ARTIFICIAL INTELLIGENCE (AI) AGENTS IN CONSTRUCTION: A CONCEPTUAL FRAMEWORK ACROSS AUTONOMY TYPES AND APPLICATION DOMAINS

Reihaneh Samsami, Ph.D., P.E. (Corresponding Author)
Department of Civil and Environmental Engineering
University of New Haven
300 Boston Post Road, West Haven, CT 06516, USA
Email: rsamsami@newhaven.edu

Submitted: 6/23/2026

ABSTRACT

Construction is one of the least digitized and most hazard-prone industries, and signs of risk often go unconnected to timely action, as the 2025 NIST findings on the Champlain Towers South collapse illustrate. Artificial intelligence (AI) agents have been proposed as a response, but the term spans computer vision monitors, physical robots, and large language model (LLM) reasoning systems studied in separate communities. This paper develops a conceptual framework connecting agent autonomy to the construction problem it addresses. The study is a structured review and synthesis of peer-reviewed and industry literature published mainly between 2014 and 2026, drawn from construction engineering, robotics, and computer science. The framework organizes agents along two axes: autonomy type (perceptual, cognitive, embodied) and application domain (safety monitoring, design and modeling, project and contract management, progress monitoring, on-site automation), rating the maturity of each pairing by weight of evidence. Perceptual agents are mature where the task is to see and report, such as safety and progress monitoring. Embodied agents are mature only for narrow physical tasks such as aerial survey. Cognitive agents reach every domain but are mature in none, and remain unreliable for safety-critical calculation. The most capable systems combine agent types rather than one alone. The framework provides the first two-axis view that unifies three separate agent literatures and makes explicit where construction capabilities are mature, overlapping, or absent. It gives transportation agencies a tool for matching an agent type to a construction need and prioritizing investment, and identifies three deployment requirements, namely human oversight, system integration, and trust, that practitioners must meet for AI agents to move from pilots to practice.

1 INTRODUCTION

In September 2025, the National Institute of Standards and Technology (NIST) reported that its investigation into the 2021 partial collapse of the Champlain Towers South building was nearing completion (NIST 2025). Using photographs, maintenance records, and eyewitness reports, the investigation found that signs of the building's distress had been visible for years and even in the final weeks before the collapse. Cracks had been recorded in the first-floor slab in the months and years prior, concentrated in the pool deck and parking areas, and in the weeks before the failure a planter wall cracked, a gate shifted and jammed, and water leaked from the garage ceiling at locations that had cracked and been repaired many times before (NIST 2025). These were observable conditions. The difficulty was not that the warnings were invisible, but that no continuous and reliable process connected what was seen to a timely response.

This difficulty is not limited to one building. Across construction and infrastructure work, the volume of information on a project often exceeds what a human team can observe, interpret, and act on in time. A great deal depends on human observation, and a single missed detail can be very costly.

This pattern sits behind several long-standing problems in construction. The first is low productivity. Over the past two decades, labor productivity growth in construction has averaged about 1 percent a year, compared with 2.8 percent for the world economy and 3.6 percent for manufacturing (Barbosa et al. 2017). Construction is also among the least digitized sectors in the world, ranking near the bottom of the McKinsey Global Institute digitization index (Barbosa et al. 2017). The second problem is a shortage of skilled and experienced staff, which has grown worse as a large share of the workforce nears retirement and the number of job vacancies rises (Barbosa et al. 2017). The third problem is safety. Construction remains one of the most dangerous industries, and many of its injuries and fatalities are tied to hazards that are difficult to monitor continuously by eye (Pan and Zhang 2021).

These problems share a common structure. Each one appears where the amount and speed of site information is greater than the capacity of a human team to track it, reason about it, and respond in time. This is the class of problem that AI agents are designed to address. An AI agent is a system that perceives its environment, reasons toward a goal, and takes or recommends action, often continuously and with limited human intervention (Luo et al. 2025). This is different from the analytical AI that construction has used for years, such as models that predict concrete strength or classify schedule risk. Those tools inform a human decision, but they do not act. An agent monitors, decides, and acts.

Interest in AI agents for construction has grown quickly since 2023. Two developments drove this growth. The first is the maturing of computer vision systems that can detect hazards and track progress in real time (Pan and Zhang 2021). The second is the rapid rise of agents built on large language models (LLMs), which can read unstructured text such as instructions, contracts, and codes, and can call digital tools to act on what they read (Kampelopoulos et al. 2025). Industry sources report sizable benefits from AI adoption, including reductions in accidents, project cost, and engineering hours (Datagrid 2025). At the same time, adoption remains shallow, and many firms report no AI use at all (RICS 2025).

A central obstacle to progress is conceptual rather than technical. The term AI agent is applied to at least three very different kinds of system. The first is the perceptual or software agent, built on computer vision and machine learning. The second is the embodied or robotic agent, which physically acts on the world. The third is the cognitive agent, built on an LLM, which reasons over language and coordinates tools (Pan and Zhang 2021; Parascho 2023; Luo et al. 2025). These kinds differ in their autonomy, their capabilities, their failure modes, and the tasks they suit. They are usually studied in separate research communities. As a result, a project manager or researcher has no clear way to connect the kind of agent autonomy that is available to the construction problem that needs solving.

The objective of this paper is to address this gap by developing a conceptual framework that organizes AI agents in construction along two axes, namely the agent autonomy type and the construction application domain. The framework is meant to serve two purposes. For researchers, it is a lens that brings three separate literatures into one view. For practitioners, it is a tool for matching an agent type to a construction need. The primary contributions of this paper are threefold. First, it presents a taxonomy of agent autonomy types relevant to construction, and places perceptual, embodied, and cognitive agents on one autonomy spectrum. Second, it maps these agent types onto the major construction application domains, and shows where capabilities are mature, where they overlap, and where they are absent. Third, it synthesizes recent peer-reviewed and industry evidence through the framework, and from this identifies the requirements and open research needs that matter most for transportation infrastructure construction.

2 BACKGROUND

2.1 Defining the AI Agent

The construction industry is widely recognized for its information-intensive nature (Memon et al. 2005). On an active site, the volume of information often exceeds what a human team can observe, interpret, and act on in time. This gap is where AI agents are positioned to help. Before reviewing their applications, it is important to define what an AI agent is and how it differs from the analytical tools already used in construction.

In classical AI, an agent is a system that perceives its environment through sensors and acts on that environment through actuators, in pursuit of a goal (Pan and Zhang 2021). What has changed is not this definition but the capability behind it. Advances in computer vision, robotics, and LLMs have made agents more perceptive, more general, and more autonomous than earlier rule-based systems (Luo et al. 2025). Three properties separate an agent from a passive tool. First, an agent is goal-directed, which means it acts toward an objective rather than producing a single output on request. Second, an agent is situated, which means it perceives and responds to a changing environment, whether that environment is a video feed, a job site, or a stream of project documents. Third, an agent is autonomous to some degree, which means it can take or recommend action with limited human intervention (Luo et al. 2025).

This distinction matters in construction. The sector has used analytical AI for years, including predictive models for concrete strength, schedule-risk classifiers, and cost estimators (Pan and Zhang 2021). These tools inform human decisions, but they do not act. The shift toward agents is a shift from systems that record and predict toward systems that monitor, decide, and act, often continuously. An agent turns information that enters a system into a timely response rather than a record that remains unexamined.

2.2 A Spectrum of Agent Autonomy

The systems labeled AI agents in construction are not a single technology. They span three classes, separated by what they sense, what they act on, and how they reason. This paper organizes them as (1) perceptual agents, (2) embodied agents, and (3) cognitive agents. **Figure 1** places these three classes on a common autonomy spectrum.

Perceptual agents, also called software agents, are built mainly on computer vision and machine learning. They perceive the site through cameras, drones, or sensors, and they act by classifying, flagging, and alerting. The dominant technology is deep-learning object detection, and the You Only Look Once (YOLO) family is the most widely adopted because it balances speed and accuracy in real-time conditions (Redmon et al. 2016; Delhi et al. 2020). These agents are mature and narrowly capable. They detect predefined objects and events well, but they are limited in their ability to reason about context (Kampelopoulos et al. 2025).

Embodied agents, also called robotic agents, physically act on the world. They include autonomous mobile robots, robotic arms for bricklaying and welding, automated guided vehicles for material transport, autonomous earthmoving equipment, and unmanned aerial vehicles (UAVs) for survey and inspection (Melenbrink et al. 2020; Parascho 2023). Over time, the field has moved from pure automation of repetitive tasks toward collaborative robots that share the workspace with human workers (Parascho 2023). Their main challenge is operating reliably in the dynamic and unpredictable construction environment (Melenbrink et al. 2020).

Cognitive agents, also called LLM-based agents, are the newest class. They reason over unstructured language, including instructions, specifications, contracts, and regulations, and they increasingly control external tools through software interfaces (Kampelopoulos et al. 2025). In construction, a domain-adapted LLM can interpret an engineer's instruction, retrieve relevant knowledge such as codes and building information modeling data, and call digital tools through application programming interfaces (Kampelopoulos et al. 2025). Their strength is general reasoning. Their main weakness is reliability, because errors and weak numerical precision make them unsuited, for now, to safety-critical calculation (Kampelopoulos et al. 2025).

These three classes are increasingly combined. A growing body of work pairs a perceptual agent with a cognitive one, for example by linking a YOLO detector to an LLM so that the system not only detects a hazard but also reasons about its context (Poudel et al. 2025). Embodied systems also embed perceptual and cognitive parts, as in drone platforms that combine real-time object detection with language-based risk assessment for site supervision (Poudel et al. 2025). The boundaries between the classes are narrowing, which is the reason a framework that holds all three in one view is useful.

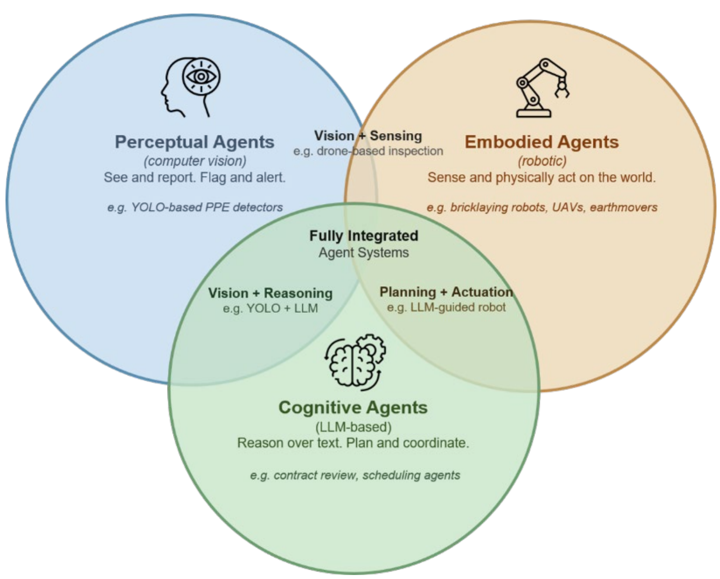

Figure 1 Spectrum of AI agent autonomy in construction.

2.3 Current Adoption in Construction

Although research interest in AI agents is rising quickly, their adoption on real projects is still uneven. A survey of more than 2,200 professionals worldwide by the Royal Institution of Chartered Surveyors found that about 45 percent of firms reported no AI implementation, and about 34 percent were in an early pilot phase (RICS 2025). A common explanation is that project data is trapped in separate systems, including building information models (BIMs), equipment sensors, and contractor communications, which prevents the single, real-time view that agents depend on (Datagrid 2025).

Where AI has been adopted, the reported gains are sizable. Industry analyses report improvements on the order of 10 to 35 percent across safety, cost, and efficiency measures, including accident reductions near 30 to 35 percent, project cost savings near 10 to 15 percent, and reductions near 10 to 30 percent in engineering hours (Datagrid 2025). Predictive maintenance is reported as one of the fastest-growing applications, with equipment downtime reductions of up to 45 percent among early adopters (RTS Labs 2026). These figures come from industry reporting rather than controlled study, so they should be read as indicative rather than exact. Their consistency nonetheless points to a real and sizable opportunity.

Peer-reviewed reviews give a more structured picture of where AI research is concentrated. A large critical review of AI in construction engineering and management found that the main focus of current research is on safety monitoring and control and on process management, supported mostly by machine learning, computer vision, and natural language processing (NLP) (Pan and Zhang 2021; Razi et al. 2026). A separate review of LLMs in construction identified seven main application areas, namely training and education, planning and scheduling, safety analysis and hazard recognition, document generation and compliance checking, virtual assistants, code and data generation, and building information modeling functions (Kampelopoulos et al. 2025). That same review found almost no work that uses LLMs for core engineering calculations such as structural analysis and load computation, which reflects known limits in numerical precision and the high stakes of safety-critical work (Kampelopoulos et al. 2025). This pattern, in which agents cluster in information-heavy coordination tasks and avoid safety-critical numerical tasks, appears again throughout the analysis in Section 4.

3 A CONCEPTUAL FRAMEWORK FOR AI AGENTS IN CONSTRUCTION

The previous section described three kinds of AI agent and noted that they are usually studied apart from one another. This section brings them together. It first explains how the framework was built, and then presents it. The framework organizes AI agents in construction along two axes. The first axis is the agent autonomy type, which describes the nature of the agent. The second axis is the construction application domain, which describes the problem the agent is applied to. The purpose of the framework is to give researchers a single view of a fragmented field, and to give practitioners a way to match an agent type to a construction need.

3.1 Framework Development

Because this paper proposes a framework rather than reporting an experiment, the framework was built through a structured review and synthesis of the literature. The review covered AI agents applied to the construction phase of building and infrastructure projects, with attention to transportation infrastructure where possible. A system was treated as an AI agent if it met the three properties defined in Section 2.1, namely goal-directed behavior, situatedness, and some degree of autonomy. Purely analytical tools that predict or classify without acting were used only as background, because they mark the line that separates agents from earlier AI. The review favored peer-reviewed literature published between 2014 and 2026, the period in which deep learning, construction robotics, and LLM agents matured. A small number of older sources were kept where they are the original source of a method or definition. A limited set of industry reports was used only for adoption statistics, which are not available in the peer-reviewed literature.

The two axes were derived from this body of work. The agent autonomy axis was formed by grouping the reviewed systems according to what they sense, what they act on, and how they reason. This grouping produced the three classes described in Section 2.2, namely perceptual, embodied, and cognitive agents. The application domain axis was formed from the application areas that recur across the major reviews of AI in construction (Pan and Zhang 2021; Kampelopoulos et al. 2025). These recurring areas were grouped into five domains, namely safety monitoring, design and modeling, project and contract management, progress monitoring, and on-site automation. The two axes were chosen because they are close to independent. The autonomy axis describes the nature of the agent, while the domain axis describes the construction problem, so a given system can be placed by one coordinate on each axis.

Each cell of the resulting grid was then rated for the maturity of its evidence, using three levels. A mature cell has multiple peer-reviewed studies or commercial deployments. An emerging cell has early studies or pilot deployments. A sparse cell has little or no published work. This rating is qualitative and reflects the weight of the evidence found in the review rather than a count of papers. It is meant to show where capabilities are strong, where they are developing, and where gaps remain. This reading has limits, which are stated plainly here.

3.2 The First Axis: Agent Autonomy Type

The first axis follows directly from the three classes defined in Section 2.2. It is ordered by how an agent reasons and whether it acts on the physical world. The three types are arranged as a spectrum rather than as separate boxes, because real systems often combine them.

Perceptual agents sit at one end. They sense the site and reason about what they see in a narrow way, and they act by reporting, flagging, or alerting rather than by changing the physical world. Their autonomy is bounded. They do one task well, such as detecting a worker without a hard hat, but they do not decide what to do about it (Kampelopoulos et al. 2025).

Cognitive agents sit in the middle of the spectrum in terms of physical action, but at the high end in terms of reasoning. They do not act on the physical world directly. Instead, they reason over language and data, make plans, and call digital tools (Kampelopoulos et al. 2025). Their autonomy is broad but abstract. They can decide and coordinate, but their actions stay inside software.

Embodied agents sit at the far end in terms of physical action. They sense the site, and they also move and build (Melenbrink et al. 2020). Their autonomy is the most complete in principle, because they close the loop from perception to physical action. In practice their autonomy is the hardest to achieve, because the construction site is dynamic and unpredictable, and current robotic technology is not yet able to build in a fully autonomous way without human oversight (Melenbrink et al. 2020).

Figure 1, introduced in Section 2.2, places the three types on this spectrum and shows that the types increasingly overlap, since perceptual and cognitive parts are now embedded inside embodied systems, and perceptual agents are now paired with cognitive ones to add reasoning.

3.3 The Second Axis: Construction Application Domain

The second axis lists the construction problems that AI agents are applied to. These domains were drawn from the application areas that recur across the major reviews of AI in construction

(Pan and Zhang 2021; Kampelopoulos et al. 2025), as described in Section 3.1. Five domains are used in this paper.

The first domain is safety monitoring, which covers the detection of hazards, unsafe behavior, and missing protective equipment, and the assessment of risk on the site (Pan and Zhang 2021). The second domain is design and modeling, which covers design review, code and clause checking, clash detection, and the creation and use of building information models (Zhang and El-Gohary 2017). The third domain is project and contract management, which covers planning and scheduling, cost and progress oversight, document routing, and the review of contracts and other text (Zhang and El-Gohary 2016). The fourth domain is progress monitoring, which covers tracking the state of work against the plan, often using site imagery, and comparing as-built conditions to as-planned conditions (Samsami et al. 2021). The fifth domain is on-site automation, which covers the physical work of building, including earthmoving, material handling, placement, and survey and inspection by air (Melenbrink et al. 2020; Parascho 2023).

3.4 The Combined Framework

The two axes combine into a grid. Each row is an agent autonomy type, and each column is a construction application domain. Each cell describes how well a given agent type serves a given domain, based on the evidence collected in the review and using the three maturity levels defined in Section 3.1. **Table 1** presents this grid as a shaded matrix so that the patterns are visible at a glance.

TABLE 1 Mapping of Agent Autonomy Types onto Construction Application Domains

Agent type	Safety monitoring	Design and modeling	Project and contract mgmt.	Progress monitoring	On-site automation
Perceptual (software)	Mature. PPE, hazard, and unsafe-behavior detection (Delhi et al. 2020).	Emerging. Vision aids as-built model creation and defect detection (Pan and Zhang 2021).	Sparse. Mainly supplies data to other agents.	Mature. Image-based tracking of work vs. schedule (Samsami et al. 2021).	Emerging. Sensing layer for robots and drones (Poudel et al. 2025).
Cognitive (LLM-based)	Emerging. Reasoning over detected hazards; paired with vision (Poudel et al. 2025).	Emerging. Code and clause checking, design review (Zhang and El-Gohary 2017).	Emerging. Scheduling, document routing, contract review (Zhang and El-Gohary 2016).	Emerging. Interpreting and reporting progress data (Kampelopoulos et al. 2025).	Sparse. Can plan and instruct, but not act on site.
Embodied (robotic)	Emerging. Drones and robots that inspect and supervise (Poudel et al. 2025).	Sparse. Consumes models to act.	Sparse. Little direct management role.	Mature. UAVs for survey, mapping, inspection (Samsami et al. 2021).	Emerging. Bricklaying, welding, earthmoving with oversight (Melenbrink et al. 2020).

Note: Mature = multiple peer-reviewed studies or commercial deployments; Emerging = early studies or pilot deployments; Sparse = little or no published work.

Three patterns stand out in **Table 1**, and they organize the analysis in Section 4. First, perceptual agents are most mature where the task is to see and report. They are strongest in safety monitoring and progress monitoring, the two domains built on detecting and tracking visual events (Delhi et al. 2020; Samsami et al. 2021). Second, cognitive agents are spread across every domain at an emerging level. This breadth reflects their general reasoning ability. They are the only agent type that touches all five domains, but in none of them are they yet mature, and they remain weak where numerical precision and safety are critical (Kampelopoulos et al. 2025). Third, embodied agents are concentrated at the two ends of the project, in physical survey and inspection and in physical building. They are mature in aerial progress monitoring, where the task is well defined, but only emerging in on-site automation, where the site is unpredictable and full autonomy is still out of reach (Melenbrink et al. 2020).

The empty/sparse cells are as informative as the full ones. They show where an agent type has little to offer a domain, which helps a practitioner avoid forcing the wrong tool onto a problem. The framework therefore works both as a map of what exists and as a guide to what does not.

4 APPLICATION DOMAIN ANALYSIS

The analysis presented in this section addresses each of the five construction application domains identified in the framework. For each domain, it describes the problem, traces how the agent types have been applied and how they have changed over time, reports what the evidence shows about their accuracy and maturity, and states the limits that remain. The analysis follows the framework in Section 3 and the mapping in **Table 1**. Two figures support the analysis. **Figure 2** reports the ranges of improvement that industry sources attribute to AI adoption across safety, cost, and efficiency. **Figure 3** reports the adoption picture, which shows wide interest alongside shallow use.

4.1 Safety Monitoring

Safety is the domain where AI agents have the longest record in construction. The Occupational Safety and Health Administration (OSHA) identifies four leading causes of construction fatalities, known as the Fatal Four: falls, struck-by incidents, electrocutions, and caught-in or caught-between accidents (OSHA 2024). These four hazards account for more than 60 percent of all construction worker deaths annually (OSHA 2024). In 2022, 1,069 construction workers died on the job, a rate of 9.6 fatalities per 100,000 full-time workers (Bureau of Labor Statistics 2022). Many of these fatalities arise from observable conditions such as a missing hard hat, a worker too close to moving equipment, or an unguarded edge. The problem is not that these conditions cannot be seen. It is that a human supervisor cannot watch every worker and every location at once, and manual safety inspection is costly and intermittent (Pan and Zhang 2021; Badhan and Samsami 2025).

Perceptual agents address this directly, and they are the most mature agents in any domain in this paper. The dominant tool is the deep-learning object detector, and the YOLO family is the most common because it processes images fast enough for real-time use (Redmon et al. 2016). The work has advanced quickly through successive versions. Delhi et al. (2020) used the third version of YOLO to classify workers into four safety states based on hard hat and vest use, and reported precision and recall near 96 percent on their test data. Later studies widened the set of detected classes and improved accuracy under harder conditions. Ferdous and Ahsan (2022) introduced an anchor-free detector trained on a purpose-built construction dataset of eight classes, and more recent lightweight models tuned for hard hat detection report mean average precision above 94 percent (Zhang et al. 2025; Samsami and Jafari Kang 2026). The trend across these studies is steady gains in accuracy and the addition of more hazard classes over time.

The limits of the perceptual agent are also clear from the evidence. Detection accuracy falls for small or distant objects, in poor light, and on blurred faces, which are common conditions on a real site (Ferdous and Ahsan 2022). More importantly, the detector sees but does not reason. It can report that a worker is near an excavator, but it does not judge whether that nearness is dangerous in context (Kampelopoulos et al. 2025). This gap is where cognitive agents are entering safety. Recent work pairs a detector with an LLM or a vision-language model (VLM), so the system not only detects objects but also reasons about the relationship between them and assesses the level of risk (Poudel et al. 2025; Samsami and Badhan 2025). Embodied agents are entering as well, mainly as UAVs that carry detectors over the site to supervise areas that are hard to reach or unsafe to enter (Poudel et al. 2025). In the framework, safety is a domain where perceptual agents are mature, while cognitive and embodied agents are emerging and tend to build on the perceptual layer.

4.2 Design and Modeling

The design and modeling domain covers the review and checking of designs, the detection of conflicts between systems, and the creation and use of BIMs. The problem is that a modern design carries a large volume of rules, codes, and clauses, and checking a design against all of them by hand is slow and prone to error (Zhang and El-Gohary 2016).

Cognitive agents are the natural fit for this domain, and the work here predates the current wave of LLMs. Zhang and El-Gohary (2016, 2017) developed semantic NLP methods that extract requirements from regulatory documents and check a building design against them automatically. These methods showed that text-heavy compliance checking could be at least partly automated, and they set the pattern that later work followed. LLMs have extended this line, because they can read codes and clauses and answer questions about a design or a model (Kampelopoulos et al. 2025). The gains reported are concrete. In one study on bridge construction schemes, a method that combined a pretrained language model with a fine-tuned LLM reached a sentence-level semantic matching accuracy of about 92 percent when checking a scheme against code provisions (Zhu et al. 2024). Work on building design has used LLMs to convert a written brief into structured specifications and to critique design options against codes in an auditable form (Jang and Lee 2024).

Perceptual agents play a supporting role in this domain, mainly by helping create as-built models from site imagery, which then feed design and quality checks (Pan and Zhang 2021). Embodied agents play little direct role in design, though some bricklaying robots now take their instructions directly from a BIM (FBR 2024). In the framework, design and modeling is a domain led by cognitive agents at an emerging level, with perceptual support and little embodied presence.

4.3 Project and Contract Management

Project and contract management covers planning and scheduling, cost and progress oversight, the routing of documents, and the review of contracts and other text. The problem in this domain is coordination. A construction project generates a constant stream of changes across schedules, costs, and documents, and each change needs review, routing, or follow-up. When this work is manual, delays become routine (CMiC 2025).

This domain is well suited to cognitive agents, because most of the work is reasoning over language and data. NLP has been applied to construction contracts to detect risky clauses, extract obligations, and check compliance, which reduces the time and error of manual review (Zhang and El-Gohary 2016). The use of generative AI tools, including large language models, in the AEC industry has also been studied from a prompt engineering perspective, showing that how a question is framed substantially affects the quality and reliability of the output (Samsami 2024b). LLMs have extended this to planning. He et al. (2024) analyzed transcripts from 94 on-site planning meetings, spanning more than 260,000 words, and built a model that classified the constraints discussed in those meetings. Their tool improved constraint classification accuracy by about 9 percent over a traditional model, and it helped reveal that managers and foremen focus on different kinds of constraints. A follow-on study used a knowledge graph to enrich an LLM’s answers to constraint questions, and reported a large gain in correctness over a standard retrieval method (He et al. 2026). Broader reviews of automated planning add an important caution. The strongest results come when an LLM is combined with a classical planner rather than used alone, because the language model supplies flexibility while the planner supplies precision (Pallagani et al. 2024). This combined pattern matters for construction, where a schedule must be both sensible and exact.

Perceptual and embodied agents play little direct role in management, though perceptual agents supply the progress and cost data that management agents act on. In the framework, project and contract management is a domain led by cognitive agents at an emerging level.

4.4 Progress Monitoring

Progress monitoring is the continuous assessment of how far the work has advanced, and the comparison of the as-built state with the as-planned state (Samsami et al. 2021). The problem is that traditional progress monitoring is manual, subjective, and slow, since it depends on staff collecting field reports and comparing them to the plan by hand (Reja et al. 2022; Samsami 2024a).

This domain is led by perceptual and embodied agents working together, and the combination is mature. The common pipeline collects site imagery, often from a UAV, reconstructs the as-built geometry as a point cloud, and compares it to a building information model to estimate progress (Reja et al. 2022). Reviews of automated computer vision-based progress monitoring describe a standard four-step process of data acquisition, three-dimensional reconstruction, as-built modeling, and progress assessment against a four-dimensional model that carries the schedule (Reja et al. 2022). UAVs are well suited to this task because they collect large volumes of visual data quickly, without contact with the surface, and at lower cost and risk than manual survey (Samsami et al. 2021). The same imagery supports quality control. For example, UAV-mounted thermal cameras can detect temperature segregation in fresh asphalt pavement, and produce a full thermal profile of the mat that a few manual samples cannot match (Samsami et al. 2023). This case is directly relevant to transportation construction, where pavement quality is tied to long-term performance. UAV-based inspection has also been applied to bridge deck condition assessment, where machine learning and deep learning approaches using UAV-collected imagery demonstrated that automated crack detection and condition rating can substantially reduce the time and safety risk of conventional manual bridge inspection (Pokhrel et al. 2024; Khatry et al. 2026). A comprehensive review of UAV platforms and sensors for bridge inspection confirms that these systems significantly reduce inspection time, labor requirements, and safety risks compared with conventional methods (Samsami 2026).

The limits of this domain are practical rather than conceptual. Automated progress methods depend on the quality and completeness of the scan, and they struggle when the view is blocked by temporary objects on a busy site (Reja et al. 2022). Cognitive agents are entering the domain to interpret and report progress data and to coordinate across the systems that hold it (Kampelopoulos et al. 2025). In the framework, progress monitoring is a domain where perceptual and embodied agents are mature, and cognitive agents are emerging as an interpretation and reporting layer.

4.5 On-Site Automation

On-site automation covers the physical work of building, including earthmoving, material handling, placement of components, and survey and inspection by air. The problem in this domain is the hardest of all, because the agent must act on the physical world, and the construction site is dynamic, unstructured, and unpredictable (Melenbrink et al. 2020).

This domain belongs to embodied agents, and the evidence shows both their potential and their constraints. Bricklaying is the clearest example. The Semi-Automated Mason, a robot that works alongside a human mason, can support a daily output of up to about 3,000 bricks, compared with the 300 to 500 a skilled worker lays by hand in a full working day (Construction Robotics 2020; Madsen 2019). A more autonomous system, the Hadrian X, can lay more than 1,000 bricks per hour and place them to sub-millimeter accuracy from a three-dimensional model (FBR 2024). Autonomous earthmoving is also moving toward practice, with firms retrofitting standard machines for self-operation to speed excavation and reduce injury risk (Built Robotics 2025). Over time the field has shifted from pure automation of fixed, repetitive tasks toward collaborative robots that share the workspace with human workers and adapt to changing conditions (Parascho 2023).

The evidence is equally clear that full autonomy remains beyond current capability. Even the most advanced bricklaying robots still need a human to load the material, to monitor the operation, and to correct any error, and they typically build only part of the structure (FBR 2024). Melenbrink et al. (2020) concluded that current robotic technology is not suitable for fully autonomous building, mainly because the level of perception and adaptive control required is still beyond what is reliable on a real site. For this reason, most on-site robots operate under human oversight. Perceptual agents serve as the sensing layer for these robots, and cognitive agents are beginning to serve as a planning layer that instructs them, but neither acts on the site by itself (Poudel et al. 2025). In the framework, on-site automation is a domain led by embodied agents at an emerging level, mature only in the narrow and well-defined case of aerial survey and inspection.

4.6 Cross-Domain Reading

Read across the five domains, the framework reveals a clear division of labor. Perceptual agents are mature where the task is to see and report, which is safety monitoring and progress monitoring. Embodied agents are mature only where the physical task is narrow and well defined, which is aerial survey, and they are still emerging in general building. Cognitive agents touch every domain, but they are mature in none, and they remain weak where numerical precision and safety are critical (Kampelopoulos et al. 2025). A second pattern cuts across the domains as well. In every domain, the newest work combines agent types rather than using one alone, whether that is a detector paired with a language model for safety, or a language model paired with a classical planner for scheduling. The reported benefits of AI adoption, summarized in **Figure 2**, are sizable but come mostly from industry sources, and the adoption gap in **Figure 3** shows that interest is far ahead of practice (RICS 2025). The next section discusses the requirements that must be met for this gap between interest and practice to close.

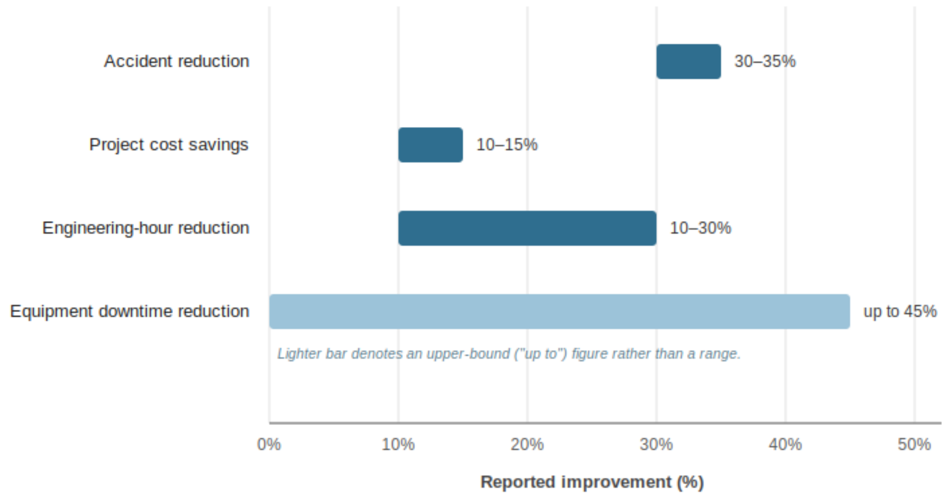

Figure 2 Reported ranges of improvement from AI adoption in construction (Datagrid 2025; RTS Labs 2026). Figures are industry- and survey-reported and are indicative, not results of controlled study.

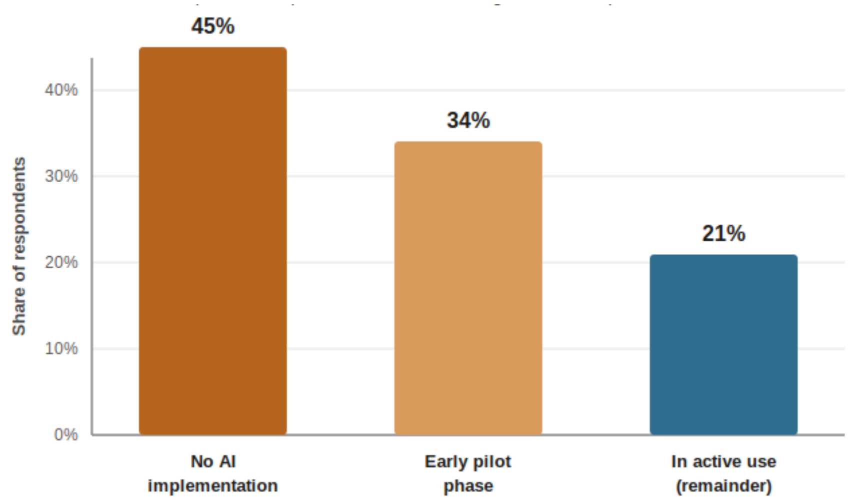

Figure 3 The adoption gap between interest and implementation, based on the RICS (2025) survey.

5 DISCUSSION

The analysis in Section 4 showed where each agent type is mature, where it is emerging, and where it is absent. Across all five domains, three requirements recur as the conditions that decide whether an agent can move from a research result to real use. These are human oversight, system integration, and trust. This section discusses each in turn, and then states the limits of current agents and of this study.

5.1 Human Oversight

The first requirement is human oversight, and the evidence for it is consistent across every domain. No agent type reviewed in this paper operates well without a human in the loop. Perceptual agents flag a possible hazard, but a person decides whether it is real and what to do about it (Kampelopoulos et al. 2025). Cognitive agents draft a schedule or review a contract, but a person checks the result, because the model can produce confident output that is wrong (Kampelopoulos et al. 2025). Embodied agents lay brick or move earth, but a person loads the material, monitors the operation, and corrects any error (FBR 2024; Melenbrink et al. 2020). The pattern is the same regardless of the agent type. The agent extends what a human team can do, but it does not replace the team.

This pattern is most important where the stakes are highest. The reviews of LLMs in construction found almost no work that uses them for core engineering calculations such as structural analysis and load computation, and they tie this gap to the known weakness of these models in numerical precision and to the high cost of an error in safety-critical work (Kampelopoulos et al. 2025). The lesson for construction is that the right level of autonomy is not the same in every domain. A detector can run with light oversight when it only reports, but an agent whose output feeds a safety-critical decision needs close human review. Oversight is therefore not a temporary limit to be removed as the technology improves. It is a design choice that must match the stakes of the task.

5.2 System Integration

The second requirement is system integration. An agent is only as useful as the data it can reach and the systems it can act on. The adoption surveys point to fragmented data as a main reason that AI use in construction stays shallow, since project information is trapped in separate systems such as BIMs, equipment sensors, and contractor communications (Datagrid 2025). An agent that cannot see across these systems cannot give the single, current view that the work depends on.

The framework makes the integration problem easier to see, because most of the value in construction comes from agents working together rather than alone. Section 4 showed that the newest work in every domain combines agent types. A detector is paired with a language model so the system can both see a hazard and reason about it (Poudel et al. 2025). A language model is paired with a classical planner so a schedule is both sensible and exact (Pallagani et al. 2024). A robot takes its instructions from a BIM so it can place a component in the right spot (FBR 2024). Each of these is an act of integration, and each depends on the agents being able to exchange data in a common form. For transportation agencies in particular, this points to the value of open data standards and of BIMs that can serve as the shared layer through which agents connect (Samsami et al. 2021).

5.3 Trust

The third requirement is trust, which follows from the first two. A project manager will not rely on an agent whose output cannot be checked or explained. Trust has two parts in the evidence reviewed here. The first is reliability, which means the agent performs well enough, and consistently enough, that its output can be acted on. The accuracy figures reported in Section 4 are part of this, but they are not the whole of it, because a model that is accurate on a clean dataset can fail on a busy, poorly lit, cluttered real site (Ferdous and Ahsan 2022). The second part is transparency, which means a person can see why the agent achieved its result. Some of the design and contract work reviewed here is valuable precisely because it documents its decisions in a form that a person can audit (Jang and Lee 2024). For cognitive agents, the risk of confident but wrong output makes transparency especially important, because a result that cannot be traced cannot be safely trusted (Kampelopoulos et al. 2025).

These three requirements are connected. Oversight is the response to imperfect reliability. Integration is what gives an agent the data it needs to be reliable in the first place. Trust is the result when oversight and integration are both in place. A construction agent succeeds not when it is most autonomous, but when it fits into a workflow that a human team can oversee, that connects to the data the agent needs, and that produces output the team can trust.

5.4 Limitations

This study has limits that should be kept in mind. The framework is built from a structured but not exhaustive review, and it favors review articles and highly cited work, so some early-stage studies are underrepresented. The maturity reading in **Table 1** is qualitative, and reflects the weight of the evidence rather than a precise count. The adoption and improvement figures come partly from industry reports that are not peer-reviewed, and they should be read as indicative rather than exact. The framework also draws a clean line between three agent types, while real systems increasingly blur that line by combining them, as Section 4 showed. This blurring is a feature of the field rather than a flaw in the framework, but it means the three types should be read as regions on a spectrum, not as sealed boxes. The framework describes the state of the field at one point in a fast-moving area, and the cognitive agent row in particular is likely to change quickly as LLM agents mature.

6 CONCLUSIONS

This paper set out to address a conceptual gap. The term AI agent is applied to at least three very different kinds of system in construction, namely perceptual agents built on computer vision, embodied agents that act on the physical world, and cognitive agents built on LLMs. These kinds are usually studied apart, which leaves researchers and practitioners without a clear way to connect the kind of agent autonomy that is available to the construction problem that needs solving.

The paper offered a conceptual framework as a response. The framework organizes AI agents in construction along two axes, namely the agent autonomy type and the construction application domain, and it maps the three agent types onto five domains of safety monitoring, design and modeling, project and contract management, progress monitoring, and on-site automation. A structured review of recent peer-reviewed and industry literature populated the framework and rated the maturity of each pairing.

Three findings follow from the framework. First, the agent types divide the work along clear lines. Perceptual agents are mature where the task is to see and report, embodied agents are mature only where the physical task is narrow and well defined, and cognitive agents reach every domain but are mature in none. Second, the most capable systems combine agent types rather than using one alone, which means the future of the field lies in integration rather than in any single technology. Third, three requirements decide whether an agent reaches real use across every domain, namely human oversight, system integration, and trust. An agent succeeds not when it is most autonomous, but when it fits a workflow a team can oversee, connects to the data it needs, and produces output the team can trust.

The framework contributes a shared vocabulary and a practical tool. For researchers, it brings three separate literatures into one view and shows where the gaps are, since the sparse cells in the mapping mark the places where an agent type has little to offer a domain. For practitioners, and for transportation agencies in particular, it offers a way to match an agent type to a construction need, and to avoid forcing the wrong tool onto a problem.

Several directions for future work follow from this study. The first is to deepen the cognitive agent row, which is the fastest-moving and least settled part of the framework, through controlled studies of LLM agents on real construction tasks rather than on synthetic data. The second is to study the combined systems directly, since the framework shows that value comes from agents working together, yet most research still studies them apart. The third is to test the framework in practice on transportation infrastructure projects, where progress monitoring and quality control are already mature and where the integration of safety and management agents could extend that base. The fourth is to develop the open data standards and shared models that the integration requirement depends on, so that agents can connect through a common layer. Progress in these directions would help close the gap, visible throughout this paper, between strong research interest in AI agents and their still shallow use on real construction sites.

ACKNOWLEDGEMENTS

Generative AI Disclosure: In preparing this manuscript, the author used a generative AI assistant (Anthropic Claude) to assist with drafting and language editing of the text, formatting of references, and data visualizations. All AI-assisted outputs were reviewed and verified by the authors. In particular, every cited source was checked against its original publication, and citations that could not be verified were removed or corrected. The author takes full responsibility for the content, accuracy, and conclusions of this paper. No data, results, or findings were generated by AI in place of original analysis.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and framework design: RS; literature synthesis and analysis: RS; draft manuscript preparation: RS. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

Badhan, Sharmin Jahan, and Reihaneh Samsami. 2025. "Artificial Intelligence (AI) in Construction Safety: A Systematic Literature Review." Buildings 15 (22): 4084. https://doi.org/10.3390/buildings15224084.

Barbosa, F., J. Woetzel, J. Mischke, M. J. Ribeirinho, M. Sridhar, M. Parsons, N. Bertram, and S. Brown. 2017. Reinventing Construction: A Route to Higher Productivity. McKinsey Global Institute.

Built Robotics. 2025. "Autonomous Construction Equipment." Built Robotics, San Francisco, CA. https://www.builtrobotics.com.

Bureau of Labor Statistics (BLS). 2022. National Census of Fatal Occupational Injuries in 2022. U.S. Department of Labor. https://www.bls.gov/news.release.

CMiC. 2025. "AI Agents in Construction Management." CMiC, Toronto, ON. https://cmicglobal.com/resources/ai-agents-construction.

Construction Robotics. 2020. "SAM100: Semi-Automated Mason." Construction Robotics, Victor, NY. https://www.construction-robotics.com/sam100.

Datagrid. 2025. "AI in Construction: Key Statistics and Trends." https://datagrid.com/blog/ai-agent-construction-statistics.

Delhi, Venkata Santosh Kumar, R. Sankarlal, and Albert Thomas. "Detection of personal protective equipment (PPE) compliance on construction site using computer vision based deep learning techniques." Frontiers in Built Environment 6 (2020): 136. https://doi.org/10.3389/fbuil.2020.00136.

FBR (Fastbrick Robotics). 2024. "Hadrian X: Autonomous Bricklaying Robot." FBR Ltd., Perth, Australia. https://www.fbr.com.au/hadrian-x.

Ferdous, Md, and Sk Md Masudul Ahsan. "PPE detector: A YOLO-based architecture to detect personal protective equipment (PPE) for construction sites." PeerJ Computer Science 8 (2022): e999. https://doi.org/10.7717/peerj-cs.999.

He, Chuanni, Bei Yu, Min Liu, Lu Guo, Li Tian, and Jianfeng Huang. "Utilizing large language models to illustrate constraints for construction planning." Buildings 14, no. 8 (2024): 2511. https://doi.org/10.3390/buildings14082511.

He, Chuanni, Min Liu, Weilin He, Zijian Wang, and Simon Hsiang. "Addressing the information silo of construction planning constraints–a graph retrieval augmented generation approach." Engineering, Construction and Architectural Management 33, no. 7 (2026): 5565-5591. https://doi.org/10.1108/ECAM-08-2024-1066.

Jang, Suhyung, and Ghang Lee. 2024. "Interactive Design by Integrating a Large Pre-Trained Language Model and Building Information Modeling." In Computing in Civil Engineering 2023, edited by Y. Turkan, J. Louis, F. Leite, and S. Ergan, 291–299. Reston, VA: American Society of Civil Engineers. https://doi.org/10.1061/9780784485231.035.

Kampelopoulos, Dimitrios, Athina Tsanousa, Stefanos Vrochidis, and Ioannis Kompatsiaris. "A review of LLMs and their applications in the architecture, engineering and construction industry." Artificial Intelligence Review 58, no. 8 (2025): 250. https://doi.org/10.1007/s10462-025-11241-7.

Khatry, Kalyan, Saida Elmi, and Reihaneh Samsami. 2026. "Automated Concrete Bridge Deck Inspection Using Unmanned Aerial System–Collected Data: A Deep Learning Approach."

ASCE OPEN: Multidisciplinary Journal of Civil Engineering 4 (1): 04026004. https://doi.org/10.1061/AOMJAH.AOENG-0091.

Luo, Junyu, Weizhi Zhang, Ye Yuan, Yusheng Zhao, Junwei Yang, Yiyang Gu, Bohan Wu et al. "Large language model agent: A survey on methodology, applications and challenges." arXiv preprint arXiv:2503.21460 (2025).

Madsen, Andrew John.. 2019. "The SAM100: Analyzing Labor Productivity." Senior project, California Polytechnic State University.

Melenbrink, Nathan, Justin Werfel, and Achim Menges. "On-site autonomous construction robots: Towards unsupervised building." Automation in construction 119 (2020): 103312. https://doi.org/10.1016/j.autcon.2020.103312

Memon, Zubair Ahmed, Muhd Zaimi Abd Majid, and Mushairry Mustaffar. "An automatic project progress monitoring model by integrating auto CAD and digital photos." In Computing in Civil Engineering (2005), pp. 1-13. 2005. https://doi.org/10.1061/40794(179)15.

NIST (National Institute of Standards and Technology). 2025. "Champlain Towers South Investigation Nears Completion of Technical Work." News release, September 2025.

Occupational Safety and Health Administration (OSHA). 2024. Construction Focus Four: Fatal Hazards. U.S. Department of Labor. https://www.osha.gov/training/outreach/construction/focus-four.

Pallagani, Vishal, Bharath Chandra Muppasani, Kaushik Roy, Francesco Fabiano, Andrea Loreggia, Keerthiram Murugesan, Biplav Srivastava, Francesca Rossi, Lior Horesh, and Amit Sheth. "On the prospects of incorporating large language models (LLMs) in automated planning and scheduling (aps)." In Proceedings of the International Conference on Automated Planning and Scheduling, vol. 34, pp. 432-444. 2024. https://doi.org/10.1609/icaps.v34i1.31503.

Pan, Yue, and Limao Zhang. "Roles of artificial intelligence in construction engineering and management: A critical review and future trends." Automation in Construction 122 (2021): 103517. https://doi.org/10.1016/j.autcon.2020.103517.

Parascho, Stefana. "Construction robotics: From automation to collaboration." Annual Review of Control, Robotics, and Autonomous Systems 6, no. 1 (2023): 183-204. https://doi.org/10.1146/annurev-control-080122-090049.

Pokhrel, Rojal, Reihaneh Samsami, Sayda Elmi, and Colin N. Brooks. "Automated concrete bridge deck inspection using unmanned aerial system (UAS)-collected data: A machine learning (ML) approach." Eng 5, no. 3 (2024): 1937-1960. https://doi.org/10.3390/eng5030103.

Poudel, Oscar, Rayan H. Assaad, and Mohamad Awada. 2025. "Integrating Large Language Models (LLMs) with Autonomous Aerial Drone Robotics and Computer Vision for Contextual Adaptive Construction Site Safety Management and Risk Assessment." In Computing in Civil Engineering 2025. Reston, VA: American Society of Civil Engineers. https://doi.org/10.1061/9780784486443.056.

Razi, Neda, Sharmin Jahan Badhan, and Reihaneh Samsami. 2026. "Artificial Intelligence (AI) in Construction Management (CM): A Systematic Review of Models and Methods." Buildings 16 (11): 2225. https://doi.org/10.3390/buildings16112225.

Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788. 2016.

Reja, Varun Kumar, Koshy Varghese, and Quang Phuc Ha. "Computer vision-based construction progress monitoring." Automation in Construction 138 (2022): 104245. https://doi.org/10.1016/j.autcon.2022.104245.

RICS (Royal Institution of Chartered Surveyors). 2025. AI in Construction: Global Survey Report. London: RICS. https://www.rics.org/news-insights/research-and-insights/ai-in-construction_report.

RTS Labs. 2026. "AI Agents for Construction." RTS Labs, Richmond, VA. https://rtslabs.com/ai-agents-for-construction.

Samsami, Reihaneh, Amlan Mukherjee, and Colin N. Brooks. "Automated Pavement Construction Inspection Using Uncrewed Aerial Systems (UAS)—Hot Mixed Asphalt (HMA) Temperature Segregation." Drones 7, no. 7 (2023): 419. https://doi.org/10.3390/drones7070419.

Samsami, Reihaneh, Amlan Mukherjee, and Colin N. Brooks. "Mapping unmanned aerial system data onto building information modeling parameters for highway construction progress monitoring." Transportation Research Record 2676, no. 4 (2022): 669-682. https://doi.org/10.1177/03611981211064277

Samsami, Reihaneh, and Saeed Jafari Kang. "Lightweight YOLO-Based Detection of Rooftop Thermal Bridges Using Uncrewed Aerial System Thermal Imagery: Comparative Evaluation of Modern YOLO Architectures." ASCE OPEN: Multidisciplinary Journal of Civil Engineering 4, no. 1 (2026): 04026007. https://ascelibrary.org/doi/full/10.1061/AOMJAH.AOENG-0117.

Samsami, Reihaneh, and Sharmin Bahdan. "Automated Hazard Detection on Construction Sites Using Contrastive Language–Image Pre-Training (CLIP): A Zero-Shot Learning Approach with Human-Centric Evaluation." In Computing in Civil Engineering 2025: Computational and Intelligent Technologies, pp. 538-542. 2025. https://doi.org/10.1061/9780784486436.05.

Samsami, Reihaneh. 2024a. "A Systematic Review of Automated Construction Inspection and Progress Monitoring (ACIPM): Applications, Challenges, and Future Directions." CivilEng 5 (1): 265–287. https://doi.org/10.3390/civileng5010014.

Samsami, Reihaneh. 2024b. "Optimizing the Utilization of Generative Artificial Intelligence (AI) in the AEC Industry: ChatGPT Prompt Engineering and Design." CivilEng 5 (4): 971–1010. https://doi.org/10.3390/civileng5040049.

Samsami, Reihaneh. 2026. "Uncrewed Aerial System (UAS) Applications in Bridge Inspection: A Comprehensive Review of Platforms, Sensors, and Operational Effectiveness." Drones 10 (2): 144. https://doi.org/10.3390/drones10020144.

Zhang, Jiansong, and Nora M. El-Gohary. "Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking." Automation in construction 73 (2017): 45-57. https://doi.org/10.1016/j.autcon.2016.08.027.

Zhang, Jiansong, and Nora M. El-Gohary. "Semantic NLP-based information extraction from construction regulatory documents for automated compliance checking." Journal of computing in civil engineering 30, no. 2 (2016): 04015014. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000346.

Zhang, Lieping, Hao Ma, Jiancheng Huang, Cui Zhang, and Xiaolin Gao. 2025. “An Improved Lightweight Safety Helmet Detection Algorithm for YOLOv8.” Computers, Materials & Continua 83 (2): 2245–2265. https://doi.org/10.32604/cmc.2025.061519.

Zhu, Yifan, Gen Zhao, Hao Liu, Dongtao Sun, and Yunze He. "Refining bridge engineering-based construction scheme compliance review with advanced large language model integration." In Proceedings of the 2024 8th International Conference on Big Data and Internet of Things, pp. 297-305. 2024. https://doi.org/10.1145/3697355.3697404.