

BELLHAVEN / DECISION BRIEF

How Bellhaven Should Adopt AI to Improve Productivity and Profitability

- FINDING**

Pilot AI for sorting and matching bookkeeping transactions. Test tax document extraction and review assistance on historical files, and compare their projected financial return with the return from adding a seasonal tax reviewer. Introduce any new live tax process after April 15.
- WHAT COULD WEAKEN THIS FINDING**

Only eight of the 20 planned bookkeeping clients have signed. Bellhaven has not yet measured how much time the tools save, how much accountant review they require or whether those savings remove the need for planned staff positions.

Bellhaven is a 64 person accounting firm with \$9.6 million in annual revenue and \$1.3 million in operating profit after partner salaries. It seeks \$250,000 in additional annual operating profit within 24 months, attributable to AI after all AI costs. First-year additional cash is capped at \$120,000.

ANNUAL PROFIT GOAL	FIRST-YEAR CASH CEILING	INITIAL DECISION
+\$250,000	\$120,000	Within 45 days

RECOMMENDATION	WHY IT MERITS ACTION	CONDITION
Test AI for sorting and matching bookkeeping transactions	Saved staff hours could cover the work for 20 additional clients and remove the need for planned staff positions.	Measure hours saved after review and corrections, and confirm how many clients sign.
Provide staff with an approved AI service	Four survey respondents reported client information in personal AI tools.	Approve terms, access, permitted uses and training.
Test tax extraction and accountant review together	Extraction may save preparation time while adding senior verification work.	Use historical files first; compare accepted completions and total effort.
Evaluate a seasonal reviewer	A \$49,400 alternative could address a review shortage without AI.	Confirm demand, preparation capacity and the sufficiency of 520 review hours.

BASIS OF THIS WORKED EXAMPLE

Bellhaven and its operating inputs are case assumptions from the supplied dossier. Pilot results below are illustrative, not observed results. External research informs the tests; it does not substitute for firm evidence.

What Bellhaven Must Confirm Before Expecting the Projected Profit

Do not promise the \$250,000 target from the supplied tools. The strongest original recurring scenario reaches \$220,500 under favorable assumptions using an approved office feature with no incremental license fee. That leaves a \$29,500 gap, before any deterioration in prices or performance.

IMPLEMENTATION / FIRST 45 DAYS

Decide Within 45 Days Whether Bellhaven Should Continue the Limited AI Pilot

Bellhaven must make an initial decision within 45 days. The schedule below identifies what can reasonably be established during that period and which decisions require a longer pilot.

PERIOD	WORK TO COMPLETE	DECISION EVIDENCE
Days 1 to 10	Assign owners; approve data terms; audit existing office features; confirm hiring plans and baseline workload.	Approved scope, information rules, baseline and a reconciled cost model.
Days 11 to 20	Configure restricted access; train pilot staff; score historical bookkeeping and tax files; rehearse fallback.	Quality results, review burden, failure cases and readiness for a limited live pilot.
Days 21 to 35	Begin the authorized bookkeeping comparison; continue historical intake and review tests; price the reviewer alternative.	Early complete-workflow results, actual usage costs and a credible alternative.
Days 36 to 45	Review evidence with operations, finance and the responsible partner.	Stop, repair or continue the bounded pilot; specify what remains to be learned.

What the 45-Day Decision Can Establish

It can approve a controlled next stage, reject an unsafe approach or identify the information needed to decide. It cannot honestly establish a full-year profit effect from a few weeks of observations. Keep annualized estimates separate from realized results.

Preserve Bellhaven’s Staffing, Review and Busy-Season Requirements

No layoffs are authorized for 12 months. Partner project time is limited to six hours weekly. Client-facing deliverables require credentialed review. Counsel must approve terms before client information is used. Do not introduce a new live tax workflow between January 15 and April 15.

Assign Specific Project Responsibilities Within the Partner’s Six Weekly Hours

A proposed weekly allocation is two hours for decisions and exceptions, two for quality and risk review, one for the financial evidence and one for staff issues. Operational managers must own routine data collection and configuration work. Confirm this allocation against actual demand rather than assuming unlimited partner oversight.

THE 45-DAY DECISION PRECEDES COMPLETION OF THE PLANNED PILOT

The bookkeeping cash plan assumes two months at 90 clients, then two at 180. The first 45-day review occurs before that sequence can finish. A later rollout requires the remaining evidence and a separate approval.

EVIDENCE / INITIAL ASSUMPTIONS

Identify What Bellhaven Must Establish Before Expecting AI to Increase Profit

Bellhaven’s general expectation that AI will increase productivity must be translated into questions about specific work and financial results. Each question below identifies information needed before management can rely on the projected benefit.

INITIAL EXPECTATION	SHARPER QUESTION	EVIDENCE NEEDED
AI will make everyone more productive	Which tasks improve after checking and corrections, and for which staff?	Accepted output and total time by task and experience.
Saved hours will increase profit	Will a hire or contractor invoice actually disappear, or will extra work be sold and collected?	Assignments of saved staff hours, the staffing plan without AI, and financial records.
Tax extraction is the obvious first use	Does additional senior verification consume the benefit or worsen the review queue?	Contractor hours avoided and displaced senior contribution.
An included office AI feature is free automation	Does it perform the required bookkeeping workflow under approved terms?	Feature inspection, permissions, test results and total operating cost.
A higher agent benchmark means readiness	Can the system complete Bellhaven’s work safely with its tools and information?	Repeated local end-to-end tests and a restricted pilot.
High adoption proves success	Does use improve quality, service or economics?	Eligible-work adoption paired with outcomes.

Investigate the Financial and Operational Questions That Could Change the Decision

Confirm which combinations of tools and review staff fit the budget. Test how client-fee changes affect projected profit, inspect functions included in existing software, identify the work staff would perform during saved hours and measure the complete cost of agent-assisted tasks. Review assistance also merits testing because it could change the need for senior-accountant time.

Keep Untested Tool Capabilities and Costs Separate From Supplied Case Assumptions

The case supplies a paid bookkeeping tool; it does not establish an included alternative that performs the same work. Additional insurance cost is unknown until the firm obtains applicable terms or a quotation. Review capacity must be measured by service and period. A combined AI and staffing program can be evaluated with a suitable comparison, and unresolved professional guidance should not be converted into an assumed universal fee reduction.

FOCUS THE INVESTIGATION ON RESULTS THAT COULD CHANGE BELLHAVEN’S DECISION

A useful finding identifies what Bellhaven should do differently, which assumption must be confirmed or how much improvement is needed to justify spending. Prioritize questions whose answers could change the proposed implementation.

EVIDENCE / CONFIDENCE AND DISSENT

Assess Which Bellhaven Recommendations Are Supported and Which Depend on Untested Assumptions

The figures supplied with the Bellhaven case describe its assumed operating situation. The effects of the proposed AI tools remain untested. The table identifies which recommendations depend on those performance assumptions.

CLAIM	STATUS	WHAT COULD OVERTURN IT
Bookkeeping has a plausible route to avoided hiring	Supported by the case workload and staffing plan.	Demand fails to convert, capacity is unusable or the hire was not actually needed.
Two hires can be avoided	Conditional on sufficient net capacity across roles and months.	Net sorting-time improvement below 16.9% at 200 clients, or a local staffing bottleneck.
Tax intake can reduce contractor spending	Conditional; vendor performance is an upper claim.	Savings fall on salaried staff or verification displaces valuable senior work.
A reviewer is a strong comparison	Plausible from declined demand and a review queue.	Demand, preparer capacity or the 520-hour estimate proves inadequate.
The supplied combination of AI uses reaches the annual profit target	Not supported by the best supplied recurring scenario.	A verified additional benefit or lower supplier charges could reduce the shortfall.

Why Bellhaven Might Prioritize Scheduling, Document Collection and Review Staffing

A partner could reasonably argue that Bellhaven should first improve scheduling, client-document collection and the availability of qualified reviewers. Of 60 sampled delayed files, 34 awaited review and 19 awaited client documents. Compare those changes with the proposed AI applications. The sample alone does not establish which cause of delay matters most across the full year.

What Would Change the Recommendation

If bookkeeping cannot avoid an actual staffing commitment, reduce or stop that deployment unless another measured benefit covers its cost. If review assistance safely releases scarce senior capacity, re-evaluate the tax route. If the reviewer alternative produces superior contribution with less disruption, give it priority without labeling its gain AI-generated.

Record the strongest opposing recommendation, the evidence behind it and the next observation that would resolve the disagreement.

D1 / DEFINE THE DECISION

Define Bellhaven’s Profit Target and the Requirements for a Successful AI Implementation

The decision is how to improve staff productivity and add \$250,000 in annual operating profit within 24 months, after AI costs, while protecting confidential information, preserving accurate work, staying within \$120,000 of additional first-year cash spending and avoiding operational disruption.

REQUIREMENT	OPERATIONAL DEFINITION
Productivity	More work that meets the required quality standard or fewer total labor hours for comparable work, including review and rework.
Profit	Additional collected contribution plus actual avoided costs, less all incremental program costs and lost contribution.
Confidentiality	Client information remains within approved services, identities, permissions and contractual terms.
Accuracy	Credentialed review and predefined quality tests; critical failures trigger suspension.
Continuity	A usable fallback, staged rollout and no new live tax changes during the restricted period.
Adoption	Staff can perform eligible work safely through the approved route and explain how they verify results.

Determine Whether the Proposed Uses Can Actually Produce the Targeted Profit

The \$250,000 target does not establish that the available AI applications can produce that amount. The analysis may support a smaller positive return, revised supplier or customer terms, or rejection of an application. Calculate the financial result from supported assumptions before deciding whether the target is achievable.

Specify Which Spending and Staffing Decisions the Pilot Must Inform

The firm needs to decide which tasks to test, whether the results justify a second stage and whether saved hours change planned spending on staff. It must also decide which tax-processing changes can be introduced after the busy-season restriction and whether the combined applications produce enough annual benefit to justify their continuing costs.

Record the accountable partner, operational lead, finance reviewer and specific decision due at the first checkpoint.

D2 / ECONOMICS

Examine Bellhaven’s Revenue, Costs and Staff Workload Before Estimating AI Savings

The tables show the revenue and operating costs supplied for Bellhaven. These figures establish the starting position for the analysis. They do not show which costs AI can reduce; that requires examining the relevant work and spending decisions.

SERVICE	VOLUME AND PRICE	ANNUAL REVENUE
Individual tax	2,600 returns at \$650	\$1,690,000
Business tax	420 returns at \$3,200	\$1,344,000
Planning and advisory	Hourly work	\$1,766,000
Monthly bookkeeping	180 clients at \$1,435 monthly	\$3,099,600
Audit and review	Hourly work	\$1,700,400
Revenue		\$9,600,000

OPERATING COST	ANNUAL AMOUNT
Staff salaries	\$5,332,000
Partner salaries	\$1,560,000
Contractors: 3,000 hours at \$70	\$210,000
Software	\$310,000
Occupancy	\$520,000
Insurance and other	\$368,000
Operating costs	\$8,300,000
Operating profit	\$1,300,000

Explain How Fixed Fees and Hourly Billing Change the Value of Time Savings

Individual tax and bookkeeping have fixed prices in the case. Faster delivery can improve contribution if prices hold and costs change. Hourly business tax, advisory and audit work require a separate check: fewer billable hours can reduce revenue unless capacity is redeployed or the pricing arrangement changes.

STAFFING CONTEXT

The 64 people comprise six partners, ten managers, 22 senior or staff accountants, 14 juniors with under two years of experience, and 12 administrative or client-service staff. Savings must be mapped to the people qualified to perform the next task.

D2 / CAUSES OF DELAYED WORK

Determine Whether Tax Returns Are Delayed by Preparation, Accountant Review or Missing Client Documents

The case includes a week-nine sample of 60 delayed tax files. The table shows the recorded reasons for delay. Use the sample to choose what to investigate, while preserving the distinction between one week’s waiting work and the firm’s annual workload.

SAMPLE OF 60 DELAYED FILES	COUNT	SHARE
Awaiting review	34	56.7%
Awaiting client documents	19	31.7%
Other reasons	7	11.7%
Sample	60	100%

Shares are rounded and sum to 100.1% before the total is rounded. This is the supplied week-nine sample, not an estimate of the full year.

Why Accountant Review and Missing Client Documents Deserve Investigation

Review and missing documents deserve direct attention. Producing more prepared files may simply increase waiting work. The case also reports 395 extensions, but those cannot all be attributed to reviewer capacity without tracing their causes.

Why One Sample of Delayed Returns Cannot Establish the Full-Year Financial Effect

A queue is a stock of waiting work at a point in time. It does not by itself reveal annual lost profit, the service rate at each stage or the effect of adding one reviewer. Track arrivals, completions, waiting time, age and reasons for delay across comparable weeks.

CANDIDATE INTERVENTION	EXPECTED CHANGE TO MEASURE
Better document collection	Fewer files enter preparation incomplete; fewer follow-up cycles.
Seasonal reviewer	More qualified review capacity at the constrained time.
Review assistance	Less senior effort per accepted file without weaker quality.
Tax extraction	Less preparation effort, with additional verification measured.

THE CAUSE OF DELAY MAY DIFFER BY SERVICE AND TIME OF YEAR

Insufficient accountant review time may delay tax returns during busy season even when reviewers have spare time later in the year. Bookkeeping may require different staff or skills. Measure which step limits completion for each service and period before choosing where to introduce AI.

D3 / OPTIONS

Compare AI Applications With Staffing and Process Changes That Could Improve Bellhaven’s Results

The table compares proposed uses of AI with staffing and process changes that could address the same business goals. The initial treatment identifies whether Bellhaven should test, investigate, defer or reject each approach at this stage.

OPTION	INITIAL TREATMENT	REASON
Bookkeeping sorting assistance	Controlled pilot	A measurable route to avoiding planned hires.
Tax document extraction	Historical test first	Potential contractor saving with senior verification risk.
Review assistance	Test with intake and alone	Could reduce the senior-accountant time required per completed return.
Approved staff AI service	Establish governed access	Addresses existing informal use; value still needs measurement.
Included office AI features	Audit and compare	Potential license savings; functions and terms unverified.
Document collection improvements	Compare or combine explicitly	Could reduce a major source of waiting work.
Seasonal credentialed reviewer	Obtain a concrete proposal	A meaningful \$49,400 non-AI alternative.
Scheduling and work allocation	Test operational changes	Could improve utilization without AI.
Client-service drafting	Low-risk scoped test	Useful only after checking and response handling.
Internal knowledge retrieval	Permission-aware test	Potential faster answers; outdated sources and access matter.
Advisory research support	Restrict to sourced assistance	Hourly revenue and professional judgment complicate value.
Autonomous ledger posting	Defer broad authority	Higher consequence; require restricted tests and approvals.
Autonomous tax filing	Do not authorize in this plan	Professional review and irreversible actions remain critical.
Firm-wide immediate rollout	Reject	Evidence and controls do not justify the commitment.

CHOOSE AN APPROACH THAT FITS THE TASK AND CAN PRODUCE A MEASURABLE BUSINESS BENEFIT

Choose the smallest intervention that can resolve a valuable business problem. Agentic implementation belongs in the comparison when tool use and multi-step execution are useful; it is not a mandatory architecture for every task.

D3 / BOOKKEEPING THRESHOLDS

Calculate How Much Bookkeeping Time AI Must Save to Avoid Adding One or Two Staff Positions

Each bookkeeping client requires 11 hours monthly: 6.5 for sorting and matching, 2.5 for corrections and review, and two for questions and reporting. The existing 180 clients are divided among four teams of 45 with no spare capacity. The planned 20 additional clients would otherwise require two positions costing \$98,000 each and providing 1,400 client hours each annually.

CALCULATION	RESULT
Additional work: $20 \times 11 \times 12$	2,640 hours annually
Sorting at 200 clients: $200 \times 6.5 \times 12$	15,600 hours annually
Save enough to need only one new position: $(2,640 - 1,400) \div 15,600$	7.95% net sorting-time improvement
Save enough to need neither position: $2,640 \div 15,600$	16.92% net sorting-time improvement
At 20%: $15,600 \times 20\%$	3,120 hours released
Capacity remaining after 20 new clients: $3,120 - 2,640$	480 hours annually

Determine Whether the Saved Hours Remove the Need for a Whole Staff Position

At ten-percent improvement, the model avoids one \$98,000 position. At 20%, it avoids two. A 15% improvement still leaves enough work to require one position under the supplied assumptions. Do not interpolate a fractional salary saving as if employment cost changed smoothly.

Confirm Which Teams Save Time and When They Can Use Those Hours for Additional Clients

The aggregate calculation is necessary but insufficient. Each team must have the right skills and availability when client work arrives. Review, correction and customer-contact time cannot be silently removed. Include the burden of AI exceptions in the net saving.

CONFIRM HOW MANY ADDITIONAL BOOKKEEPING CLIENTS WILL SIGN

Only eight new clients are signed. At 188 clients, eight add 1,056 annual hours and the sorting base is 14,664 hours. Avoiding a genuinely required 1,400-hour position needs about 7.2% net sorting improvement, but the case must first establish that this hire would otherwise occur.

D3 / AGENTIC BOOKKEEPING

Compare an AI Bookkeeping Agent With Software That Follows a Predetermined Sequence of Steps

The proposed bookkeeping agent would choose which permitted software steps to perform. A simpler alternative would follow a sequence specified in advance. The table shows the permitted actions and controls for the agent test; compare its total results and costs with those of the simpler alternative.

STEP	PERMITTED BEHAVIOR	CONTROL OR EVIDENCE
Receive records	Read only the authorized client’s approved inputs.	Software enforces access restrictions for the firm’s account and each client’s records.
Classify and match	Suggest categories and matches with references to source transactions.	Use approved rules; identify missing or conflicting evidence.
Handle exceptions	Send uncertain, unusual or material items to a qualified reviewer.	Thresholds and escalation categories set before the test.
Prepare proposed changes	Create a reviewable batch; do not post automatically in the initial pilot.	Show before-and-after values and supporting evidence.
Approve and record	An authorized person approves permitted changes.	Log identity, action, source, time and outcome.
Reconcile and recover	Check the resulting ledger state and resolve partial failures.	Safe retries, duplicate prevention and a tested rollback procedure.

Compare the Current Process, Predetermined Software Steps and an AI Agent

Compare the current workflow, a structured assisted workflow and a restricted agent where feasible. An agent chooses steps and tool calls; a predetermined workflow may be simpler to control. Evaluate completed tasks that meet the required quality standard, total staff effort, exception burden and cost rather than awarding credit for autonomy itself. [G04, G05]

Retain the Quoted Bookkeeping Tool Cost Until a Suitable Alternative Is Verified

The supplied bookkeeping tool costs \$25 per client per month. No included bookkeeping agent has been established. If the office-software audit discovers a credible alternative, test it and obtain its actual terms before changing the base case.

REQUIRE A SEPARATE DECISION BEFORE PERMITTING AUTOMATIC LEDGER ENTRIES

A later proposal for automatic low-risk posting would require a separate action policy, evidence of reliable reconciliation, limits on value and volume, and explicit approval. The initial recommendation does not authorize it.

D3 / TAX INTAKE ECONOMICS

Calculate Whether Tax Data-Entry Savings Exceed Tool Costs and the Cost of Additional Senior Review

The 2,600 individual returns require 1.1 hours of data entry each, or 2,860 hours. Of these, 1,700 are contractor hours and 1,160 are salaried staff hours. The vendor claims up to 50% savings. Verification adds 0.25 senior hours per return, or 650 hours. Treat 50% as a scenario to test.

AT THE CLAIMED 50% REDUCTION	CONTRACTOR-FIRST ALLOCATION	PROPORTIONAL ALLOCATION
Gross entry hours released	1,430	1,430
Contractor hours avoided	1,430	850
Contractor spending avoided at \$70	\$100,100	\$59,500
First-year fees: \$15,600 + \$4,000	\$19,600	\$19,600
First-year gain with spare senior capacity	\$80,500	\$39,900
Recurring gain after \$15,600 fees	\$84,500	\$43,900
If all 650 senior hours displace \$175 hourly contribution	-\$113,750	-\$113,750
First-year gain after lost senior contribution	-\$33,250	-\$73,850

Determine Whether Data-Entry Savings Reduce Contractor Spending or Salaried Staff Work

A reduction in salaried entry work does not automatically cut cash spending. Contractor-first allocation is favorable and must be operationally achievable. Proportional allocation is a distinct assumption, not a conservative guarantee.

Calculate How Many Hours of Displaced Senior Work Would Eliminate the Tax Savings

Suppose each hour of additional verification prevents a senior accountant from completing other paid work that would contribute \$175 after its additional delivery costs. The first-year tax savings would then be eliminated after 460 hours in the contractor-first scenario or 228 hours in the proportional scenario. If senior accountants have spare time or no additional paying work is available, that lost contribution does not arise. The 650 verification hours still need to fit the work schedule.

RECALCULATE THE TAX SAVINGS IF THE TOOL SAVES LESS TIME THAN THE VENDOR SUGGESTS

At 35% gross entry savings, the first-year gain before loss of other productive senior-accountant work falls to \$50,470 under allocation of time savings to contractor work before salaried staff work and \$22,050 under allocation of time savings in proportion to contractor and salaried entry hours. Both still require quality and workload validation.

D3 / INTAKE AND REVIEW

Test Whether AI Review Assistance Improves the Financial Return From Tax Document Extraction

Review assistance could reduce the 650 additional verification hours, improve detection or create new rework. It deserves a separate test because its effect may interact with extraction. Do not assume that adding a second AI tool automatically repairs the first tool’s economics.

TEST CONDITION	QUESTION ANSWERED
Current preparation and review	What is the baseline accepted output, labor and error rate?
Extraction with current review	What does extraction change by itself?
Current preparation with review assistance	What does review assistance change without extraction?
Extraction with review assistance	Does the combination improve the full workflow, and at what total cost?

Determine Whether Combining Extraction and Review Assistance Changes Their Individual Benefits

Compare the combined change with the sum of the two separate changes. A favorable interaction would mean the combination performs better than those separate effects suggest. Use comparable historical files, randomized assignment where practical, independent scoring and enough observations to detect a decision-relevant effect. [G03]

Measure All Senior Time Required to Verify and Approve a Tax Return

Measure time spent locating evidence, checking extraction, identifying exceptions, correcting outputs and signing off. Test planted errors and naturally difficult files. A correct suggestion is useful only if the reviewer can verify and act on it efficiently.

Obtain the Review-Assistance Price and Measure Its Effect Before Estimating Profit

The case does not supply a review-assistance quotation or measured savings. Record these as unresolved inputs. Do not insert a convenient price or credit all 650 hours as recovered. Test on historical material before considering a live change after April 15.

MEASURE THE COMBINED RESULT AND IDENTIFY THE SEPARATE EFFECTS WHERE THE TEST PERMITS

The full package can have a measurable effect even when its components interact. Report the package effect and, if the design supports it, separate and conditional effects. Do not label staffing or process gains as AI-only gains.

D3 / NON-AI COMPARISON

Calculate Whether Adding a Seasonal Tax Reviewer Could Increase Bellhaven’s Profit

Bellhaven reportedly declined individual and business tax work. The table estimates the additional contribution if that work can still be obtained and completed with a seasonal reviewer. Contribution here means additional revenue after the additional costs of delivering the work.

CALCULATION	AMOUNT
140 declined individual returns × \$650	\$91,000
30 declined business returns × \$3,200	\$96,000
Potential additional revenue	\$187,000
Delivery costs at 45%, excluding the reviewer	−\$84,150
Contribution before the reviewer	\$102,850
520 reviewer hours × \$95	−\$49,400
Illustrative incremental contribution	\$53,450

Confirm That Additional Clients and Staff Capacity Can Support the Reviewer’s Projected Return

The declined customers must still be available and willing to pay. Preparation capacity and the delivery-cost estimate must support the work. The 520 hours must be sufficient for the required review at acceptable quality. Confirm the timing, credentials and actual engagement terms.

Calculate the Additional Revenue Needed to Cover the Seasonal Reviewer’s Cost

At a 55% contribution margin before the reviewer, approximately \$89,818 of additional collected revenue covers the \$49,400 reviewer cost. This is a useful decision threshold, not proof that the historical declined work will convert.

Distinguish the Reviewer’s Contribution From Profit Attributable to AI

The \$53,450 is a modeled non-AI contribution. Adding it to the favorable \$220,500 AI scenario gives \$273,950 for a combined program only if the resources and demand do not overlap. That arithmetic does not meet a \$250,000 AI-attributable target by itself.

COMPARE THE FINANCIAL RETURN FROM AI WITH THE RETURN FROM ADDITIONAL REVIEW STAFF

The firm should compare ways to improve profit and service, not only competing AI products. A conventional staffing change may solve a more valuable problem, or make an AI intervention useful. Either outcome should be visible.

D4 / CONFIDENTIALITY AND CONTROLS

Approve How AI Services Will Handle Bellhaven’s Client Information Before Use

Before client records enter an AI service, Bellhaven must approve the service, information use and access arrangements. The table assigns practical requirements to each area of control. These requirements also apply to connected tools and any agent acting on the firm’s behalf.

CONTROL	BELLHAVEN IMPLEMENTATION
Approved service	Counsel reviews terms before client information is used, including retention, training use, subprocessors and deletion.
Identity and access	Named accounts, appropriate authentication and least-privilege client access. No shared broad credentials.
Information boundaries	Separate historical test copies from production; minimize client information in prompts and logs.
Professional review	A credentialed reviewer remains accountable for client-facing work and its supporting evidence.
Actions and approvals	Initial agents prepare proposals; consequential external actions require explicit authorization.
Incident response	A named owner can suspend the workflow, preserve records, assess exposure and invoke the existing response process.
Continuity	Staff can retrieve work status and complete the task through the prior approved process.

Investigate Reported Client-Information Use in Personal AI Accounts

The survey reports four people entering client information into personal AI tools. Establish what information and services were involved through the appropriate internal process. Provide a safe reporting route and an approved alternative. Do not treat the survey count as a complete incident assessment.

Apply Confidentiality, Security and Professional Duties to AI-Assisted Accounting Work

The FTC Safeguards Rule includes covered tax preparation firms. Professional duties still apply when AI assists work. Review relevant IRS guidance and the unresolved AICPA objection concerning fees; do not convert that debate into an automatic fixed-fee reduction assumption. [G07, G08, G09]

CONFIRM THE APPLICABLE INSURANCE TERMS BEFORE ASSUMING AN ADDITIONAL PREMIUM

Ask the broker how the proposed uses fit existing professional and cyber coverage. Record actual conditions and quotations. The model does not assume a mandatory new policy or a supplied premium.

D5 / PILOT COMPARISON

Compare Bookkeeping Quality and Total Staff Time With and Without AI Assistance

Use the four existing teams of 45 clients to create a practical staged comparison. The initial paid scope is 90 clients for two months, followed by 180 for two months only if the next stage is justified. Assignment, work mix and timing must be documented before the pilot starts.

MEASURE	DEFINITION
Accepted work	Completed client-month work that passes the same quality standard.
Total staff time after including AI-related work	Sorting, matching, review, corrections, exceptions, training during operation and recovery, by role.
Quality	Errors by severity, missed exceptions, reviewer catches and customer-impacting defects.
Time taken at each stage	Arrival, preparation completion, review completion and final delivery times.
Cost	Actual tool charges, additional labor, setup, support and failed runs.
Use of saved staff hours	Hours that can cover new clients or remove a documented hiring or contractor commitment.

Account for the Limited Number of Independent Teams in the Pilot

Randomize team assignment where practical, inspect prior trends and work mix, and report the limited number of independent groups. Repeated client observations within a team do not create the same information as many independent teams. Use the initial results to estimate variability and refine the next test.

Use Historical Files to Diagnose Errors Before Expanding the Quality Test

The original starting set includes 30 bookkeeping and 30 tax files, with sample production review across teams. Retain it as a diagnostic starting point, then expand according to risk, variability and the decision threshold. It is not enough to certify rare failures are absent.

REQUIRE ACCEPTABLE QUALITY AND A JUSTIFIED FINANCIAL RETURN BEFORE EXPANSION

Critical failures stop the affected use. Ordinary quality must be acceptable before favorable time savings justify expansion. Economic thresholds must reflect actual hiring or contribution, rather than a generic five- or ten-percent target.

D5 / ILLUSTRATIVE RESULTS

Determine Whether the Illustrative Bookkeeping Results Justify Expanding the AI Pilot

The results on this page are an illustrative example from the supplied case. They are not measurements collected from Bellhaven. The comparison shows how to deduct changes seen in untreated teams and the extra review work introduced by AI.

TEAM	BASELINE SORTING HOURS	PILOT-PERIOD HOURS	CHANGE
Pilot team A	6.5	5.7	-0.8
Pilot team B	6.5	5.5	-1.0
Comparison team C	6.5	6.4	-0.1
Comparison team D	6.5	6.0	-0.5

The pilot teams improve by 0.9 hours on average; the comparison teams improve by 0.3. The difference is 0.6 hours. Subtract 0.2 hours of additional review and the illustrative net improvement is 0.4 hours per client-month, or about 6.2% of the original 6.5 sorting hours.

IMPLICATION	ILLUSTRATIVE RESULT
Annual capacity at 200 clients: $200 \times 0.4 \times 12$	960 hours
Additional-client work remaining: $2,640 - 960$	1,680 hours
Positions still required at 1,400 hours each	Two
Spending on additional staff demonstrably avoided	None
Errors per file: pilot versus comparison	0.8 versus 0.6
Planted critical errors detected	4 of 4
Planted minor errors detected	13 of 16

Investigate Quality Problems and Insufficient Staffing Savings Before Expanding

The result does not reach the 7.95% threshold for avoiding one position under the 20-client growth assumption. Quality also needs investigation. A limited repair test can be reasonable if it addresses a specific cause at a defined cost; continued spending is not justified merely because the point estimate is positive.

REPORT THE LIMITS OF THE SMALL ILLUSTRATIVE PILOT

The four pairwise net comparisons range from 0.1 to 0.7 hours. This is a descriptive range, not a confidence interval. Four of four critical catches do not establish reliable detection of rare critical errors.

D5 / EXPANSION RULES

Use Quality Results, Staff-Time Savings and Costs to Decide Whether to Expand the Bookkeeping Pilot

Use the conditions below to decide whether the bookkeeping pilot should stop, undergo a limited revision or expand. The time-saving thresholds come from the 20-client growth scenario explained earlier and remain conditional on client demand and actual staffing needs.

OBSERVED CONDITION	DECISION
Critical confidentiality or unauthorized-action failure	Suspend the affected workflow, preserve evidence and investigate.
Quality worse or review burden unexplained	Repair and retest before expansion; do not offset quality concerns with speed.
Net saving below the first hiring threshold	Do not claim avoided hiring. Identify another verified benefit or stop.
Net saving clears 7.95% but not 16.92%	Model at most one avoided position, subject to demand and team capacity.
Net saving clears 16.92%	Test whether both positions can actually be avoided across roles and months.
Demand or prices differ from plan	Recalculate capacity, staffing and profit before committing.

Set a Cost Limit and Specific Question for One Additional Month of Testing

One additional month at 90 clients costs \$2,250 in bookkeeping tool fees before staff time and other charges. Authorize it only with a named defect, proposed fix and decision criterion. If an extension replaces a later rollout month, rephrase the cash plan rather than mechanically adding both charges.

Confirm Whether Bellhaven Would Actually Add Staff Without AI

Finance and operations should sign off on the staffing plan that would exist without AI. If the firm would defer hiring regardless because only eight clients sign, the model cannot credit AI with avoiding two positions.

Record the measured net saving, quality result, confirmed demand, staffing decision and authorized next-stage cost.

RETAIN FINDINGS THAT HELP BELLHAVEN AVOID AN UNSUITABLE DEPLOYMENT

A negative or inconclusive result is useful when it identifies a specific limitation before a larger contract or operational commitment. Retain the tested cases and failure explanations.

D6 / STAFF ADOPTION

Address the Reasons Bellhaven Staff Use AI, Avoid AI or Use Unapproved Services

The survey has 58 responses out of 64 staff. Nineteen report weekly AI use. Among the 39 other respondents, 22 are unsure what is permitted, 14 do not know how to use it and nine do not find it useful. These reasons can overlap; they must not be added as distinct people.

BARRIER OR BEHAVIOR	RESPONSE	EVIDENCE
Unclear permission	Publish permitted tasks, information rules and the approved route.	Staff can correctly classify practical examples.
Lack of skill	Provide role-specific exercises with checking and escalation.	Successful work samples, not attendance alone.
Low perceived usefulness	Examine the actual tasks and friction before pushing use.	Quality and total effort on eligible work.
Unsafe personal-tool use	Investigate through the appropriate process and provide a workable alternative.	Reduced unsafe workarounds and prompt reporting.
Concern about jobs or workload	Explain the 12-month no-layoff commitment and planned use of capacity.	Staff feedback, workload distribution and manager follow-through.

Ask Accountants With Different Experience Levels to Test AI Outputs and Verification Procedures

Experienced accountants should help identify plausible errors and exceptions. Junior staff should learn to verify evidence and recognize limits. External accounting and training research supports testing differences by experience; it does not justify assuming every group benefits equally. [R06, R12]

Assign the Hours AI Saves to Specific Client Work

For bookkeeping, the proposed destination is new client work that would otherwise require hiring. For tax intake, distinguish contractor hours from salaried hours and senior verification. Managers must schedule the actual work; a spreadsheet total does not transfer capacity by itself.

MEASURE WHETHER APPROVED AI USE IMPROVES COMPLETED WORK

Track eligible work completed safely, completed work that meets the required quality standard, support needs and staff effort. Avoid targets based only on logins or prompts, which can reward unnecessary use and additional cost.

D7 / MEASURING FINANCIAL BENEFITS

Distinguish Staff Hours Saved From Actual Spending Reductions and Additional Profit

An hour saved, an expense avoided and additional profit are different results. The table identifies how Bellhaven should report each proposed benefit and which records would be needed to support a financial conclusion.

CLAIM	PERMITTED TREATMENT	EVIDENCE REQUIRED
Bookkeeping time released	Capacity until a staffing commitment changes.	Role and month allocation; signed client demand.
One or two avoided hires	\$98,000 or \$196,000 annual avoided cost when established.	A credible no-AI hiring plan and the actual decision.
Tax entry hours released	Cash only for contractor or other spending actually avoided.	Assignment records and reduced commitments or invoices.
Additional verification by senior accountants	Opportunity cost only when it displaces valuable work; still a capacity burden otherwise.	Available hours, paying demand and incremental contribution.
Additional tax clients	Collected contribution after delivery and reviewer costs.	Conversion, completion, billing and collection.
Improved quality or service	Report directly until a financial effect is supported.	Comparable errors, deadlines, retention or loss records.

Do Not Assign the Same Saved Hours to Both Client Growth and Contractor Savings

At 20% bookkeeping improvement, 3,120 hours cover 2,640 hours of growth and leave 480. Those 480 could hypothetically cover 120 hours during tax season, worth \$8,400 at \$70, only if timing, skills and actual contractor avoidance align. That benefit is excluded from the main scenarios because those conditions are unverified.

Compare AI Results With What Bellhaven Would Have Done Without AI

Compare the pilot with similar untreated work and reconcile operating changes with actual finances. Record seasonality, price changes, demand and scheduling improvements. The estimate can concern the AI intervention or a clearly defined combined package, but its label must match the comparison.

Record the benefit, responsible manager, hours or spending affected, expected result without AI, supporting financial record and possible double counting.

D7 / RECURRING SCENARIOS

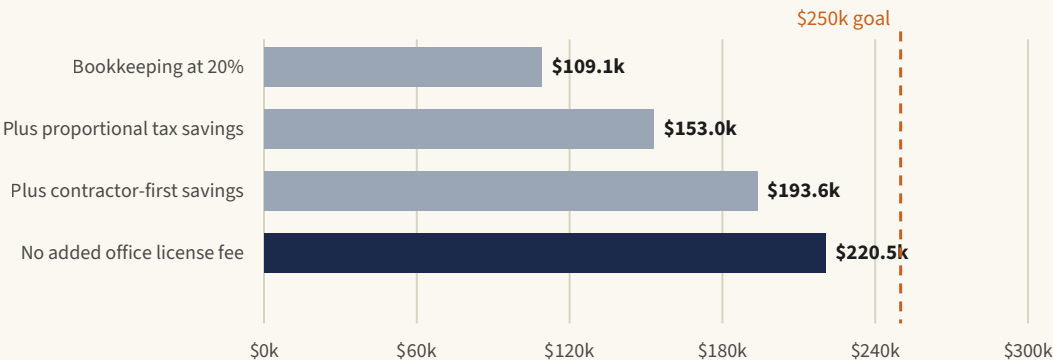
Calculate How Far the Proposed AI Applications Could Take Bellhaven Toward Its \$250,000 Annual Profit Target

The scenarios below combine bookkeeping and tax assumptions to estimate annual operating profit after the tools are fully in use. Each scenario holds the other stated assumptions constant so the financial effect of a particular change remains visible.

ANNUAL RECURRING SCENARIO	AVOIDED HIRING	TAX GAIN	TOOLS AND LICENSES	NET GAIN
A: 10% bookkeeping improvement	\$98,000	\$0	-\$86,880	\$11,120
B: 20% bookkeeping improvement	\$196,000	\$0	-\$86,880	\$109,120
C: B plus proportional tax savings	\$196,000	\$43,900	-\$86,880	\$153,020
D: B plus contractor-first tax savings	\$196,000	\$84,500	-\$86,880	\$193,620
D with approved office access at no incremental license fee	\$196,000	\$84,500	-\$60,000	\$220,500

Bookkeeping tools at 200 clients cost \$60,000 annually. The 64-seat AI service costs \$26,880. Tax gains already deduct the \$15,600 recurring tax fee. Scenarios assume acceptable quality, all 20 additional clients, staff hours available for the required work, unchanged prices and no loss of contribution from other senior-accountant work. One-time costs are excluded from recurring results.

The Remaining Distance to the \$250,000 Goal



These are modeled annual operating-profit improvements, not observed outcomes.

Identify What Would Be Needed to Close the Remaining \$29,500 Profit Gap

The favorable \$220,500 scenario leaves \$29,500 to find through measured additional contribution, lower costs or a better-performing alternative. Do not fill it with a hypothetical free ledger agent or unverified time savings.

DISTINGUISH FIRST-YEAR IMPLEMENTATION COSTS FROM ANNUAL COSTS AFTER FULL DEPLOYMENT

The recurring table assumes mature operation at 200 clients and omits one-time costs. The first-year cash schedule uses phased deployment and must be reconciled with when clients actually arrive.

D7 / SENSITIVITY

Determine How Changes in Performance, Client Growth and Pricing Affect the AI Recommendation

The table changes important assumptions in the recurring financial scenarios. Each change is assessed separately unless the row says otherwise. The purpose is to identify which uncertainty could change Bellhaven’s decision to proceed.

CHANGE FROM THE FAVORABLE ASSUMPTIONS	ANNUAL CONSEQUENCE OR REVISED RESULT
Bookkeeping improvement is 15%, not 20%	Only one hire avoided; scenario D with 64 licenses falls to \$95,620.
Only eight new clients sign and one hire is genuinely avoided	At 188 clients and \$56,400 tool cost, scenario D with 64 licenses becomes \$99,220.
All 650 senior verification hours displace \$175 contribution	Scenario D with 64 licenses falls from \$193,620 to \$79,870.
Bookkeeping fees fall 1% at 200 clients	Revenue falls \$34,440; the favorable \$220,500 becomes \$186,060.
Bookkeeping tool price alone must close the \$29,500 gap	Annual tool cost must fall from \$60,000 to \$30,500, about \$12.71 per client-month at 200 clients.

Calculate the Effect of a One-Percent Reduction in Bookkeeping Fees

At 200 clients, annual bookkeeping revenue is \$3,444,000. A one-percent fee change is larger than the remaining \$29,500 gap in the most favorable supplied scenario. This does not predict a fee cut; it shows why price and engagement terms belong in the business case.

Treat a Free Bookkeeping Tool as a Hypothetical Scenario Until a Suitable Product Is Identified

Removing the entire \$60,000 bookkeeping fee would produce \$280,500 under the other favorable assumptions. A one-percent fee reduction would bring that hypothetical result to \$246,060. Neither calculation establishes that a free suitable tool exists; both illustrate sensitivity only.

INVESTIGATE DEMAND, STAFFING, SENIOR REVIEW AND TOOL COSTS FIRST

Confirm demand, the hiring threshold, loss of other productive senior-accountant work and actual tool terms before refining minor cost items. These variables can move the result by tens of thousands of dollars or remove the benefit entirely.

D8 / FIRST-YEAR CASH

Calculate Whether the Proposed First-Year AI Program Fits Within Bellhaven’s \$120,000 Cash Budget

The first-year cash schedule includes phased tool use and the supplied setup, training and legal allowances. Compare the total with the \$120,000 cash limit. Later rows show how the total changes under different arrangements for staff access to AI.

FIRST-YEAR ITEM	CALCULATION	CASH
Staff AI service	$64 \times \$35 \times 12$	\$26,880
Bookkeeping stage one	$90 \times \$25 \times 2 \text{ months}$	\$4,500
Bookkeeping stage two	$180 \times \$25 \times 2 \text{ months}$	\$9,000
Bookkeeping later operation	$180 \times \$25 \times 7 \text{ months}$	\$31,500
Setup and training	Case allowance	\$8,000
Historical tax-intake test	$200 \times \$6 + \$4,000 \text{ setup}$	\$5,200
Legal review	Case allowance	\$10,000
Base first-year cash		\$95,080
Remaining below the ceiling	Before additional items	\$24,920

ALTERNATIVE STAFF ACCESS	TOTAL FIRST-YEAR BASE CASH	REMAINING
30 licenses instead of 64	\$80,800	\$39,200
Approved included office feature, no added license fee	\$68,200	\$51,800

Include Existing Staff Time When Assessing the Full Economic Cost

Existing staff and partner time still have opportunity cost and must be measured. The cash allowance does not establish that legal review or setup work will fit the estimates. Costs of any new review-assistance product are unresolved.

ADD TOOL CHARGES FOR NEW CLIENTS WHEN THEY ENTER THE FIRST-YEAR SCHEDULE

The supplied first-year bookkeeping budget uses 180 clients in later months. Each additional client-month costs \$25. Rephase the budget when the 20 new clients arrive; the recurring 200-client model cannot be silently substituted without adjusting cash.

D8 / BUDGET CHOICES

Determine Which Combinations of AI Tools and a Seasonal Reviewer Fit Within the First-Year Budget

The table compares combinations of the first-year AI program, live tax processing and a seasonal reviewer. “Included-office route” means the base program uses an approved feature in existing office software with no additional staff license fee; the separate bookkeeping tool still has its stated cost.

FIRST-YEAR PACKAGE	CASH TOTAL	POSITION AGAINST \$120,000
Base with 64 licenses	\$95,080	\$24,920 available
Base with 64 licenses plus live tax fees	\$110,680	\$9,320 available
Base with 64 licenses plus reviewer	\$144,480	\$24,480 over
Included-office route plus reviewer	\$117,600	\$2,400 available
Included-office route plus reviewer and live tax fees	\$133,200	\$13,200 over
Stop after stage two, with 64 annual licenses	\$63,580	\$56,420 available

Reviewer: \$49,400. Added live tax fees: \$15,600. The included-office route requires approval and functional suitability; it is not assumed to contain the bookkeeping tool. Stopping after stage two removes the \$31,500 later bookkeeping operation from the stated schedule.

Confirm That the Included-Office Route Plus a Reviewer Totals \$117,600

The included-office base of \$68,200 plus a \$49,400 reviewer is \$117,600. It fits the stated ceiling by \$2,400. Rejecting that combination as automatically unaffordable would remove a potentially useful alternative for the wrong reason.

Identify Additional Costs That Could Exceed the Remaining \$2,400 Budget

Additional client-months, a review-assistance subscription, higher legal costs or a pilot extension can consume \$2,400 quickly. The route should not be described as comfortably funded. Confirm timing and quotes before committing the reviewer alongside the pilot.

Decide Which Expenditures to Authorize at Each Stage

The firm can compare a reviewer, a tax rollout and further AI testing without authorizing all of them concurrently. Use the first checkpoint to decide which uncertainty is worth the next cash commitment.

MEASURE AND LIMIT THE FULL COST OF AGENT-ASSISTED ACCOUNTING WORK

Track total cost per successfully completed client-month or return, including failed runs, retries, review and recovery. Set spending and action limits. The charge for processing text excludes other tool fees, staff review and recovery needed to complete an accounting task.

SYNTHESIS / COMBINED FINDINGS

Explain How Staffing, Tax Review, Client Fees and AI Costs Affect Bellhaven’s Recommendation

The recommendation depends on how the findings affect one another. Time savings matter financially when they change work assignments or spending; review requirements can consume those savings; prices and client demand determine whether the projected return remains achievable.

FINDING
Bellhaven’s financial result depends on whether AI saves usable staff hours, whether client demand justifies the work and whether the firm changes actual spending. These conditions must be confirmed together before the projected profit can be relied on.

WHAT COULD WEAKEN THIS FINDING
The combined program can still fail if demand, prices, senior capacity or implementation costs move against the case assumptions.

COMBINED FINDING	WHY IT CHANGES THE RECOMMENDATION
Bookkeeping savings reduce staffing costs only when enough hours are saved to avoid a planned position	Modest speed gains can be real yet avoid no spending. Pilot precision matters most near 7.95% and 16.92%, and only if demand converts.
Tax document extraction and the additional accountant review must be evaluated together	Preparation savings can consume senior capacity. Review assistance or a reviewer may change the result, but their costs and interactions must be measured.
Staying within the first-year budget does not establish that the annual profit target will be achieved	The \$117,600 reviewer package can fit first-year cash while the best supplied AI-only recurring scenario still misses the \$250,000 goal.
A lower license cost leaves task performance and staff-review requirements unresolved	The included-office route saves \$26,880 if approved, but it neither proves bookkeeping automation nor removes verification work.
Client fee changes can have a larger financial effect than modest improvements in AI performance	A one-percent bookkeeping fee reduction costs \$34,440, more than the favorable scenario’s remaining target gap.
AI and conventional staffing or process changes may both contribute to the best result	A reviewer or improved document collection may produce superior returns or enable AI. Keep the combined value and AI attribution distinct.

What Bellhaven Should Authorize and What It Should Test Next

Authorize the governed staff route and a bookkeeping pilot with a defined scope and spending limit. Test tax intake and review on historical files. Obtain a concrete reviewer proposal and verify the included-office alternative. Make expansion depend on work that meets the required quality standard, staff hours available for the required work, actual hiring decisions and a reconciled cash plan.

Which Financial Results Bellhaven Cannot Yet Promise

Do not promise two avoided hires, the full contractor-first tax saving or the \$250,000 AI-attributable target. Each requires evidence that is presently absent. A disciplined smaller return can be a better decision than a larger forecast assembled from incompatible assumptions.

D9 / STAYING CURRENT

Monitor AI Advances That Could Improve Bellhaven’s Bookkeeping and Tax Work

Bellhaven should monitor advances that could improve the tasks or costs identified in this analysis. The table names the capability, explains its relevance and states what would justify testing it.

WATCH ITEM	WHY BELLHAVEN CARES	TRIGGER FOR A TEST
Bookkeeping integrations and restricted agents	Could improve matching, exception handling or cost.	A credible product can pass the same client-month evaluation at better total economics.
Document and multimodal processing	Could reduce extraction errors or difficult-file handling.	Improvement on Bellhaven’s historical document set.
Review assistance	Could reduce scarce senior effort or improve detection.	A testable claim with evidence, usable output and a quoted cost.
Included office features	Could change license needs or action permissions.	New capability, changed default or material terms change.
Model and agent reliability	Could improve completed workflows.	Better local end-to-end results, not a leaderboard alone.
Security, dependencies and professional guidance	Could invalidate an approved operating arrangement.	Material incident, changed integration or applicable guidance.

Assign Time and Responsibility for Monitoring Relevant AI Changes

Use a short monthly scan and a quarterly operating review. The original plan allocates two hours for monitoring and five manager hours for review activity monthly; confirm actual availability and record the opportunity cost. Escalate urgent security or permission changes immediately rather than waiting for the next meeting.

Use External Research to Identify Tests Relevant to Bellhaven’s Work

Task experiments, accounting studies and agent benchmarks can suggest where to test. Their populations, tools and outcomes differ. Maintain a source register with the claim, design, limitation and Bellhaven decision it could affect. [R01, R03, R06, R14, R15]

RETAIN WORK SAMPLES SO NEW AI SYSTEMS CAN BE COMPARED WITH THE CURRENT PROCESS

Retain representative files, difficult cases, expected results and failure traces. Re-run them when models, prompts, connectors or permissions change. This makes staying current a controlled comparison rather than a repeated restart.

IMPLEMENTATION / PILOT CHARTER

Record the Scope, Budget and Conditions for Bellhaven’s AI Pilot

Complete this authorization before the limited pilot begins. Record the actual people, services, information and spending approved, rather than assuming that the recommendations in this report constitute implementation approval.

Workflow, participating teams, eligible client work, permitted information and excluded actions.

Assignment and comparison, baseline, sample, dates and treatment of client or workload differences.

Acceptance standard, critical failures, independent reviewer and quality sampling.

Cash limit, staff time, approved services, access scope and next-stage commitments.

Stop, repair and expansion criteria; fallback owner; next decision and required evidence.

REQUIREMENTS THAT REMAIN IN FORCE DURING THE PILOT

Counsel-approved terms before client data; credentialed review of client-facing work; no new live tax workflow during January 15 to April 15; six partner project hours weekly; no layoffs for 12 months. Changes require an explicit management decision.

IMPLEMENTATION / STAFF COMMUNICATION

Explain the AI Pilot, Staff Responsibilities and Employment Commitment to Bellhaven’s Staff

We are beginning a limited AI pilot to learn whether selected bookkeeping tasks can be completed more efficiently while maintaining the quality and confidentiality our clients expect.

The first pilot will involve selected teams and approved client work. We will compare the full process, including review and corrections, with our current approach. We have not assumed that the tool will save a particular amount of time or justify a wider rollout.

Use only the services and information categories approved for your role. Do not enter client information into personal AI accounts or connect an unapproved tool. If you have already done so, contact the designated internal lead promptly so we can assess what happened and respond appropriately.

AI output must be checked against the underlying records. A credentialed reviewer remains responsible for client-facing work. If information is missing, an answer seems wrong, or a tool attempts an unexpected action, stop that step and contact the pilot lead. The existing approved workflow remains available.

Training and practice time will be provided. We want staff to report errors, confusing instructions and tasks for which the tool is not useful. We will judge the pilot by work that meets the required quality standard, total effort and service quality, not by how many prompts people enter.

The firm’s commitment is no layoffs for 12 months. The proposed use of bookkeeping capacity is to serve additional clients and reduce the need for planned hiring. Managers will explain any changes to work assignments.

We will review the first evidence within 45 days and explain whether we will stop, change or continue the limited pilot. Wider use will require further evidence and approval.

Before sending: insert the named pilot lead, approved services, help channel, training arrangements and first review date.

DRAFT FOR INTERNAL USE

This message is ready for management adaptation. It has not been sent. Confirm that operational arrangements exist before making the commitments in the message.

DECISION / MANAGEMENT RECORD

Record Which AI and Staffing Expenditures Bellhaven Has Authorized

Use this record to document management’s actual authorization. Separate spending approved now from proposals that still require a quotation, performance test or further decision.

Decision and scope: what is authorized now, what is deferred and what is the next checkpoint?

Record the authorized expenditures, actual quotations, expected arrival of new clients, remaining cash budget and finance approval.

Record quality results, staff hours saved after review, confirmed client demand, the no-AI staffing plan, and estimated versus achieved benefits.

Operational, security, legal, finance and training owners; partner-time allocation and fallback responsibility.

Conditions that would stop or change the decision: incidents, weak results, demand changes, pricing, costs or a better alternative.

RECOMMENDED INITIAL AUTHORIZATION

Approve preparation and a pilot with a defined scope and spending limit subject to information and quality controls. Do not approve full rollout or promise the annual profit target from the current assumptions.

METHOD / CORRECTIONS AND INTERPRETATION

Explain the Financial Model’s Assumptions, Corrections and Unresolved Questions

The table records corrections and interpretive limits applied to the supplied financial model. These distinctions explain how to use the figures without treating an assumption or illustrative result as an observed business outcome.

ISSUE	TREATMENT IN THIS DOCUMENT
Reviewer budget	Corrected: \$68,200 + \$49,400 = \$117,600, leaving \$2,400 before additional items.
Included office software	A conditional license alternative. No free bookkeeping agent is inserted into the supplied case.
Profit attribution	Separates AI, non-AI changes and possible interactions. Mixed interventions are measurable with suitable designs.
Illustrative pilot	A 0.4-hour net gain does not avoid one planned hire at 20-client growth. The pairwise range is not a confidence interval.
Tax verification	Separates spare senior capacity from contribution lost because senior accountants cannot perform other paid work and contractor-first from proportional savings.
Growth timing	Recurring economics use 200 clients; first-year supplied tool charges use 180 in later months and need rephasing for growth.
Fee guidance	Acknowledges professional guidance and unresolved interpretation. No universal fixed-fee reduction is assumed.
Research and benchmarks	Uses source-specific limitations; task results and benchmark scores are not Bellhaven profit forecasts.

How to Use the Numbers

All dollar amounts are US dollars. Annual scenarios assume the stated workload and prices throughout the modeled period. Cash schedules and recurring profit are deliberately separate. Rounded percentages may not sum exactly. Negative values represent costs or reductions.

Which Inputs Bellhaven Still Needs Before Authorizing Wider Use

The dossier does not establish actual AI performance, review-assistance pricing, approved vendor terms, customer conversion, realized staffing avoidance or the precise timing of new clients. The worksheets and gates specify how to obtain that evidence.

HOW THE BLUEPRINT AND WORKED EXAMPLE RELATE TO EACH OTHER

The AI Adoption Blueprint supplies the general nine-decision method. This worked example applies it to Bellhaven, preserving the supplied operating inputs while making assumptions and decision thresholds explicit.

CASE SOURCES / OPERATING INPUTS

Review the Business and Staffing Assumptions Used in the Bellhaven Analysis

The table preserves the operating assumptions supplied for the Bellhaven case. Use these inputs to reproduce the analysis and identify which figures would need verification in a real implementation.

INPUT GROUP	CASE VALUES RETAINED
Firm	64 people; Columbus, Ohio; \$9.6 million revenue; \$1.3 million operating profit after partner salaries.
Staff mix	Six partners, ten managers, 22 senior or staff accountants, 14 juniors with under two years of experience, 12 administrative or client-service staff.
Objective and limits	\$250,000 additional annual operating profit within 24 months; \$120,000 additional first-year cash spending; first decision in 45 days; six partner hours weekly.
Operating commitments	No layoffs for 12 months; credentialed review of client deliverables; counsel-approved terms before client data; no new live tax changes January 15 to April 15.
Bookkeeping	180 clients, four teams of 45; \$1,435 monthly fee; 6.5 sorting hours, 2.5 correction and review hours, two client-question and reporting hours monthly.
Growth	20 planned clients: eight signed and 12 in pipeline; two otherwise-planned positions at \$98,000 and 1,400 client hours each annually.
Individual tax	2,600 returns at \$650; 1.1 entry hours each, including 1,700 contractor and 1,160 salaried hours; 0.25 added senior verification hours per return.
Contract and senior rates	\$70 per contractor hour; \$175 per senior hour only where it represents contribution lost because senior accountants cannot perform other paid work.

Source: the supplied AI Adoption Worked Example dossier and its internal case-source register. These are scenario inputs, not independently verified company records. The original ZIP remains the source record.

IDENTIFY ANY PROPOSED CHANGE TO THE SUPPLIED CASE ASSUMPTIONS

A different tool, staffing arrangement or pricing model may improve the decision. It must be presented as an alternative with new evidence, not substituted into the original case as though it had always been supplied.

CASE SOURCES / COMMERCIAL AND PILOT INPUTS

Review the Tool Prices, Survey Responses and Illustrative Pilot Results Used in the Analysis

The table records the supplied commercial terms, staff survey responses and illustrative test results. The quoted prices and modeled performance must be distinguished from terms and results that Bellhaven would need to verify before purchasing.

INPUT GROUP	CASE VALUES RETAINED
AI service and bookkeeping tool	\$35 per staff license monthly; \$25 per bookkeeping client monthly. An included office capability has no incremental license fee but requires review.
Tax intake tool	\$6 per individual return; \$4,000 setup; vendor claim up to 50% entry-time reduction.
First-year allowances	\$8,000 setup and training; \$10,000 legal review; historical tax test of 200 returns.
Bookkeeping cash stages	90 clients for two months; 180 for two months; 180 for seven later months.
Reviewer alternative	520 hours at \$95; 140 declined individual returns and 30 business returns; 45% delivery costs excluding the reviewer.
Delay evidence	Week-nine sample: 34 of 60 files waiting for review, 19 for client documents, seven for other reasons; 395 extensions reported.
Staff survey	58 responses; 19 weekly users; four report client information in personal tools; among 39 others, 22 unclear on permission, 14 unsure how, nine find it unhelpful. Reasons overlap.
Illustrative pilot	All teams start at 6.5 sorting hours; pilot 5.7 and 5.5; comparison 6.4 and 6.0; additional review 0.2 hours; errors 0.8 versus 0.6; planted catches four of four critical and 13 of 16 minor.

Source: the supplied worked example. The pilot observations are an illustrative result embedded in that case, not evidence collected from a live implementation. Proposed additional tests and operational controls in this rebuild are recommendations.

RESEARCH AND GUIDANCE SUPPORTING THE ANALYSIS

The next pages identify research and guidance used in the two-document analysis. Their inclusion is not an assertion that every external finding applies directly to Bellhaven.

SOURCES / RESEARCH AND GUIDANCE

Consult the Sources Behind the Recommendations, Part 1

Source codes throughout both documents identify the material below. Research findings, vendor guidance and legal interpretation have different strengths. Read the underlying source before relying on an external result.

R01 Brynjolfsson, Li and Raymond: Generative AI at Work

Field research on customer support finds higher productivity, with gains varying by experience. The setting is not a forecast for every occupation.

<https://academic.oup.com/qje/article/140/2/889/7990658>

R02 Dell’Acqua and colleagues: AI assistance in consulting tasks

Randomized task research finds gains on some work and worse results on a task outside the model’s capabilities. Task selection and independent verification matter.

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4573321

R03 Cui and colleagues: Three Field Experiments With Software Developers

Randomized experiments across 4,867 developers report about 26% more completed tasks on average. Effects differ across workers and settings.

<https://pubsonline.informs.org/doi/10.1287/mnsc.2025.00535>

R04 METR: Experienced developer productivity and the follow-up

The original small randomized study found slower completion despite participants’ expectations. The follow-up reports selection problems; it does not supply a reliable universal replacement estimate.

<https://metr.org/blog/2026-02-24-uplift-update/>

R05 Humlum and Vestergaard: Large language models and labor outcomes

Administrative labor records provide a different outcome from task tests. The study found no detectable near-term effect on recorded earnings or hours in its setting.

<https://www.nber.org/papers/w33777>

R06 Choi and Xie: Human + AI in Accounting

Published accounting research combines a survey, platform records and a classification experiment. Most productivity findings are associations; the experiment supports causal inference for its specific accuracy task.

<https://onlinelibrary.wiley.com/doi/10.1111/1475-679x.70052>

SOURCES / RESEARCH AND GUIDANCE

Consult the Sources Behind the Recommendations, Part 2

Source codes throughout both documents identify the material below. Research findings, vendor guidance and legal interpretation have different strengths. Read the underlying source before relying on an external result.

R07 Dillon and colleagues: Shifting Work Patterns With Generative AI

Random access across 7,137 knowledge workers reduced email time. The current paper separates the effect of access from the estimated effect of induced use. It does not establish company profit.

<https://arxiv.org/html/2504.11436v4>

R08 Ju and Aral: Collaborating With AI Agents

The revised experiment reports 2,234 participants and higher advertising output per worker. Text quality, image quality and diversity move differently; one average obscures those differences.

<https://arxiv.org/abs/2503.18238>

R09 McKinsey: The State of AI

Survey results associate greater value with changes to workflows and operating practices. Self-selection and self-reporting prevent a causal interpretation of those associations.

<https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>

R10 BCG: AI at Work

Worker survey results describe adoption, reported time savings and limited guidance on how released time should be used. Self-reported time is not verified cash savings.

<https://www.bcg.com/publications/2026/ai-at-work-why-strategy-matters-more-than-tools>

R11 Faros: The AI Productivity Paradox

Vendor telemetry associates higher code output with longer review times and no significant company-level performance association. The results concern software teams and are observational.

<https://www.faros.ai/blog/ai-software-engineering>

R12 Chen and Bao: AI adoption and training

Experimental research examines whether structured training changes adoption and performance. The narrow professional education setting limits transfer to a whole business.

<https://arxiv.org/abs/2603.04982>

SOURCES / RESEARCH AND GUIDANCE

Consult the Sources Behind the Recommendations, Part 3

Source codes throughout both documents identify the material below. Research findings, vendor guidance and legal interpretation have different strengths. Read the underlying source before relying on an external result.

R13 Research on AI assistance and patent drafting expertise

A three-month professional field experiment examines assisted work and subsequent unaided performance. The results do not justify a universal assertion that AI inevitably erodes or improves expertise.

<https://research.google/pubs/does-ai-assistance-enhance-or-erode-expertise-evidence-from-a-three-month-field-experiment-in-patent-drafting/>

R14 Scale and the Center for AI Safety: Remote Labor Index

A benchmark of extended freelance projects uses expert judgments of final deliverables. Scores depend on task distribution, tools and budgets; they are not accounting-firm completion forecasts.

<https://labs.scale.com/leaderboard/rli>

R15 METR: Time Horizon 1.1

Tracks task duration at specified success probabilities on selected software and research tasks. A 50% success horizon is not a safe autonomous operating duration for business work.

<https://metr.org/blog/2026-1-29-time-horizon-1-1/>

G01 NIST: AI Risk Management Framework

Voluntary risk-management guidance supports governance, measurement and monitoring. It does not certify that a particular deployment is safe or lawful.

<https://www.nist.gov/itl/ai-risk-management-framework>

G02 UK government: Impact evaluation of AI interventions

Explains how to plan evaluation early and distinguish technical performance from impact. Experimental and other methods have different requirements and limitations.

<https://www.gov.uk/government/publications/the-magenta-book/guidance-on-the-impact-evaluation-of-ai-interventions-html>

G03 NIST: Full factorial experimental designs

Explains designs that vary factors independently and permit estimates of separate effects and interactions. Adequate observations and an appropriate design remain necessary.

<https://www.itl.nist.gov/div898/handbook/pri/section3/pri3331.htm>

SOURCES / RESEARCH AND GUIDANCE

Consult the Sources Behind the Recommendations, Part 4

Source codes throughout both documents identify the material below. Research findings, vendor guidance and legal interpretation have different strengths. Read the underlying source before relying on an external result.

G04 Anthropic: Building Effective Agents

Distinguishes predetermined workflows from agents that choose how to proceed. Vendor engineering guidance recommends matching complexity to the task.

<https://www.anthropic.com/engineering/building-effective-agents>

G05 Anthropic: Evaluating AI agents

Explains repeated trials, evaluation of actual outcomes and the relationship between offline tests and production monitoring. This is engineering guidance, not measured financial impact.

<https://www.anthropic.com/engineering/demystifying-evals-for-ai-agents>

G06 OWASP: Risks in agentic applications

Identifies threats involving tools, identities, instructions, memory and interconnected agents. Threat guidance supports testing and controls; it does not quantify local incident probability.

<https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026/>

G07 IRS: Responsible AI use in federal tax practice

Applies professional duties to AI use, including verification, confidentiality and billing. Its discussion of fees should be read with the unresolved professional debate in G08.

<https://content.govdelivery.com/accounts/USIRS/bulletins/41d6e70>

G08 AICPA reporting: Request for clarification on AI and fees

Describes the AICPA's objection to interpreting AI efficiency guidance as requiring automatic reductions in value-based fees. It is not an IRS resolution of the issue.

<https://www.journalofaccountancy.com/news/2026/sep/aicpa-seeks-irs-clarity-on-ai-guidelines-cpa-fees/>

G09 FTC: Safeguards Rule guidance

Explains information-security program and service-provider duties for covered financial institutions, including tax preparation firms. Applicability and implementation require professional review.

<https://www.ftc.gov/business-guidance/resources/ftc-safeguards-rule-what-your-business-needs-know>

SOURCES / RESEARCH AND GUIDANCE

Consult the Sources Behind the Recommendations, Part 5

Source codes throughout both documents identify the material below. Research findings, vendor guidance and legal interpretation have different strengths. Read the underlying source before relying on an external result.

G10 Microsoft: Agentic editing in office applications

Documents changes in product behavior and availability. A vendor release is a reason to examine approved configurations, not proof that a particular firm's permissions changed.

<https://www.microsoft.com/en-us/copilot/blog/2026/04/22/copilots-agentic-capabilities-in-word-excel-and-powerpoint-are-generally-available/>

G11 European Commission: AI regulatory implementation

Official material explains changes to implementation. Duties depend on jurisdiction, the use, and the organization's role. Reconfirm applicable timing before deployment.

<https://digital-strategy.ec.europa.eu/en/news/ai-omnibus-enters-force>

G12 Colorado legislature: AI legislation

Primary legislative material for consequential automated decisions. Counsel should examine the enacted text, affected decisions and applicable commencement provisions.

<https://www.leg.colorado.gov/bills/SB26-189>

F01 LiteLLM: Supply chain incident disclosure

The project identifies compromised package releases and affected installation paths. The incident supports dependency inventory, controlled updates, credential isolation and response planning.

<https://docs.litellm.ai/blog/security-update-march-2026>