



# AI for construction and project delivery

Improving project information flow, documentation,  
coordination, and risk visibility.

WORKS WITH THE ASSISTANTS TEAMS ALREADY USE



EDITION 19 / EMBER & INDIGO

EXECUTIVE WORKING EDITION - 08 OCT 2026

# 19

## AT A GLANCE

Improving project information flow, documentation, coordination, and risk visibility.

Construction projects fail less often from a lack of effort than from information that arrives late, in the wrong place, or in a form nobody has time to read. This playbook shows where AI helps with RFIs, submittals, document control, site reporting and project communication, and where it does not. It covers how the tools work, who owns what, how to keep a human accountable for every contractual and safety-relevant output, and how to measure value against a baseline. It ends with a 30/60/90 plan a project executive can start on this quarter.

### WHO IT IS FOR

Project executives, project managers, document controllers, superintendents, preconstruction and operations leaders at general contractors, specialty contractors and owner-side delivery teams.

### TIME TO USE

60 minutes to read; one working session to apply

## What you will have at the end

- A shortlist of two or three project-information workflows worth automating first
- A clear line between AI drafting and human sign-off for contractual documents
- A baseline and metric set so you can show value or stop
- A 30/60/90 plan with named owners

## Contents

|           |                                                         |    |
|-----------|---------------------------------------------------------|----|
| <b>01</b> | Why project information flow is the real constraint     | 03 |
| <b>02</b> | Where AI fits in project delivery and where it does not | 04 |
| <b>03</b> | How these tools work, in plain language                 | 06 |
| <b>04</b> | Roles, ownership and a workable operating rhythm        | 08 |
| <b>05</b> | Risks, controls and human review                        | 09 |
| <b>06</b> | Measuring quality and value against a baseline          | 11 |
| <b>07</b> | An implementation guide you can follow                  | 12 |
| <b>08</b> | Common pitfalls and an illustrative scenario            | 14 |
| <b>09</b> | Questions to ask vendors and your own team              | 15 |
| <b>10</b> | A first quarter that produces evidence                  | 16 |
| <b>11</b> | Terms used in this publication                          | 17 |

## EXECUTIVE BRIEF

# Why project information flow is the real constraint

01



Most delivery risk shows up first as an information problem: a question unanswered, a drawing revision missed, a note that never reached the crew.

A typical commercial project generates thousands of documents: drawings, specifications, RFIs, submittals, meeting minutes, daily reports, change requests, photos and emails. The people who need an answer rarely have time to search all of it. A superintendent wants to know which revision of a detail governs. A project manager wants to know whether a notice deadline is approaching. A cost engineer wants to know which open items carry exposure.

AI is useful here because it reads and drafts text quickly. It can summarize a 200-page specification section, compare two revisions of a document, draft a first-pass RFI response for a reviewer, or turn a week of daily reports into a short risk summary. It is not a substitute for a licensed professional's judgment, for contract administration, or for site safety management. The value comes from giving experienced people their time back and surfacing things earlier, not from removing them from the loop.

- Start with information-heavy, review-friendly work: RFI drafting support, submittal checking, meeting minutes, daily report summaries and document search.
- Keep AI output as a draft with a named human reviewer on anything that carries contractual, safety, structural or payment consequences.
- Measure cycle time, rework and missed-item rates against a baseline you capture before launch.
- Treat your project data as an asset: clean naming, current revisions and clear ownership matter more than the choice of tool.



**SCOPE NOTE**

This playbook covers administrative and coordination work. It does not cover engineering design decisions, structural calculations or safety determinations, which remain with qualified professionals under applicable codes and contracts.

**WORKING QUESTION**

Which single information delay cost your last project the most time, and who found out last?

**WHERE IT FITS**

Where AI fits in project delivery and where it does not

02

The best candidates are repetitive reading and drafting tasks with a clear reviewer and a cheap way to check the result.

| USE CASE                                                                                  | FIT                   | HUMAN CHECK REQUIRED                                          |
|-------------------------------------------------------------------------------------------|-----------------------|---------------------------------------------------------------|
| RFI drafting support: find relevant spec and drawing references, propose a draft question | Strong                | Project engineer confirms references and wording before issue |
| Submittal review support: compare submittal data to spec requirements, flag gaps          | Strong with checklist | Reviewer signs the response; tool never approves              |
| Meeting minutes and action items from transcripts                                         | Strong                | Meeting owner corrects names, dates and commitments           |

| USE CASE                                                    | FIT                         | HUMAN CHECK REQUIRED                                                            |
|-------------------------------------------------------------|-----------------------------|---------------------------------------------------------------------------------|
| Daily report and photo log summaries for weekly risk review | Good                        | Superintendent confirms what was actually observed                              |
| Document search across drawings, specs and correspondence   | Good if documents are clean | User opens the source before relying on it                                      |
| Schedule or cost forecasting from historical data           | Mixed                       | Scheduler and estimator own assumptions; treat output as a prompt for questions |
| Structural, life-safety or code compliance determinations   | Not suitable                | Licensed professionals decide; do not use AI as the decision maker              |

The pattern is consistent. AI performs well when the answer lives in documents you already hold, when a knowledgeable person can verify it in minutes, and when a wrong answer is caught before it leaves the building. It performs poorly when the task needs site-specific observation it cannot see, a professional seal, or a judgment about risk that nobody has written down.

## Two filters before you choose

- **Reviewability:** can the person who checks the output do so faster than doing the work from scratch?
- **Consequence:** if a wrong answer slipped through, would it cause a rework, a dispute, a delay claim or an injury? The higher the consequence, the tighter the review and the narrower the tool's role.
- **Data readiness:** are current documents in one place with consistent names and revision control?
- **Volume:** does the task recur often enough that a time saving of minutes per item adds up across a project portfolio?

### COMMON MISTAKE

Starting with the most dramatic use case, such as automated claims analysis, instead of the duller one. Meeting minutes and daily report summaries build trust, data discipline and review habits that harder use cases need.

**WORKING QUESTION**

Which two tasks on your current project score high on volume and reviewability and low on consequence?

**HOW IT WORKS**

# How these tools work, in plain language

# 03

Knowing the mechanism helps you predict where a tool will fail before it does.

A large language model predicts useful text based on patterns learned from a very large body of writing. It does not know your project. To make it useful on one, a tool shows the model the right passages at the moment of the question, usually by searching your documents first and then asking the model to answer from what it found. This approach is called retrieval-augmented generation. The model then drafts an answer, ideally with references to the source pages so a reviewer can verify it.

## 01 Ingest

Documents are pulled from your common data environment or document system, split into passages and indexed. Drawings are the hard part: many are scanned or vector files where text and linework mix, so results on drawings are weaker than on specifications.

## 02 Retrieve

When someone asks a question, the tool finds the passages most likely to contain the answer. If the wrong revision or a superseded addendum is in the index, the wrong passage can be retrieved.

## 03 Draft

The model writes an answer or document using only the retrieved material, and should cite where each statement came from.

## 04 Review

A named person checks the citations, edits the draft and approves or rejects it. Their corrections become your test data.

### Failure modes to expect

- Stale sources: superseded drawings or unincorporated addenda sit in the index and get quoted as current.
- Confident gaps: the tool answers when the answer is not in the documents, or when two documents conflict, instead of saying so.
- Drawing blindness: dimensions, revision clouds and detail callouts are misread or missed entirely.
- Numerical slips: quantities, dates and unit conversions are copied incorrectly.
- Context leakage: information from one project or one party surfaces in another because permissions were not respected.

#### DESIGN RESPONSE

Require citations, restrict each assistant to one project's current documents, instruct it to say not found when sources are silent, and make the reviewer open the cited page every time during the pilot.

#### WORKING QUESTION

If your document system were indexed tomorrow, which superseded or draft files would be retrieved by mistake?

OPERATING MODEL

## Roles, ownership and a workable operating rhythm

04

Tools without owners drift. Name who owns the data, the prompts, the review and the results.

| ROLE                         | OWNS                                                    | TYPICAL PERSON                                  |
|------------------------------|---------------------------------------------------------|-------------------------------------------------|
| Executive sponsor            | Scope, budget, decision to scale or stop                | Operations or project executive                 |
| Workflow owner               | The process being improved, revision standards, metrics | Senior project manager or document control lead |
| Data steward                 | Source quality, revision control, permissions, naming   | Document controller or BIM/VDC coordinator      |
| Reviewers                    | Check and approve every output, their discipline        | Project engineers, superintendents, estimators  |
| Technology and security lead | Vendor access, data handling, logging, contract terms   | IT lead or outside advisor                      |

Run the pilot on one or two live projects with willing project teams, not across the portfolio. A weekly fifteen-minute review is enough: what was produced, what the reviewer changed, what was missed, and what to adjust. Record each reviewer correction in a simple log. After a month, that log shows which instructions need tightening, which document sources are unreliable and which use cases to drop.

### Standard work that should exist before launch

#### Confirm before the first live use

- A written statement of what the tool may draft and what only a person may issue
- A named reviewer for each document type, with a backup
- A rule that no AI draft is sent to an owner, architect, subcontractor or authority without human approval
- A single place where approved prompts and templates live
- A log of corrections and errors that someone reads weekly

**SUBCONTRACTOR REALITY**

Your subcontractors and consultants will use their own tools. Agree on expectations in kickoff meetings: who may upload project documents to external tools, and what must be disclosed in contract-relevant correspondence.



**WORKING QUESTION**

Who on your current project would own the correction log, and do they have time this month?

**RISKS AND CONTROLS**

## Risks, controls and human review

05



Construction documents create obligations. The control framework must be proportionate to the consequence of each output.



Contracts, notices and responses to information requests can create or waive rights. An AI draft that misquotes a specification, misses a notice period or commits the company to a scope item can cost far more than the time it saved. Safety documents carry a different kind of weight: a hazard analysis, a toolbox talk or an incident summary must reflect what actually happens on site and must be owned by the people responsible for it.

| RISK                                    | CONTROL                                                                | OWNER           |
|-----------------------------------------|------------------------------------------------------------------------|-----------------|
| Wrong or superseded source quoted       | Index current revisions only; show source and revision on every answer | Data steward    |
| Contractual commitment made by accident | AI drafts only; named person issues, no auto-send                      | Project manager |



| RISK                                                  | CONTROL                                                                             | OWNER           |
|-------------------------------------------------------|-------------------------------------------------------------------------------------|-----------------|
| Confidential or owner data exposed to a public tool   | Approved tools only, contract terms reviewed, access by project                     | Technology lead |
| Over-reliance and skill erosion in junior staff       | Juniors draft first, then compare AI output; periodic unaided review                | Workflow owner  |
| Safety content that sounds right but is generic       | Site-specific input required; safety manager approves                               | Safety lead     |
| Discoverable records created without a retention plan | Decide what is a record; set retention for prompts, drafts and outputs with counsel | Legal and IT    |

## Review tiers

- Tier 1, low consequence: internal summaries, meeting notes, search. Spot check weekly.
- Tier 2, medium consequence: draft RFIs, submittal comments, owner updates. Reviewer approves every item.
- Tier 3, high consequence: notices, change order positions, claims, safety plans, anything requiring a professional seal. AI may assist with research, but a qualified person authors and signs.
- Escalate any item where the tool and the reviewer disagree.

### CHECK CURRENT REQUIREMENTS

Contract language on confidentiality, data location, records and professional responsibility varies by project and jurisdiction. Have counsel review your standard contract terms and your vendor agreement. This playbook is not legal advice.

### WORKING QUESTION

Which of your outgoing documents are Tier 3 today, and does everyone who drafts them know it?

## MEASURING VALUE

# Measuring quality and value against a baseline

06

If you do not capture the before, you cannot prove the after.

Before the pilot, spend two weeks measuring how long target tasks take now and how often they go wrong. Sample real items: for example, twenty recent RFIs or ten submittal reviews. Record time to first draft, number of review cycles, and defects found later. Then run the same measures during the pilot, alongside reviewer ratings of draft quality. Compare like with like: similar project phases, similar complexity.

| METRIC              | HOW TO MEASURE                                                  | WHAT GOOD LOOKS LIKE                     |
|---------------------|-----------------------------------------------------------------|------------------------------------------|
| Time to first draft | Minutes from request to reviewable draft, sampled               | Falls without review time rising         |
| Reviewer edit rate  | Share of drafts needing substantive changes, logged by reviewer | Falls over weeks as instructions improve |
| Citation accuracy   | Share of cited passages that actually support the statement     | Near complete before widening use        |
| RFI cycle time      | Days from issue to closed response                              | Shorter, and fewer RFIs reopened         |
| Missed-item rate    | Open items found late in weekly review versus caught on time    | Fewer late discoveries                   |
| Reviewer confidence | Short monthly rating from the people reviewing                  | Stable or improving                      |

## Build a test set

- 01 Collect**  
Pull 30 to 50 real questions or documents from closed projects where the correct answer is known.
- 02 Mark**  
Have a senior person record the correct answer and the source page for each.

### 03 Run

Run the tool and score each answer as correct, partly correct, wrong or properly declined.

### 04 Set a pass mark

Agree in advance what accuracy lets you widen use. Re-run the set whenever sources, tool settings or the model change.

#### ILLUSTRATIVE ARITHMETIC

If a project engineer drafts 15 RFIs a week and a tool saves 20 minutes on each, that is 5 hours a week. If review adds 8 minutes per item, the net is about 3 hours. Use your own numbers, and count review time.

#### WORKING QUESTION

What would you need to see after 60 days to say this is worth scaling, and have you written it down?

#### IMPLEMENTATION

## An implementation guide you can follow

# 07

A small, supervised pilot on a live project beats a portfolio rollout planned in a conference room.

### 01 Choose one workflow

Pick a single workflow such as RFI drafting support or daily report summaries. Define inputs, outputs, reviewer and turnaround target.

### 02 Prepare the data

Gather current specifications and drawings, remove superseded files, confirm naming and permissions, and decide what the tool may see.

### 03 Select and configure the tool

Choose a tool that keeps data within agreed boundaries, cites sources and logs activity. Write instructions that name the document types, tone and what to do when sources are silent.



### 04 Pilot with a small team

Run for four to six weeks on one project. Reviewers log every correction. Hold the weekly fifteen-minute review.

### 05 Decide

Compare results to baseline and the pass mark. Scale, adjust or stop, and write down why.

### 06 Train and standardize

Turn what worked into a one-page standard, short training and a template library for the next project.



## Pilot readiness

- Sponsor, workflow owner, data steward and reviewers are named
- Baseline measures are captured for at least two weeks
- Document sources are current and permissions are correct
- Vendor terms on data use, retention and confidentiality are reviewed
- Staff know what they may and may not paste into any tool
- A stop rule is agreed: what result ends the pilot early

Train the people who will use and review the tool, not just the champions. Short, role-based sessions work best: document controllers on source hygiene, engineers on prompt structure and verification, superintendents on how to feed accurate field information in. Include a session on what the tool gets wrong, using examples from your own test set. People trust a tool more appropriately when they have seen it fail.



#### WORKING QUESTION

Which live project has a supportive project manager, clean documents and a repetitive information task?



## PITFALLS

# Common pitfalls and an illustrative scenario

08

Most failures are predictable. They come from data, ownership and review, not from the model.

- Indexing everything, including old drafts, and then wondering why answers conflict.
- Letting AI output go to the owner or architect without a named reviewer because it looked polished.
- Skipping the baseline, so nobody can show the value when budgets are reviewed.
- Treating drawings like text and trusting results on dimensions or details.
- Rolling out to every project at once with no correction log or owner.
- Ignoring the field: tools used only in the office miss what the crew actually saw.

## ILLUSTRATIVE SCENARIO (HYPOTHETICAL)

A mid-size contractor pilots a drafting assistant for RFIs. Early drafts quote a superseded addendum because old files remain in the index. The data steward removes them, adds a revision field to every citation, and reviewers begin checking the cited page. Edit rates fall over several weeks. The lesson: the fix was data hygiene and review discipline, not a different model.

Another frequent pattern is the unowned tool. A vendor demonstrates well, a few enthusiasts use it, and then the champion moves to another project. Within a quarter nobody reviews outputs, instructions are stale and trust has gone. Assign the workflow owner a named backup, and put the tool in your project start-up checklist so each new project sets it up deliberately or decides not to.

Finally, resist the urge to measure only speed. A faster RFI that is wrong creates a dispute later. Pair every speed metric with a quality metric such as reviewer edit rate or reopened items, so improvement in one cannot hide damage in the other.

## WORKING QUESTION

If your pilot failed in six months, what is the most likely reason, and what can you do about it now?

## BUYING AND VENDORS

# Questions to ask vendors and your own team

09

Demos show the best case. These questions reveal what happens on a bad day.

| TOPIC         | ASK                                                                                             | WHY IT MATTERS                         |
|---------------|-------------------------------------------------------------------------------------------------|----------------------------------------|
| Data handling | Where is our data stored, who can access it, is it used to train models, and how is it deleted? | Project and owner data is confidential |
| Integration   | Does it connect to our document system and respect existing permissions?                        | Avoids shadow copies and leakage       |
| Citations     | Can every answer link to the exact page and revision?                                           | Makes review fast and honest           |
| Drawings      | How does it handle scanned drawings and revision clouds? Show a real example.                   | Drawings are the weakest area          |
| Logging       | Can we see who asked what and what was produced?                                                | Supports audit and learning            |
| Exit          | How do we export our data and configuration if we leave?                                        | Prevents lock-in                       |

Ask each vendor to run your own test set on your own documents during a trial, and have your reviewers score the results blind where possible. Be cautious about claims of accuracy that cannot be reproduced on your data. Also ask what happens when the underlying model

changes: you want notice, the ability to re-test, and a rollback option.

### Before you sign

- Trial results on your documents are scored by your reviewers
- Contract terms on confidentiality, retention and breach notice are reviewed by counsel
- Pricing is understood at pilot scale and at full scale
- Support, training and named contacts are agreed
- You have a documented exit path

#### WORKING QUESTION

What would you ask a vendor to prove on your documents, not theirs?

30 / 60 / 90

A first quarter that produces evidence

10

Use the next 90 days to prove one workflow on one project, then decide with evidence.

### First 30 days: ground it

- Select one workflow and one live project, and name the sponsor, workflow owner, data steward and reviewers
- Capture two weeks of baseline timing and quality samples
- Clean the document source: current revisions only, correct permissions
- Write the tier rules for what AI may draft and who must approve

## Days 31 to 60: test it

- Run the pilot with a weekly fifteen-minute review and a correction log
- Build and score a 30 to 50 item test set from closed projects
- Adjust instructions and sources based on logged errors
- Train reviewers and field staff on verification and on what not to paste into tools



## Days 61 to 90: decide

- Compare results to baseline and the agreed pass mark
- Decide to scale, change or stop, and document the reasoning
- Turn the working setup into a one-page standard and template library
- Add the tool setup decision to the project start-up checklist

### GLOSSARY

## Terms used in this publication

11



| TERM                           | MEANING                                                                                                                   |
|--------------------------------|---------------------------------------------------------------------------------------------------------------------------|
| RFI                            | Request for information: a formal question from the contractor to the design team about unclear or conflicting documents. |
| Submittal                      | Product data, shop drawings or samples sent for review to confirm they meet the specification.                            |
| Common data environment        | The shared, controlled place where project information is stored and exchanged.                                           |
| Document control               | The discipline of keeping documents current, named consistently and distributed to the right people.                      |
| Retrieval-augmented generation | A method where the tool finds relevant passages first and then drafts an answer from them.                                |
| Citation                       | A reference from a statement back to the exact page or section that supports it.                                          |
| Hallucination                  | A confident but unsupported statement produced by a model.                                                                |
| Revision control               | Tracking which version of a drawing or document is current and which are superseded.                                      |



| TERM              | MEANING                                                                          |
|-------------------|----------------------------------------------------------------------------------|
| Test set          | A fixed group of real questions with known correct answers used to score a tool. |
| Baseline          | A measurement of current performance taken before a change, used for comparison. |
| Daily report      | A site record of work, labor, equipment, weather and issues for a given day.     |
| Human in the loop | A design where a named person reviews and approves output before it has effect.  |

### A note on use

This playbook is general guidance for administrative and coordination work. It is not engineering, safety or legal advice. Check your contracts, local codes and current requirements.

#### TAKE THE NEXT STEP

Take the free Halden assessment at [halden.company/assessments](https://halden.company/assessments), or talk to us at [halden.company/contact](https://halden.company/contact) about turning this into a plan for your organization.