conditions (e.g., "price fluctuations in Iran" and "cost differences in provincial cities"), while ChatGPT and Grok provided standard, repetitive explanations, and Claude offered general, non-specialized content.

Conclusion: This study concludes that no single approach to construction cost estimation is sufficiently reliable for independent application. Human experts, while confident, exhibit significant variability. Online systems lack both accuracy and consistency. AI models show promise but vary widely in performance and stability. The hybrid "human-in-the-loop" approach is proposed as the most effective strategy, wherein a stable AI model (such as Grok) generates an initial estimate, which is then validated, refined, and completed by human experts based on their contextual knowledge and experience. Engineers and project managers are advised to evaluate AI model stability across multiple iterations and select models with $CV < 15\%$ for practical applications. Professional engineering organizations should develop formal guidelines and training programs to enhance trust and adoption of AI technologies. Future research should expand the sample size, replicate the study across diverse projects and cities, and develop explainable AI tools to improve transparency and trust. Despite rapid AI advancements, engineering judgment remains indispensable, and new technologies should serve as complementary tools augmenting, not replacing, human expertise.

Key words: Construction cost estimation, Human experts, Online systems, Generative artificial intelligence, Reliability.



This Journal is an open access Journal Licensed under the Creative Commons Attribution 4.0 International License

(CC BY 4.0)