EvolveData AI Solutions

Giving Frontline Staff Time Back

AI document processing in home and community care: why review, not drafting, decides whether it saves time

For executive directors, operations, clinical and IT leaders in home and community care, social care and community health

> **Who this is for.** If your coordinators re-key referrals into two systems, your nurses search several systems for a client's history before a visit, or someone has proposed an AI tool to "fix the paperwork", this paper is for you. It explains how to tell whether the tool will give time back, and how to check its work.

The argument in brief

AI can give frontline staff time back, but only if checking its work costs less than it saves. Paperwork takes a large share of the care day in community care [1][2][3], and the published studies of AI documentation tools show real but modest savings: about 10% of note-writing time in the best randomized trial [4]. A margin that small is easy to lose. Four ideas keep it.

1. **Review cost decides the return.** A tool that drafts quickly but needs close checking can save no time at all. Net Time Saved (Section 6) counts review, rework, exceptions and assurance against the drafting gain.
2. **A summary can be fluent and wrong, so check the facts before the summary reaches anyone.** A system that compares facts across documents can flag what a reader would miss, including handwritten changes that were never keyed into the system (Section 5).
3. **A handful of planted errors measures nothing.** A team that catches four of five planted errors cannot say whether its true catch rate is 40% or 95%. Telling takes closer to 50 or 100 (Section 5).
4. **Give AI autonomy in proportion to what an error costs and how likely it is to be noticed.** The Consequence–Detectability Check (Section 4) turns that into a decision for each document task.

AI is rarely the first step. Some paperwork should be retired and some systems connected before any model is used, and Section 3 says how to tell. Where AI does belong, the four ideas above decide whether it pays.

> **A summary can be fluent and wrong.** On a claims study the author worked on, OCR extracted facts and figures from documents held in several places, and a language model summarized them for a reviewer. When OCR misread a number, the summary was just as confident as when it was right. Reviewers had to go back to the source documents and read everything closely, which removed most of the time the automation was meant to save. The team fixed this by checking the facts before any summary reached a reviewer, as Section 5 describes. The same failure applies to a faxed referral or a scanned intake form: if extraction goes wrong and the summary still reads confidently, the cost lands on the reviewer.

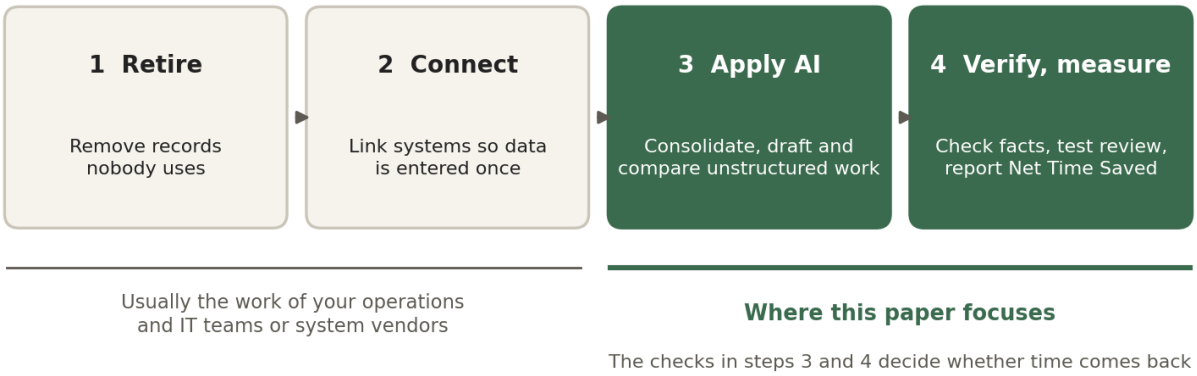


Figure 1. The order of work. The last two steps are where AI is used and where its results are checked.

1. Where frontline time goes

Information work takes a large share of the care day, and staff see much of it as redundant. In a 2019 survey of 195 Dutch community nurses, respondents estimated that they spent on average 8 hours a week on clinical documentation and a further 3.6 hours on organizational documentation, which serves care organizations, managers, policymakers and insurers more than the client [1]. Finnish home care professionals reported an average of 258 minutes a week on finding, coordinating and communicating information [2], and observation of Australian aged care case managers found that office time went mainly on communication and documentation [3]. Appendix A lists these and other sources.

Not all of it is waste, and the burden is not evenly felt. Dutch nurses regarded clinical documentation as necessary and useful, while many called organizational records cumbersome, redundant and repetitive. In the same data, time on organizational documentation was linked to feeling overloaded and time on clinical documentation was not [1]. The studies point to two causes. One is separate systems that make staff enter the same information twice [1][2]. The other is rules from employers and funders that add records [1][5]. Documentation time is also easy to underestimate: in Norwegian home care, the time needed for statutory documentation, along with driving, appeared to be underestimated in staff schedules [6].

What 10% means in practice. Dutch nurses reported a mean of 3.6 hours a week on organizational documentation (216 minutes) and a median of 2.0 hours (120 minutes) [1]. Every 10% removed, net of any review time, returns about 12 minutes a week to a typical nurse and about 22 at the mean. For a Finnish registered nurse reporting 518 minutes of information work [2], 10% is about 52 minutes a week. In each case the gain counts only after review time is subtracted, which is the subject of the rest of this paper.

2. What AI saves, and what it doesn't

The evidence on AI time savings is indirect, and the gains are modest. We are not aware of controlled studies of AI on organizational tasks such as re-entry between systems, referral intake or care coordination. The nearest evidence is clinical note-writing. In a randomized trial of 238 physicians, one AI scribe cut note-writing time by 9.5% and the other produced no significant change [4]. A study of 8,581 clinicians found about 16 fewer minutes of documentation per eight hours of scheduled patient time, with after-hours time unchanged [7]. Outside healthcare, writing tasks ran about 40% faster in a controlled experiment [8], and customer support agents became 15% more productive, mostly the least experienced ones [9]. Appendix B has the details.

The benefit depends on the product, the task and the people. In the trial, physicians rated inaccurate content as occurring close to "occasionally" for both products. In the larger study, heavier users saw bigger reductions but were a self-selected third of users. Evening charting is a case in point: in the larger study, after-hours documentation time did not change significantly [7], so these tools have not yet shown that they reduce it. None of the studies measured the time spent checking AI output either, and that gap is why this paper argues for measuring review directly. An organization's own measurement therefore matters more than any published average.

> **What the evidence can and can't tell us.** The sources show that information work is substantial, partly driven by fragmented systems and rules, and felt as a burden. They do not show that AI reduces it in community care. Most time figures are self-reported, with response rates of 38.4% in the Dutch survey and 17% in the Finnish one [1][2], and there is no standard validated measure of documentation burden [10]. The English social care figures come from a sector report commissioned by local-government bodies and covering children's social care [11]. The Consequence–Detectability Check is a planning heuristic that has not been validated as a measurement instrument, and the judgments in this paper about where AI belongs should be tested locally.

3. Where AI fits

Settle what can be removed or connected before choosing AI. Retiring a record that serves no current purpose saves more time than automating it. Where data is structured and the target system has an interface, an integration, a master-data fix or a workflow rule is usually cheaper to run and to review than a language model. Integration errors are systematic and testable, but a wrong field mapping repeats across every record until someone reconciles the systems, so every integration needs

reconciliation checks between source and target. These two steps are normally work for an organization's own operations and IT teams or its system vendors. They come first because they change what is left for AI.

AI earns its place on what remains. It has three main jobs.

- **Consolidation.** Assembling a record from several systems or from unstructured inputs such as faxed referrals, intake forms, hospital discharge summaries and free-text notes, which an integration cannot do. Accuracy has to be measured field by field on the organization's own documents, and conflicts between systems flagged for a person instead of resolved silently.
- **Drafting.** Visit notes, shift handovers, care coordination notes and case recordings follow templates, so a model can turn a checklist or short voice memo into a draft for the worker to correct. This has the nearest measured evidence, though from physician note-taking, and the share of drafts needing substantive edits shows whether it saves time.
- **Comparison.** Checking documents against a standard, such as agreements against approved terms. A missed difference leaves no trace in the output, so comparison is the hardest to review. We are not aware of independent benchmarks, so test comparison tools on your own documents with known changes planted.

Where a target system has no interface at all, screen automation or AI agents are usually proposed. They act directly on other systems and should be treated like the agents in Section 4.

4. Where AI shouldn't decide: the Consequence–Detectability Check

Two questions set how much autonomy AI gets. How costly would an error be, and how likely is it that someone notices? The higher the cost, or the lower the chance of detection, the less autonomy the system should have. We call this the Consequence–Detectability Check, or the Check. It borrows from failure mode and effects analysis, which has rated the severity and detectability of failures for decades and is standardized internationally [12], and from research on levels of automation [13]. The NIST AI Risk Management Framework and the EU AI Act also scale controls to the level of risk [14][15]. The addition here is applying these ideas to individual document tasks, with detectability as an explicit axis.

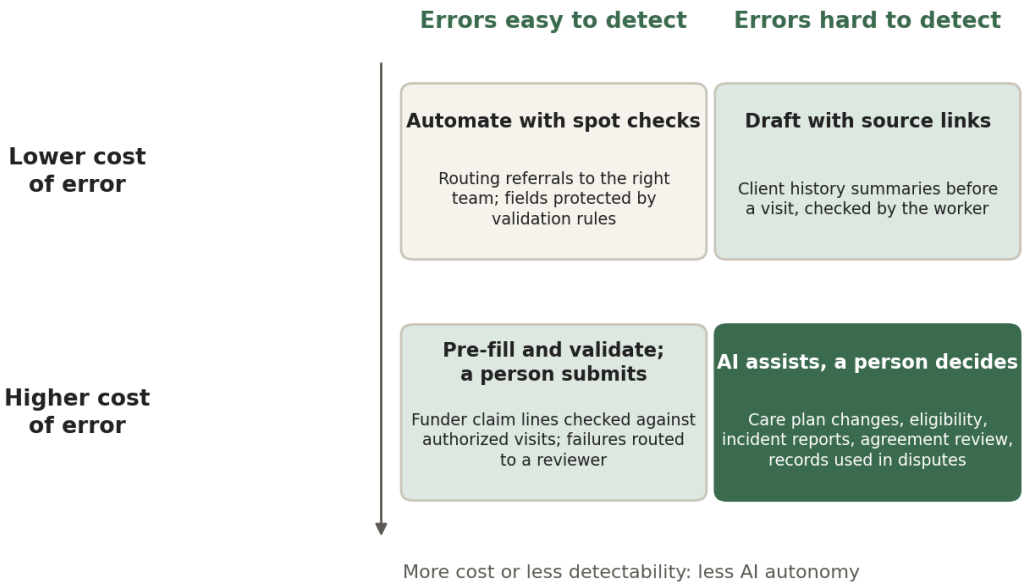


Figure 2. The Consequence–Detectability Check. Placements are examples; each organization places its own tasks.

An integration with reconciliation checks sits on the left; AI on unstructured inputs usually sits further right. Without reconciliation checks a mapping error can spread silently, and the task moves right. AI work on the left of the Check relies on a calibrated definition of "low confidence". We set the confidence threshold from labelled data so that the observed error rate among items passing automatically stays below a tolerance agreed in advance, then confirm it with a regular audit sample. Without that calibration, the task belongs in the bottom-right cell.

Agents need approval before any high-risk action. An agent that reads a referral, updates the schedule and notifies the funder removes the human pause between AI output and its consequence. Each action an agent can take, whether updating a record, sending a notification or submitting a claim, should be placed on the Check separately, and actions in the bottom row should require a person's approval before they run. Organizational work can still be high-risk: an incident report counts as organizational documentation in the Dutch study, yet it belongs in the bottom-right cell. Decisions that materially affect people, such as changing a care plan or determining eligibility, stay with an accountable person however the workflow is automated.

5. Review: the discipline that decides whether time comes back

Reviewers miss what automation misses. Every deployment described in this paper relies on a person checking the AI's work, and research on how people check automated output shows clear limits. In their widely cited review, Parasuraman and Manzey describe automation bias as producing two kinds of error when a decision aid is imperfect: missing a problem the system did not flag, and following advice that is wrong. It occurs in experts as well as novices, and training or instructions do not prevent it [16]. Omissions are the hardest case in document work, because a summary that left something out looks exactly like one that did not.

The claims team checked the facts, not the summary. In the claims study described at the start of this paper, two kinds of check ran before a summary was shown. The first tested whether related facts made sense together, for example that an end time was later than its start time. The second compared each fact across every document where it appeared and produced a consistency score. Facts below a threshold were highlighted for the reviewer. The threshold was set from labelled data: the team ran the checks across a range of documents, analyzed the results and chose the threshold from them.

Some of the best catches were not extraction errors. People sometimes wrote the actual value by hand over the printed one, and nobody keyed it back into the system. The flags exposed those disagreements between the paper and the system, which is the same duplicate-entry problem that Section 3 describes, in another form. Figures from this study are confidential, so none are given here.

Citations do not guarantee accuracy. An audit of four AI search engines found that only 51.5% of generated sentences were fully supported by their citations [17], and a preregistered study of legal research tools from LexisNexis and Thomson Reuters found hallucinations on 17% to 33% of queries, despite vendor claims that retrieval-augmented generation avoided them [18]. Both studies tested systems from 2023 and 2024, and independent faithfulness tests on operational documents such as referrals, care plans and visit notes remain scarce.

Review works when it has four properties. Each one answers a failure mode described above.

1. **Every claim or extracted value links to the passage that supports it,** so checking is fast enough to be done properly.
2. **Uncertainty shows in the draft itself.** Items below the agreed confidence threshold are highlighted, so reviewer attention goes to the riskiest content.
3. **Anything the system cannot support is marked unsupported** and routed to a person.
4. **The review step is itself tested.** Known errors, including deliberate omissions, are planted in the queue at a set rate, and the proportion caught is tracked.

A handful of planted errors measures nothing. Planted items must be marked so they can never be filed to a client record, and the planted set should mirror the errors that matter most. The test also needs enough of them. Catching four of five planted errors is consistent with a true catch rate anywhere from 38% to 96%, while catching 90 of 100 narrows it to 83% to 94% (Appendix C). A falling catch rate signals that the real error rate in approved output is higher than dashboards suggest. A random audit of approved items, weighted toward those approved fastest, provides a second check.

Scale assurance to the organization. A small community agency should pool planted errors over a quarter, report the catch rate with its interval, treat any missed planted error as a prompt to look at how review is being done, and rely mainly on a supervisor checking a random weekly sample of approved items, recorded in a simple decision log. A large organization processing high volumes, or using AI in the bottom row of the Check, can plant enough errors to track a trend and needs labelled test sets, field-level precision and recall, regression tests on every model or prompt change and drift monitoring. Either way, assurance is a real cost, and Section 6 counts it.

6. Measuring Net Time Saved

Net Time Saved counts every step between the old process and the new one, including the ongoing cost of assurance spread across the documents processed.

> **Net Time Saved = Baseline − (Generation + Review + Rework + Exceptions + Assurance)**

Review and assurance decide the result. Generation takes roughly the same time whether errors are easy or hard to spot. Review time rises sharply when reviewers cannot trust the draft and must read everything against the source, and assurance adds a fixed overhead that matters most at low volumes. Against savings of the size measured in Section 2, around a tenth of drafting time in the strongest trial, these costs can turn a gain into a loss.

Use a break-even worksheet, not a vendor estimate. Measure each term in the formula on the organization's own work, on the same document type before and after the change. The Net Time Saved worksheet that accompanies this paper sets out how to measure each one, converts the result into hours and dollars, shows the review time at which the saving disappears, and checks whether a planted-error test is large enough to mean anything.

A worked example from published figures. Suppose a drafting tool matched the better product in the UCLA trial, a 9.5% reduction in note-writing time, and was applied to the Dutch nurses' average of 8 hours a week of clinical documentation. The gross saving would be about 46 minutes per nurse per week. Review, rework, exception handling and assurance would together have to stay below 46 minutes a week for any net saving at all, and below a lower figure once licence and implementation costs are included. The trial concerned physicians and a different kind of note, so the 9.5% may not transfer; the example shows the scale of the margin, not a forecast.

Measure before and after, and decide in advance where the time goes. Record the baseline before anything changes and repeat the measurement after a few months, once review times have settled. To convert minutes into money, multiply net minutes saved per document by annual volume and the loaded cost of a staff minute, then subtract licence and implementation costs. Tell staff whether recovered time goes into client-facing work, reduced overtime or added capacity. Nurses in the Dutch focus groups were skeptical of burden-reduction programs for exactly this reason; one expected any reduction to "come back in spades some other way" [1].

> **Run it on your own numbers.** We complete this worksheet with clients as part of a Document AI Readiness Review (Section 9). To try it yourself, download the Net Time Saved worksheet at [worksheet link]. To talk it through with us, write to contact@evolvedataai.com.

7. Governance, security and Canadian rules

Standards give technical and risk teams a shared vocabulary. The NIST AI Risk Management Framework and its Generative AI Profile cover generative AI risks, including confabulated output [14].

ISO/IEC 42001 specifies requirements for an AI management system and appears increasingly in procurement questionnaires [19].

A hidden instruction in an uploaded document can steer the model. The OWASP Top 10 for LLM Applications ranks this risk, prompt injection, first [20]. It matters most when an agent can act on what it reads. Buyers should establish how personal information is handled at each step and whether the vendor uses client data to train its models, and should secure both answers in the contract.

Canadian rules depend on the organization. For health information custodians and public-sector bodies, health privacy and public-sector rules are usually the most immediate obligations; privately owned care providers may also be subject to federal or provincial private-sector privacy law. Appendix D summarizes the main instruments as of October 2026. It is a starting point for discussion with legal counsel and is not legal advice.

8. Getting started

Start with one high-volume document type that many staff handle every day. Time the current process on real work. Sort its tasks using Section 3 and place any AI task on the Check. Involve frontline staff from the first week, and agree accuracy targets and the level of assurance before testing begins. Set a decision point as well: a pilot that cannot show positive Net Time Saved after a few weeks of measured use should be redesigned or stopped. Appendix E lists questions to settle before building or buying.

9. Working with EvolveData AI Solutions

EvolveData is an applied AI company, and our work starts where the first two steps of Figure 1 end. We help organizations decide which document tasks are candidates for AI, then design, build and measure the AI and the checking around it. We don't build system integrations. Where the right answer for a task is to retire a record or connect two systems, we say so, and we specify the reconciliation checks an integration needs so that your IT team or system vendor can implement it.

The checking is part of what we deliver. AI output links to its source, uncertain items are highlighted in the draft, and anything the system cannot support is marked unsupported and routed to a reviewer. Before go-live we run a planted-error test on the client's own documents, using marked test items kept out of client records, and report the catch rate by error type, including omissions. After go-live a random sample of approved items is audited and reviewer decisions are logged, so the client can see who accepted or overrode each item. We size these steps to the organization, so that assurance does not consume the time it is meant to protect.

We report Net Time Saved, not demonstrations. We measure the baseline before anything changes and report Net Time Saved, including assurance time, at go-live and again after a few months, showing medians alongside means. Where the client requires it, deployments run inside the client's own cloud environment. We review the model provider's terms on data use and training with the client before

deployment, and we map the applicable privacy and public-sector obligations, such as those in Appendix D, with the client's legal counsel.

The Document AI Readiness Review. For organizations with high-volume paperwork, an engagement usually begins with a two-week review of one workflow. You receive:

1. A list of records that could be retired and duplicate entry that an integration could remove, for your own team or vendor to act on.
2. The remaining tasks sorted into AI-suited, AI-assisted with heavy review, and person-only, each placed on the Check.
3. A measured baseline for the document type.
4. A test plan sized to your organization, including the planted-error design.
5. The completed Net Time Saved worksheet, with assumptions stated, and a recommendation on whether a pilot is justified.

> **Next step.** If the review shows a pilot is justified, the next stage is a measured pilot on the same workflow, reporting Net Time Saved against the baseline. To book a 30-minute walkthrough of the method, or to arrange a readiness review, write to contact@evolvedataai.com.

Frequently asked questions

What is AI document processing?

It is the use of AI models, including large language models, to read, classify, extract, summarize, compare and draft documents such as visit notes, referrals, care plans, case recordings, claims and agreements. In well-designed deployments a person reviews and signs off the result.

Do we need to connect our systems before using AI?

Not always, but it is worth checking first. Where data is structured and the target system has an interface, an integration is usually cheaper to run and review than a language model. AI is most useful for unstructured inputs, such as faxed referrals and free-text notes, and for consolidating records from systems that cannot be connected.

Does AI document processing save frontline staff time?

Published studies, mostly of clinical note-writing, show real but modest savings that vary with the product and with how consistently staff use it [4][7]. Net savings depend on review, rework and assurance time, so they have to be measured on the organization's own work.

Is this approach too heavy for a small organization?

The principles are the same at any size, but the effort should scale. A small organization can rely mainly on a weekly random sample check and a simple decision log, pool planted errors over a quarter instead of reading meaning into a monthly handful, and count all of that time in its Net Time Saved calculation.

Why is human review so important?

Automation bias leads reviewers, experts included, to miss problems a system did not flag and to accept wrong advice [16]. Review therefore has to be designed and tested, with source links, visible uncertainty and planted-error checks.

What privacy rules apply to AI documentation tools in Ontario?

For health information custodians, PHIPA applies, and the Information and Privacy Commissioner of Ontario has published guidance specific to AI scribes [21]. Public sector entities may also be subject to the Enhancing Digital Security and Trust Act, 2024 [22], and commercial organizations to PIPEDA [23]. Confirm obligations with legal counsel.

About the author

Atri Mahapatra is Founder & CEO of EvolveData AI Solutions. Over 10 years he has taken AI from strategy to production in operationally complex organizations, including Fortune 500 companies, across manufacturing, consumer goods, travel technology and optimization consulting. He has led the build of an LLM-based document agent for claim validation that shortened the validation cycle and cut invalid deductions, set up data science functions from the ground up, and put data and AI governance in place so that models stay accountable after deployment. He began his career as a metallurgical engineer leading a 50-person team on a steel plant floor. That frontline experience, together with a grounding in operations research, is why he insists on measured outcomes over demonstrations, and it is the discipline EvolveData AI Solutions now brings to document-heavy work in home and community care. He has served as an industry expert on the steering committee of the Toronto Machine Learning Society conference.

About this paper

Research, structuring and editing were assisted by AI tools. The author is responsible for the content and conclusions. EvolveData AI Solutions offers services in this area, and readers should weigh the paper with that in mind. Peer-reviewed studies, standards and legislation are cited to the original publication or official page. Sources that are not peer-reviewed, such as sector reports, are identified as such in the text. Corrections are welcome at contact@evolvedataai.com. This paper is informational and is not legal advice.

Appendix A. Evidence on information work

Source	Method	Finding	Main caveat
Netherlands [1]	Survey of 195 community nurses, 2019	About 8 hours a week on clinical documentation and 3.6 hours on organizational documentation	Self-reported time; 38.4% response rate
Finland [2]	Survey of 625 professionals in four home care organizations, 2024	Average of 258 minutes a week on retrieving, coordinating and communicating information; registered nurses 518 minutes	Self-reported time; 17% response rate
Australia [3]	Direct observation of community aged care case managers	Office time went mainly on communication (43.7%) and documentation (33.3%)	Case managers only; percentages are shares of office time
Systematic review [24]	Review of studies of time spent with electronic health records	Nurses spent about 22% of the workday with the record; documentation was 19% to 35% of nurse activities in paper and electronic chart studies	Pooled across settings; not specific to community care
England [11]	Time logs and observation, reported by a sector body	County council social workers spent about half their time on case recording, forms, meetings and IT tasks	Not peer-reviewed; commissioned by local-government bodies; covers children's social care

Table A1. Evidence on information work in community care and related settings.

Appendix B. Evidence on AI and time savings

Study	Design	Result	Main caveat
Lukac et al. 2025, *NEJM AI* [4]	Randomized trial: 238 outpatient physicians in 14 specialties, two AI scribes against usual care, two months	One product cut note-writing time by 9.5%. The other's 1.7% reduction was not statistically significant.	Clinical notes, not organizational tasks. Inaccuracies were rated close to "occasionally".
Rotenstein et al. 2026, *JAMA* [7]	Observational study: 8,581 ambulatory clinicians across five US academic health systems	About 16 fewer minutes of documentation per eight hours of scheduled patient time. After-hours time did not change significantly.	Heavier users saw larger reductions, but they were a self-selected third of users.
Noy & Zhang 2023, *Science* [8]	Controlled experiment: 453 professionals on short writing tasks	Tasks completed about 40% faster.	Self-contained tasks, and a first-generation tool now several generations old.
Brynjolfsson et al. 2025, *QJE* [9]	Staggered rollout of a live-chat assistant to 5,172 support agents	Productivity up 15% on average, with gains concentrated among less experienced agents.	Customer support, not documentation.

Table A2. Published evidence on AI and time savings.

Appendix C. Why sample size matters for planted-error tests

The table shows the range within which the true catch rate plausibly lies, at 95% confidence, for different numbers of planted errors. Intervals are Wilson score intervals.

Planted errors caught	Observed catch rate	Plausible true catch rate (95% interval)
4 of 5	80%	38% to 96%
18 of 20	90%	70% to 97%
45 of 50	90%	79% to 96%
90 of 100	90%	83% to 94%

Table A3. Plausible true catch rates by number of planted errors.

Appendix D. Canadian instruments relevant to document AI

Applicability depends on the organization and use case; confirm with counsel. Status as checked in October 2026.

Instrument	Applies to	Relevance to document AI	Status (checked Oct. 2026)
Personal Health Information Protection Act, 2004 (Ontario)	Health information custodians and their agents	Governs collection, use and disclosure of personal health information, including through AI tools	In force
IPC Ontario, *AI Scribes: Key Considerations for the Health Sector* [21]	Custodians that develop, procure or use AI scribes	Non-binding guidance on governance, risk management, privacy impact assessments, security, accuracy, transparency and human oversight	Published 28 Jan. 2026
Enhancing Digital Security and Trust Act, 2024 (Ontario) [22]	Ontario public sector entities; AI requirements apply as prescribed by regulation	Enables requirements for accountability, risk management and technical standards in public-sector AI use	In force; detail set by regulation
PIPEDA (federal) [23]	Private-sector organizations in commercial activities, subject to provincial laws	Consent, purpose limitation and safeguards for personal information	In force
Bill C-36, Protecting Privacy and Consumer Data Act (federal) [25]	Would replace Part 1 of PIPEDA	Proposed modernization of federal private-sector privacy law	Introduced 15 June 2026; at second reading

Instrument	Applies to	Relevance to document AI	Status (checked Oct. 2026)
Act respecting the protection of personal information in the private sector (Quebec), s. 12.1 [26]	Organizations subject to the Quebec act	Individuals must be informed of decisions based exclusively on automated processing of their personal information	In force

Table A4. Canadian instruments relevant to document AI.

Appendix E. Questions to settle before building or buying

Question	A good answer includes
Which task will this speed up, and how long does it take today?	A named document type and a baseline measured on real work, including searching and re-keying
Does every output link to its source?	Links to the exact passage, page or field, demonstrated on your documents
How was accuracy measured, and on whose documents?	Field-level precision and recall on a labelled sample of your own documents, with targets agreed before testing and results reported by field type
How does it handle faxes, scans and handwriting?	Separate accuracy figures for poor-quality inputs, and a clear route to a person when confidence is low
Who reviews, and how will we know review works?	Named reviewers, measured review time, planted-error tests and an audit sample sized to the organization
Where is data processed, and is it used for training?	Processing location, deployment inside your cloud environment where needed, and a contractual answer on training use
How will we detect accuracy slipping?	Scheduled re-testing, regression tests on model or prompt changes, and drift alerts where volumes justify them

Table A5. Questions to settle before building or buying.

References

[1] De Groot, K., De Veer, A. J. E., Munster, A. M., Francke, A. L., & Paans, W. (2022). Nursing documentation and its relationship with perceived nursing workload: A mixed-methods study among community nurses. *BMC Nursing*, 21, 34. https://doi.org/10.1186/s12912-022-00811-7

[2] Sylgren, I., Jämsen, E., & Torkki, P. (2026). Time well spent? Time use for health information management across home care roles: A cross-sectional survey study in Finland. *International Journal of Integrated Care*, 26(3), 17. https://doi.org/10.5334/ijic.9881

[3] Prgomet, M., Walter, S., Jorgensen, M., Georgiou, A., & Westbrook, J. (2020). Understanding the work of case managers in Australian community aged care: A longitudinal time and motion study. *Australian Health Review*, 44(6), 853–861. https://doi.org/10.1071/AH20078

[4] Lukac, P. J., Turner, W., Vangala, S., Chin, A. T., Khalili, J., Shih, Y.-C. T., Sarkisian, C., Cheng, E. M., & Mafi, J. N. (2025). Ambient AI scribes in clinical practice: A randomized trial. *NEJM AI*, 2(12). https://doi.org/10.1056/AIoa2501000

[5] Ethier, A., Dubois, M.-F., Savaria, V., & Carrier, A. (2024). Tensions experienced by case managers working in home care for older adults in Quebec: First level analysis of an institutional ethnography. *BMC Health Services Research*, 24. https://doi.org/10.1186/s12913-024-10709-6

[6] Holm, S. G., & Angelsen, R. O. (2014). A descriptive retrospective study of time consumption in home care services: How do employees use their working time? *BMC Health Services Research*, 14, 439. https://doi.org/10.1186/1472-6963-14-439

[7] Rotenstein, L. S., Holmgren, A. J., Thombley, R., Sriram, A., et al. (2026). Changes in clinician time expenditure and visit quantity with adoption of artificial intelligence–powered scribes: A multisite study. *JAMA*. Published online April 1, 2026. https://doi.org/10.1001/jama.2026.2253

[8] Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187–192. https://doi.org/10.1126/science.adh2586

[9] Brynjolfsson, E., Li, D., & Raymond, L. R. (2025). Generative AI at work. *The Quarterly Journal of Economics*, 140(2), 889–942. https://doi.org/10.1093/qje/qjae044

[10] Moy, A. J., Schwartz, J. M., Chen, R., Sadri, S., Lucas, E., Cato, K. D., & Rossetti, S. C. (2021). Measurement of clinical documentation burden among physicians and nurses using electronic health records: A scoping review. *Journal of the American Medical Informatics Association*, 28(5), 998–1008. https://doi.org/10.1093/jamia/ocaa325

[11] Newton, County Councils Network, & Association of County Chief Executives. (2022, March 9). *The Future of Children's Social Care* [sector report]. https://www.countycouncilsnetwork.org.uk/wp-content/uploads/Newton-The-Future-of-Childrens-Social-Care.pdf

[12] International Electrotechnical Commission. (2018). *IEC 60812:2018 Failure modes and effects analysis (FMEA and FMECA)*.

[13] Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 30(3), 286–297. https://doi.org/10.1109/3468.844354

[14] National Institute of Standards and Technology. (2023, 2024). *AI Risk Management Framework (NIST AI 100-1)* and *Generative Artificial Intelligence Profile (NIST AI 600-1)*. https://doi.org/10.6028/NIST.AI.100-1; https://doi.org/10.6028/NIST.AI.600-1

[15] European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). https://eur-lex.europa.eu/eli/reg/2024/1689/oj

[16] Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3), 381–410. https://doi.org/10.1177/0018720810376055. Open repository copy: https://depositonce.tu-berlin.de/handle/11303/8923

[17] Liu, N. F., Zhang, T., & Liang, P. (2023). Evaluating verifiability in generative search engines. *Findings of the Association for Computational Linguistics: EMNLP 2023*. https://aclanthology.org/2023.findings-emnlp.467/

[18] Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C. D., & Ho, D. E. (2025). Hallucination-free? Assessing the reliability of leading AI legal research tools. *Journal of Empirical Legal Studies*, 22(2), 216–242. https://doi.org/10.1111/jels.12413

[19] International Organization for Standardization. (2023). *ISO/IEC 42001:2023 Information technology – Artificial intelligence – Management system*.

[20] OWASP Foundation. (2025). *OWASP Top 10 for LLM Applications 2025*, LLM01:2025 Prompt Injection. https://genai.owasp.org

[21] Information and Privacy Commissioner of Ontario. (2026, January 28). *AI Scribes: Key Considerations for the Health Sector*. https://www.ipc.on.ca/en/resources/ai-scribes-key-considerations-health-sector

[22] Government of Ontario. *Strengthening Cyber Security and Building Trust in the Public Sector Act, 2024*, S.O. 2024, c. 24, Sched. 1 (Enhancing Digital Security and Trust Act, 2024). https://www.ontario.ca/laws/statute/s24024

[23] Government of Canada. *Personal Information Protection and Electronic Documents Act*, S.C. 2000, c. 5. https://laws-lois.justice.gc.ca/eng/acts/p-8.6/

[24] Pinevich, Y., Clark, K. J., Harrison, A. M., Pickering, B. W., & Herasevich, V. (2021). Interaction time with electronic health records: A systematic review. *Applied Clinical Informatics*, 12(4), 788–799. https://doi.org/10.1055/s-0041-1733909

[25] Parliament of Canada. Bill C-36, An Act to enact the Protecting Privacy and Consumer Data Act (45th Parliament, 1st session), LEGISinfo status page. https://www.parl.ca/LegisInfo/en/bill/45-1/C-36

[26] Quebec. *Act respecting the protection of personal information in the private sector*, CQLR c. P-39.1, s. 12.1 (as amended by S.Q. 2021, c. 25).