A PRACTICAL FIELD GUIDE FOR UK SMALL AND MEDIUM-SIZED BUSINESSES

Where AI Actually Helps

Choosing one worthwhile AI use case, scoping a 30-day pilot and estimating the return before buying software.

AUTHOR
Ted Milton
Founder, Expedient AI

EDITION
July 2026
Version 1.0

FORMAT
A4 print edition
26 pages, worksheets included

Expedient AI Ltd

INDEPENDENT · BRISTOL · UK

BEFORE YOU START

You already have the software. Or you are about to buy it.

This guide does not start with a platform. It starts with one question about your work.

Maybe someone in the business signed up for ChatGPT Enterprise, Copilot or a vertical tool and now wants to know what to do with it. Maybe a consultant has proposed a six-figure transformation and you need a way to evaluate it without becoming an AI expert. Maybe you have read enough headlines to feel you should be doing something, but every guide starts with "choose a platform" and you do not know which one.

This guide starts with a different question: *is there a piece of work in your business that is repeated often enough, structured enough and safe enough to test?*

If the answer is yes, the next pages will help you scope a 30-day pilot and estimate the return before you spend anything. If the answer is no — or not yet — this guide will tell you that too, and suggest what to do instead. Stopping is a valid outcome.

Who this is for

- owners and managers of UK small and medium-sized businesses;
- operations, finance, HR or client-service leads who own a repeated process;
- anyone evaluating an AI proposal and needing a structured way to say yes, no or not yet;
- teams that already have a tool and want to find one worthwhile use for it.

Who this is not for

- organisations that need autonomous decision-making, regulated advice, clinical judgement or safety-critical control — those require specialist procurement;
- anyone looking for a prompt library or a list of chatbot tricks — though one page covers the basics of what to tell the system;
- anyone who needs legal, privacy, security or employment conclusions — this guide gives operational prompts, not advice.

> **The test is simple.** Can you name one repeated task, show five real examples of its input and output, and identify the person who checks the result? If yes, you have a candidate. The rest of this guide helps you decide whether to test it.

WHAT YOU WILL LEAVE WITH

A short, useful toolkit, not a software purchase.

Five decisions, seven worksheets, three worked examples, one Monday plan.

One shortlisted use case. One 30-day pilot canvas.
An honest return range. A human-review checklist.
A practical Monday plan.

What this guide is

A printable field guide for UK small and medium-sized businesses that want to choose one worthwhile AI use case, run a 30-day pilot and estimate the return before committing to a software purchase. The structure follows a single page per decision so you can work through it with the process owner in the room.

How to use it

Set aside 60 to 90 minutes for the first pass. Print or photocopy the worksheets before you start. Include the person who does the work and the person accountable for the result.

Contents

1 THE CORE RULE

Start with the work, not the software.

Five decisions, in order. No platform choice before the use case is chosen.

AI can help a small or medium-sized business. It can also add cost, risk and another half-used subscription. The difference usually starts with how the opportunity is chosen.

A worthwhile use case has a clear job, a sensible input, a useful output and an accountable owner. It saves time, improves consistency, reduces delay or helps staff make better decisions. Its errors can be spotted before they cause harm.

This guide helps you find that kind of use case. It does not ask you to choose a platform first. It asks you to examine real work.

> **The core rule:** automate preparation before judgement. Let AI draft, extract, classify, compare or summarise. Keep a named person responsible for checking important outputs and approving consequential actions.

This rule is especially important where personal data, money, contracts, safety, employment or customer commitments are involved.

The five decisions

Work through them in order. Each one narrows the options before the next.

Are the business and process ready?

Owner, inputs, test data and a stop rule must already exist.

Which repeated tasks deserve attention?

List processes, not departments.

Which single opportunity scores best?

Use five factors and the same arithmetic.

Can it be tested safely in 30 days?

Define success and stop measures before the test.

Is the likely benefit worth the full cost?

Build a low, expected and high return range.

> CAUTION Do not try to redesign the whole business. Choose one bounded problem and obtain evidence.

2 SEQUENCE AND ROOM

How to use this toolkit.

60 to 90 minutes, three viewpoints, evidence over impressions.

Set aside 60 to 90 minutes for the first pass. Include the person who does the work and the person accountable for the result.

Suggested sequence

First 15 minutes: check readiness. Complete the readiness worksheet on page 7. Stop if the process has no owner, no stable inputs or no safe test data.

Next 25 minutes: inventory the work. List repeated processes on page 9, not vague departments. “Prepare first draft of weekly debtor emails” is useful. “Use AI in finance” is not.

Next 20 minutes: score opportunities. Score up to five candidates on page 11 using the same five factors. Show the arithmetic. Discuss large scoring differences rather than averaging them silently.

Final 30 minutes: scope one pilot. Complete the pilot canvas on page 16, the ROI range on page 18 and the review checklist on page 20. Give each action an owner and a date.

Who should be in the room

For a small pilot, three viewpoints are usually enough:

- the process owner, who knows the real work;
- the accountable manager, who owns quality and risk;
- a data or systems contact, who knows access and storage.

One person may cover two roles. Do not omit the employee who performs the task.

Evidence beats enthusiasm

During the pilot, retain a small comparison set. Record the original input, the AI-assisted output, the final approved output, the time spent and the corrections made.

Do not rely on impressions such as “it felt quicker”. Use the same measures before and during the test.

> RULE If the pilot cannot be measured, it cannot be decided.

The person who does the work

AI pilots fail more often from silence than from errors. The person who currently performs the task knows where the exceptions live, which shortcuts are safe and which outputs matter most. If they are not in the room, the pilot will miss the cases that cause harm.

Include them from day one — not as a subject of the pilot, but as a co-designer. Their job changes; it does not disappear. A pilot that removes a task without addressing what happens to the person who did it will produce resistance, workarounds and hidden risk. Name the change explicitly: this task moves from manual preparation to AI-assisted preparation with human review. The reviewer role replaces the drafter role. Training and review time are included in the plan.

3 SCOPE AND LIMITS

What is included, and what is not.

Operational scope, legal limits and the UK/EU regulatory floor.

Included

This toolkit covers early operational use cases such as:

- drafting routine internal or customer-facing text;
- summarising calls, documents or case notes;
- extracting fields from consistent documents;
- classifying requests into agreed categories;
- comparing information against a checklist;
- preparing options for a person to review;
- retrieving approved internal guidance;
- assisting with repetitive administrative steps.

It also covers basic data handling, pilot controls, return estimates and human review.

Not included

This is not legal, regulatory, tax, employment or information-security advice. It is not a procurement specification. It does not certify that a particular system is safe or compliant.

This guide does not cover autonomous decisions about people, safety-critical control, clinical decisions, regulated financial advice or unsupervised contractual commitments.

> **SPECIALIST REVIEW** Mistakes that could materially affect rights, safety, employment, credit, housing, insurance or access to essential services require specialist review.

Important UK and EU context

The UK does not currently have one horizontal AI Act. Existing law still applies. For personal data, the UK GDPR and the Data Protection Act 2018 form the floor.

The Information Commissioner's Office publishes AI and data-protection guidance. The ICO has stated that parts are under review following the Data (Use and Access) Act 2025. Check the current position before relying on it.

The National Cyber Security Centre publishes guidance for secure AI system development. Its lifecycle approach covers design, development, deployment, operation and maintenance.

If your organisation has relevant EU exposure, consider the EU AI Act. Article 4 AI literacy duties have applied since 2 February 2025. Article 50 transparency duties and standalone high-risk obligations (Annex III) apply from 2 August 2026. High-risk AI embedded in regulated products (Annex I) applies from 2 August 2027. General-purpose AI model obligations have been in force since 2 August 2025 — for an SME using tools such as ChatGPT or Copilot, the GPAI and transparency rules are the relevant ones; the heavier obligations fall on the model provider. An EU digital omnibus proposal may adjust some Annex I dates — check the current position before relying on a specific deadline.

This guide gives operational prompts, not a legal conclusion.

4 READINESS CHECK

Readiness before opportunity.

Score each statement 0 to 2. Ten statements, twenty points.

A promising task can still make a poor pilot. Use the following check before scoring opportunities.

Scoring scale

- **0 =** No or unknown.
- **1 =** Partly.
- **2 =** Yes, with evidence.

NO.	READINESS STATEMENT	SCORE 0–2	EVIDENCE OR ACTION NEEDED	OWNER	DUE DATE
1	The process has a named owner.				
2	The start and end of the process are clear.				
3	Inputs are available and reasonably consistent.				
4	A good output can be described or shown.				
5	Current volume and handling time can be sampled.				
6	Common errors and exceptions are known.				
7	Data can be classified before testing.				
8	Safe test data or an approved environment exists.				
9	A person can review outputs during the pilot.				
10	Success and stop measures can be agreed.				
Total		**/20**	**See interpretation on page 8**		

Four hard stops

Do not begin a live-data pilot if any of these statements is true:

1. Nobody is accountable for the output.
2. You cannot explain what data enters the system or where it goes.
3. Errors could cause serious harm before a person can intervene.
4. You cannot stop the test and return to the current process.

> A low-risk offline test may still be possible with synthetic or properly anonymised material.

4 INTERPRET THE SCORE

Interpret the score.

Use the bands below as a decision aid, not a verdict.

16–20	11–15	0–10	Hard stop
READY TO SCOPE	PREPARE FIRST	DO NOT PILOT YET	PAUSE REGARDLESS OF SCORE
Proceed, but close any gap that affects data, security or accountability before live use.	The use case may be sound. Spend several days collecting samples, naming owners and defining quality.	The process is not sufficiently understood or controlled. Improve the process before adding AI.	Any single hard stop from page 7 must be cleared before a live pilot, even with a high readiness score.

Readiness notes

Capture what must be true before the pilot starts, the evidence that proves it, and the person who can stop the test.

WHAT MUST BE TRUE BEFORE WE TEST?

WHAT EVIDENCE WILL SHOW IT IS TRUE?

WHO CAN STOP THE PILOT?

What to do next

If you scored 16 or more and there are no hard stops, move to the process inventory on page 9.

If you scored 11 to 15, agree owners and a one-week preparation plan before continuing.

If you scored 10 or less, stop here. Improve the process or the data before adding AI.

If any hard stop applies, do not begin a live-data pilot. An offline test may still be possible.

> REMINDER A high readiness score does not override a hard stop. A high score with unsafe data remains unsafe.

5 INVENTORY REPEATED WORK

Build a process inventory.

Begin with work that happens repeatedly, not with every possible AI feature.

Walk through one typical week. Look at inboxes, handovers, spreadsheets, forms, calls, quotes, reports and approvals. Useful candidates often contain one or more of these patterns:

- staff repeatedly retype or reformat information;
- requests wait because somebody must triage them;
- first drafts follow a stable structure;
- information must be found across approved sources;
- documents must be checked against a known list;
- routine work crowds out higher-value judgement;
- demand varies sharply, causing backlogs;
- avoidable omissions create rework.

> **Avoid starting with a process only because it is disliked.** It must also be measurable, bounded and safe enough to test.

Describe the unit of work

A score is only useful if everyone scores the same unit. Complete this sentence:

```
When [trigger] occurs, [role] uses [inputs] to produce [output], which is checked by [reviewer] before [next action].
```

OUR UNIT OF WORK

Establish a baseline sample

Choose a recent, representative sample. Twenty units may be enough for an early operational test. Use more where variation or risk is high. Record low, typical and high values for each measure.

BASELINE MEASURE	LOW	TYPICAL	HIGH	SOURCE OR SAMPLE PERIOD
Units per month				
Minutes per unit				
Rework minutes per unit				
Error or correction rate				
Waiting time				
Direct cost per unit				

5 PROCESS INVENTORY WORKSHEET

Process inventory worksheet.

List five to ten repeated processes. Use observed ranges where exact figures are unavailable.

PROCESS OR TASK	TRIGGER	INPUT	OUTPUT	VOL / MONTH	MIN EACH	MAIN PAIN	CURRENT OWNER	DATA TIER	ERRORS OR EXCEPTIONS

Selecting from the inventory

From this list, pick up to five candidates to score on the next page. Prefer candidates that are frequent, time-consuming, structured and bounded. Do not score more than five at once; the arithmetic loses meaning beyond that.

PICK IF

You can name the owner, show five real examples and explain what a good output looks like.

DEFER IF

The process is unclear, ownerless or produces novel judgement each time.

EXCLUDE IF

Restricted data, regulated decisions or autonomous actions are involved.

6 FIVE FACTORS, WEIGHTED ARITHMETIC

Score opportunities using five factors.

Score no more than five candidates at once. Each factor receives 1 to 5 points.

Use the same five factors for every candidate. Show the arithmetic. Discuss large scoring differences rather than averaging them silently.

The five factors

A **Frequency.** How often does the task occur?

- **1:** less than monthly **2:** monthly **3:** weekly **4:** several times a week **5:** daily or many times daily.

B **Time burden.** How much total staff time does it consume?

- **1:** under 2 staff-hours per month **2:** 2–5 hours **3:** 6–15 hours **4:** 16–40 hours **5:** over 40 hours.

C **Standardisation.** How repeatable are the inputs, steps and acceptable outputs?

- **1:** mostly novel judgement **2:** limited recurring structure **3:** a common path with several exceptions **4:** mostly consistent with known exceptions **5:** highly consistent and rules can be stated.

D **Value of improvement.** What would a better process achieve?

- **1:** minor convenience **2:** modest local benefit **3:** useful time, quality or response improvement **4:** material operational or customer benefit **5:** major measurable benefit tied to a business priority.

E **Control and safety.** Can errors be detected and contained before harm?

- **1:** errors may cause serious or irreversible harm **2:** weak detection or difficult recovery **3:** review is possible, but exceptions are significant **4:** clear review, permissions and rollback **5:** low-consequence output with easy checking and recovery.

6 THE ARITHMETIC

The arithmetic and the worksheet.

Value and control receive double weight. Show your workings.

The weighted formula

```
Opportunity score = Frequency + Time burden + Standardisation + (Value × 2) +
(Control × 2)
```

A busy task is not worthwhile if the benefit is trivial or the risk cannot be controlled. The minimum score is 7. The maximum is 35.

> **Why value and control carry double weight.** A task that happens 100 times a day but produces trivial output is not worth automating — the benefit ceiling is too low regardless of frequency. A task with serious error consequences and no review path is dangerous to automate regardless of how often it occurs. Doubling these two factors ensures that high frequency alone cannot outscore real business value and safe controls.

Bridging the two scores. The readiness score (out of 20) asks: is the process understood and controlled enough to test? The opportunity score (out of 35) asks: is this task worth testing? A process can score 18/20 on readiness but 12/35 on opportunity — well-run, but not worth the effort. Or 9/20 on readiness but 30/35 on opportunity — high value, but the process needs work first. Both scores must clear their thresholds independently.

Worked scoring calculation

Suppose a task scores: Frequency 4 · Time burden 3 · Standardisation 4 · Value 4 · Control 5.

```
4 + 3 + 4 + (4 × 2) + (5 × 2) = 29 / 35
```

Decision bands

29–35	23–28	17–22	7–16
STRONG PILOT	PROMISING	REDESIGN FIRST	DO NOT PRIORITISE
Subject to readiness and data checks.	Resolve the weakest factor before testing.	Gather evidence or redesign the process.	Defer for now; revisit only if circumstances change.

Opportunity scoring worksheet

CANDIDATE	FREQUENCY 1–5	TIME 1–5	STANDARD 1–5	VALUE 1–5	CONTROL 1–5	CALCULATION	TOTAL /35	WEAKEST FACTOR
A						F + T + S + 2V + 2C		
B						F + T + S + 2V + 2C		
C						F + T + S + 2V + 2C		
D						F + T + S + 2V + 2C		
E						F + T + S + 2V + 2C		

> **IMPORTANT** A score does not override a hard stop. A high score with unsafe data remains unsafe.

7 FOUR TIERS OF DATA

Classify data before it enters AI.

Apply the highest relevant tier to the whole input.

Data classification turns a vague warning into a working rule. Removing a name does not always make data anonymous. A person may remain identifiable from combinations of details.

Four-tier classification

TIER 1 · PUBLIC

Information approved for anyone to see. Examples: published web pages, public brochures, public job adverts and approved press material.

TIER 2 · INTERNAL

Routine business information not intended for public release. Disclosure would cause limited harm. Examples: internal procedures, generic meeting notes, draft schedules and non-sensitive templates.

TIER 3 · CONFIDENTIAL

Commercially sensitive information or ordinary personal data. Disclosure could harm a person, customer or business. Examples: customer contact details, quotations, contracts, supplier terms, identifiable call notes and unpublished financial information.

TIER 4 · RESTRICTED

Highly sensitive, regulated, privileged or security-critical data. Exposure could cause serious harm. Examples: special category personal data, criminal-offence data, authentication secrets, bank details, detailed employee cases and security configurations.

> Do not infer privacy from a familiar brand name. Verify the terms and controls for the exact account and service configuration.

7 SAFE-USE MATRIX

Safe-use matrix.

An approved environment is one your organisation has assessed for service, contract, access, retention, data use and deletion.

Use the matrix below to decide what may enter which environment. When in doubt, raise the tier or move to a more controlled environment.

DATA TIER	UNAPPROVED PUBLIC AI SERVICE	APPROVED BUSINESS ENVIRONMENT	ADDITIONAL CONTROLS	DEFAULT PILOT POSITION
1 Public	May be acceptable if content is genuinely public and terms are understood.	Usually acceptable within policy.	Check copyright, accuracy and publication approval.	Suitable for a first pilot.
2 Internal	Do not enter by default. Use synthetic or public substitutes.	May be acceptable with access controls and policy.	Minimise input, restrict sharing and define retention.	Proceed after owner approval.
3 Confidential	Do not enter.	Only after privacy, security and contractual assessment.	Data minimisation, lawful basis, access logging, retention limits and human review.	Prefer offline or tightly bounded testing first.
4 Restricted	Do not enter.	Specialist approval required. Some uses should remain prohibited.	Formal risk assessment, strong technical controls and expert oversight.	Exclude from an initial SME pilot.

Data-use worksheet

QUESTION	ANSWER	EVIDENCE OR CONTROL	OWNER
What exact fields enter the system?			
Which tier applies?			
Is personal data necessary?			
Can fields be removed, masked or replaced?			
What is the lawful basis for personal data use?			
Is a data protection impact assessment needed?			
Is customer or staff notice required?			
Is input used to train provider models?			
Where is data processed and stored?			
Who can access prompts, files and outputs?			
How long is data retained?			
How is data deleted or exported?			
What happens if the service is unavailable?			

7 INSTRUCTION CRAFT

What to tell the system — and what to check.

The pilot's success depends less on the model and more on the instructions.

A well-scoped use case with poor instructions will fail. A modest use case with clear instructions, good examples and defined review criteria will often succeed. Write the instructions before the pilot starts, not after the first bad output.

The six-part instruction

For each AI-assisted step, write instructions that cover six things:

Role. What the system is doing. "You are preparing a first draft of a property listing from approved particulars."

Input. What it receives. "You will receive: the property address, tenure, price, room count, EPC rating and the agent's notes."

Output. What it produces. "Produce a structured draft with: headline, summary paragraph, room-by-room description, local area note and required disclosures."

Examples. One or two examples of acceptable output. Even a short example anchors quality more than a page of rules.

Prohibitions. What it must not do. "Do not invent descriptions of condition. Do not add amenities not in the source. Do not state prices not in the approved particulars."

Flags. What it should surface for human attention. "If any required field is missing from the input, list it under 'Missing information' and do not guess."

Review criteria

Write the review criteria before the pilot starts. The reviewer should be able to check each output against a list in under two minutes. If the list takes longer than the task would have taken without AI, the instructions need work.

A useful test: give the instructions and one input to a colleague who has not seen the task before. If they can produce an output that passes the review criteria, the instructions are clear enough for the system.

COMMON FAILURE Instructions that say "produce a good summary" without defining what good means. Replace with: "the summary must contain the client name, the matter reference, the decision reached, the next action and its owner. Omit opinion and speculation."

8 PILOT CANVAS

The 30-day pilot canvas.

A pilot is a time-limited test of a stated claim. It is not a quiet production launch.

Keep the first pilot narrow. One process, one team, one owner and a limited data set are usually enough.

CANVAS FIELD	YOUR ANSWER
Use case name	
Process owner	
Accountable approver	
Pilot dates	Start: End:
Problem today	
Unit of work	
Users included	
Users excluded	
Inputs included	
Inputs excluded	
Data tier	
AI-assisted step	
Human step retained	
Expected output	
Baseline sample	
Success measures	
Stop measures	
Review frequency	
Fallback process	
Decision on day 30	Stop / extend / redesign / adopt

Write the testable claim

```
For [defined users], AI assistance will reduce [measured burden] from [baseline] to
[target range], while keeping [quality measure] above [threshold].
```

OUR CLAIM

8 PRACTICAL 30-DAY RHYTHM

A practical 30-day rhythm.

Days 1–5 baseline · Days 6–10 offline · Days 11–20 limited live · Days 21–27 expand · Days 28–30 decide.

Days 1–5 · Baseline and controls

- confirm the unit of work;
- sample current time and quality;
- classify the data;
- agree examples of acceptable output;
- define prohibited inputs and actions;
- prepare a fallback process;
- brief every participant.

Days 6–10 · Offline test

- use synthetic, public or approved historical inputs;
- compare results with known good outputs;
- document common failure modes;
- revise instructions and review criteria;
- stop if a hard boundary cannot be enforced.

Days 11–20 · Limited live test

- use a small portion of eligible work;
- require review before use;
- record time, corrections and exceptions;
- hold brief twice-weekly reviews;
- retain evidence needed for the final decision.

Days 21–27 · Controlled expansion

- expand only if measures remain within limits;
- test unusual but permitted cases;
- confirm access and deletion controls;
- ask users where hidden work has increased;
- calculate the low, expected and high return.

Days 28–30 · Decision

- compare pilot and baseline evidence;
- include setup, review and exception costs;
- review incidents and near misses;
- decide to stop, extend, redesign or adopt;
- assign controls for any continued use.

Pilot measure sheet

MEASURE	BASELINE	TARGET RANGE	PILOT RESULT	EVIDENCE SOURCE	PASS?
Minutes per unit					
Review minutes per unit					
Correction rate					
Material error count					
Waiting time					
User acceptance					
Customer impact, if measured					

9 FIVE-STEP RETURN RANGE

Estimate an honest return range.

A return estimate is not a promise. It is a decision aid built from visible assumptions.

Calculate low, expected and high cases. Use measured time where possible. If a number is uncertain, widen the range.

Step 1 · Gross hours saved

```
Gross hours saved per month = monthly volume × minutes saved per unit ÷ 60
```

```
Minutes saved per unit = current minutes − assisted minutes
```

Assisted minutes must include prompting, checking, corrections and handling failed outputs.

Step 2 · Usable hours

```
Usable hours = gross hours saved × realisation rate
```

The realisation rate is the share of saved time that can be redirected. Fragmented five-minute savings may have a lower rate than removing a two-hour weekly backlog.

Realisation rate — a rule of thumb

PATTERN	TYPICAL RATE	WHY
Removes a continuous block of work (e.g. a two-hour weekly backlog)	70–85%	Freed time is contiguous and redirectable.
Speeds up a task that still happens at the same frequency	40–60%	The person still sits down to do the task; the saving is marginal per instance.
Reduces errors or rework	30–50%	Benefit is real but indirect; measure it separately as avoided cost.
Fragmented micro-savings (30 seconds here, 2 minutes there)	10–30%	Too scattered to redirect; often absorbed by context-switching.

If you cannot place your use case in one of these rows, use the lowest rate that fits and widen the range.

Step 3 · Monthly benefit

```
Capacity benefit = usable hours × loaded hourly cost
```

Loaded hourly cost may include salary, employer costs and relevant overhead. Document what is included.

```
Total monthly benefit = capacity benefit + evidenced avoided cost + evidenced contribution gain
```

Do not count the same benefit twice. Time released is not automatically extra revenue.

> CAUTION If you cannot evidence an avoided cost or contribution gain, do not include it in the calculation.

9 WORKSHEET

ROI worksheet and decision questions.

Complete steps 4 and 5 on this page after the range above.

Step 4 · Monthly and setup costs

Include service and usage costs, implementation or configuration, staff training, review and governance time, integration and support, exception handling, security, privacy or legal review, and switching and exit work.

```
Monthly equivalent setup cost = one-off setup cost ÷ allocation months
```

```
Total monthly cost = recurring cost + monthly equivalent setup cost + added labour cost
```

Step 5 · Net benefit and return

```
Monthly net benefit = total monthly benefit − total monthly cost
```

```
ROI percentage = monthly net benefit ÷ total monthly cost × 100
```

```
Payback months = one-off setup cost ÷ monthly net benefit before setup allocation
```

Only calculate payback when monthly net benefit is positive.

ROI range worksheet

ASSUMPTION	LOW	EXPECTED	HIGH	EVIDENCE OR REASON
Monthly volume				
Current minutes per unit				
Assisted minutes per unit				
Minutes saved per unit				
Gross hours saved				
Realisation rate				
Usable hours				
Loaded hourly cost				
Capacity benefit				
Other evidenced benefit				
Total monthly benefit				
Recurring monthly cost				
One-off setup cost				
Allocation months				
Monthly setup equivalent				
Added labour cost				
Total monthly cost				
Monthly net benefit				
ROI percentage				

Decision questions

10 REVIEW AND PERMISSIONS

Human review and agent permissions.

Review must be a defined control, not a phrase in a policy.

A reviewer needs enough time, information and authority to reject the output. Requiring a click without meaningful examination is not review.

Human-review checklist

Before the pilot

- A named person is accountable for each output type.
- Review criteria are written in plain language.
- Reviewers have examples of acceptable and unacceptable outputs.
- Review time is included in the return estimate.
- Reviewers can see the source material.
- Reviewers can edit, reject or escalate.
- High-impact outputs receive stronger review.
- The system identifies uncertainty or missing information where possible.

For each material output

- Facts match the source.
- Names, dates, figures and units are correct.
- Required information is present.
- Unsupported claims have been removed.
- Tone suits the audience and situation.
- Personal or confidential data is handled correctly.
- The output follows policy and legal requirements.
- Exceptions have been escalated.
- The approving person is recorded.

During operation

- Corrections and failures are logged.
- Samples are reviewed for drift.
- Access is reviewed when roles change.
- Instructions and reference material are versioned.
- Users know how to report an incident.
- The fallback process is tested.

Permission ladder

An AI agent may do more than draft text. It may browse systems, call services, update records or send messages. Start at the lowest sufficient level.

1 **Suggest.**
The system proposes an action.

2 **Draft.**
It prepares the action for review.

3 **Stage.**
It creates a reversible change in a controlled area.

4 **Execute with confirmation.**
A person approves each consequential action.

5 **Execute within limits.**
Only after evidence supports bounded autonomy.

11 WORKED EXAMPLES

Example A · Fabricator RFQ preparation.

A small metal fabricator drafts RFQ summaries. Time savings alone are weak.

EXAMPLE A

Fabricator · Requests for quotation

Situation

A small metal fabricator receives requests for quotation by email. An estimator reads attachments, extracts requirements and prepares a structured RFQ summary.

The proposed use is to prepare a first-pass summary and missing-information list. A qualified estimator checks all fields and makes the commercial decision.

Visible assumptions

- 60 RFQs per month
- current preparation time: 35 minutes each
- assisted preparation and review: 22 to 29 minutes each
- low, expected and high time saved: 6, 10 and 13 minutes
- realisation rates: 50%, 65% and 75%
- illustrative loaded hourly cost: £32
- illustrative total monthly cost: £260

Arithmetic

CALCULATION	LOW	EXPECTED	HIGH
Gross hours saved	60 × 6 ÷ 60 = 6.0	60 × 10 ÷ 60 = 10.0	60 × 13 ÷ 60 = 13.0
Usable hours	6.0 × 50% = 3.0	10.0 × 65% = 6.5	13.0 × 75% = 9.75
Capacity benefit	3.0 × £32 = £96	6.5 × £32 = £208	9.75 × £32 = £312
Monthly net benefit	**£96 – £260 = –£164**	**£208 – £260 = –£52**	**£312 – £260 = £52**

Honest interpretation

On time savings alone, this pilot is weak. The expected case is negative. The business should not invent revenue to improve the result.

It may still test the process if missing-information detection has measurable value. That value must be evidenced separately.

Controls

- exclude restricted technical or defence-related work;
- use an approved environment for confidential drawings;
- require estimator review against original attachments;
- prohibit automatic pricing or customer commitments;
- record missed specifications and false extractions.

Possible decision: proceed only as a small offline test. Adopt only if quality evidence or lower costs change the low and expected cases.

11 WORKED EXAMPLES

Example B · Estate agent listing drafts.

An estate agency drafts listings from approved particulars. Range crosses zero.

EXAMPLE B

Estate agent · Listing drafts

Situation

An estate agency prepares property listing drafts from approved particulars and staff notes. The proposed use creates a structured first draft. A named agent checks every fact before publication.

Visible assumptions

- 45 listings per month
- current drafting time: 42 minutes each
- assisted drafting, fact check and edit: 20 to 30 minutes
- low, expected and high time saved: 12, 17 and 22 minutes
- realisation rates: 55%, 70% and 80%
- illustrative loaded hourly cost: £28
- illustrative total monthly cost: £180

CALCULATION	LOW	EXPECTED	HIGH
Gross hours saved	45 × 12 ÷ 60 = 9.0	45 × 17 ÷ 60 = 12.75	45 × 22 ÷ 60 = 16.5
Usable hours	9.0 × 55% = 4.95	12.75 × 70% = 8.925	16.5 × 80% = 13.2
Capacity benefit	4.95 × £28 = £138.60	8.925 × £28 = £249.90	13.2 × £28 = £369.60
Monthly net benefit	**£138.60 – £180 = –£41.40**	**£249.90 – £180 = £69.90**	**£369.60 – £180 = £189.60**
ROI	**–£41.40 ÷ £180 = –23%**	**£69.90 ÷ £180 = 39%**	**£189.60 ÷ £180 = 105%**

Honest interpretation

The range crosses zero. The decision depends heavily on review time and usable capacity. A pilot should measure both carefully.

Controls

- use only approved property facts and images;
- check tenure, price, room details and required disclosures;
- prohibit invented descriptions of condition or local amenities;
- keep publication approval with the responsible agent;
- retain a source-to-draft checklist.

Possible decision: run a 30-day bounded pilot. Continue only if the expected case is supported and factual correction rates stay below the agreed limit.

12 EXAMPLE C, MONDAY PLAN, PATHS

EXAMPLE C

Accountancy · Call summaries

Situation

An accountancy practice prepares internal summaries after routine client calls. The proposed use drafts a summary and action list from an approved transcript. The accountant checks the summary before it enters the client record. The system does not provide tax advice or contact the client.

Visible assumptions

- 90 eligible calls per month
- current note preparation: 16 minutes each
- assisted generation and review: 7 to 12 minutes
- low, expected and high time saved: 4, 7 and 9 minutes
- realisation rates: 60%, 75% and 85%
- illustrative loaded hourly cost: £38
- illustrative total monthly cost: £210

CALCULATION	LOW	EXPECTED	HIGH
Gross hours saved	90 × 4 ÷ 60 = 6.0	90 × 7 ÷ 60 = 10.5	90 × 9 ÷ 60 = 13.5
Usable hours	6.0 × 60% = 3.6	10.5 × 75% = 7.875	13.5 × 85% = 11.475
Capacity benefit	3.6 × £38 = £136.80	7.875 × £38 = £299.25	11.475 × £38 = £436.05
Monthly net benefit	**£136.80 – £210 = –£73.20**	**£299.25 – £210 = £89.25**	**£436.05 – £210 = £226.05**
ROI	**–£73.20 ÷ £210 = –35%**	**£89.25 ÷ £210 = 43%**	**£226.05 ÷ £210 = 108%**

The expected and high cases are positive, but the low case is negative. Consent, confidentiality and review may add further cost. Pilot with a limited call category. Include consent or notice administration, transcript checks and correction time in the final calculation.

Your Monday plan

A good next step fits into the working week. It does not require a large technology decision.

Monday · 45 minutes

- Name one process owner.
- Book a 45-minute process walk-through.
- Print or copy the readiness and inventory worksheets.
- Choose a recent sample period.
- Identify where source documents are stored.

Tuesday · 60 minutes

- Map five repeated tasks.
- Record volume and time ranges.
- Mark data tiers.
- Note exceptions and failure consequences.

Wednesday · 45 minutes

- Score up to five opportunities.
- Check the arithmetic.
- Select one candidate.
- Record why the other candidates are waiting.

Thursday · 60 minutes

- Complete the pilot canvas.
- Define success and stop measures.
- Write review criteria.
- Set data and permission boundaries.
- Confirm the fallback process.

11 WORKED EXAMPLES

Example D · Stopped at the data gate.

A high score does not override a hard stop. The most valuable outcome was the decision not to proceed.

EXAMPLE D

Recruitment · CV screening

Situation

A recruitment agency wants to auto-screen CVs against job descriptions and produce a shortlist ranking. The proposed use would parse each CV, extract qualifications and experience, and score candidates against the role requirements.

Scores

- readiness: 14/20 — process has an owner and consistent inputs, but output criteria ("good fit") vary by consultant
- opportunity: 26/35 — high frequency, high time burden, good input standardisation
- data tier: **Tier 4 — Restricted**

Why it stopped

CVs contain special category personal data (ethnicity indicators, disability disclosures, age, gender), criminal-offence data and contact details. The proposed use involves automated decision-making about individuals. This places the task outside the scope of an SME pilot.

Decision

Stopped. The agency should obtain specialist legal and privacy advice before proceeding. A bounded offline test using fully synthetic CVs may validate the extraction logic, but live screening requires a data protection impact assessment and likely human-only decision-making under UK GDPR Article 22.

Lesson

The most valuable outcome of this exercise was the decision not to proceed — reached in 90 minutes with a worksheet, rather than after a six-month procurement. The opportunity score was strong. The data tier and the automated-decision-making element were the hard stop.

Not every exercise ends in a pilot. Three of the four examples in this guide produce a positive expected case. This one does not. All four are useful outcomes. The worksheet exists to make the decision visible, not to produce a predetermined answer.

12 REFERENCES AND CLOSE

Source notes

These notes support the legal, security and governance context. They are not a substitute for advice on your circumstances. Pages and guidance can change.

1. **Information Commissioner's Office,** Guidance on AI and data protection — ico.org.uk
2. **National Cyber Security Centre,** Guidelines for secure AI system development — ncsc.gov.uk
3. **European Commission,** AI literacy questions and answers — digital-strategy.ec.europa.eu
4. **European Commission AI Act Service Desk,** implementation timeline — ai-act-service-desk.ec.europa.eu
5. **OWASP,** AI Agent Security Cheat Sheet — cheatsheetseries.owasp.org
6. **OpenAI,** Enterprise privacy — openai.com
7. **Microsoft,** Enterprise data protection — learn.microsoft.com

Provider privacy pages are included as examples of the questions organisations should verify. Their inclusion is not a product recommendation.

Quick glossary

Realisation rate — The share of saved time that can actually be redirected to other work. Not all saved time is usable.

Loaded hourly cost — The fully loaded cost of a staff hour, including salary, employer contributions and relevant overhead.

Unit of work — The smallest repeatable piece of work the AI assists with. "Prepare one RFQ summary" is a unit. "Do the finance work" is not.

Hard stop — A condition that prevents a live-data pilot regardless of how high the readiness or opportunity score is.

Data tier — A four-level classification (Public, Internal, Confidential, Restricted) that determines what may enter which AI environment.

Permission ladder — The five levels of agent autonomy, from suggest to execute within limits. Most first pilots stay at levels one or two.

Reusable shortlist summary

OUR SHORTLISTED USE CASE

OUR OPPORTUNITY SCORE

OUR DATA TIER

OUR 30-DAY CLAIM

OUR EXPECTED MONTHLY NET BENEFIT

OUR LOW-TO-HIGH RANGE

THE HUMAN WHO APPROVES MATERIAL OUTPUTS

THE ACTION WE

Colophon

PUBLICATION

Where AI Actually Helps

Edition 01 · July 2026 · Version 1.0

Author: Ted Milton, Expedient AI Ltd

CONTACT

Expedient AI Ltd

expedientai.co.uk

> You can use this toolkit without contacting Expedient AI. If you would value an independent review of your shortlist or pilot canvas, Expedient AI Ltd offers practical AI scoping support for UK SMEs.