A REUSABLE METHOD FOR RESPONSIBLE AI ADOPTION

AI Delegation Guardrail Framework

Deciding where AI can automate work, where it can support people, and where human judgment and accountability must remain in control.

Automate | Assist | Never delegate

Greg Hoskins
Sr. Manager, UX Research

BEFORE YOU BEGIN

Foreword

This framework began with a practical question for my UX Research team: Which parts of our work should AI automate, where should it assist, and what should we never delegate?

To answer it, I designed a workshop grounded in the actual work of UX Research. We mapped the thirteen stages of our workflow and identified the individual tasks within each stage. We sorted each task into Automate, Assist, or Never Delegate, placing debatable tasks in a Gray Area for further discussion. We then began turning those classifications into guardrail statements that defined what AI could do, the conditions for its use, and what responsibility remained with the researcher.

The AI Delegation Guardrail Framework generalizes this process so any team can apply it to its own work. Teams map their workflow, identify its tasks, classify each task by the appropriate level of AI delegation, and write guardrails that reflect their responsibilities, risks, and context.

The resulting guardrails are intended first as a strategic framework. Together, they express the team's position on how AI should, and should not, participate in its work. They identify where AI can create value, where human review and judgment remain essential, and where accountability must never be left to an AI system. In this way, the guardrails become the basis of the team's AI strategy.

The operational layer is the logical next step. Teams can examine each Automate and Assist guardrail and brainstorm specific AI solutions for those tasks. They can evaluate tools and redesign workflows while preserving the conditions established in the guardrails. Never Delegate statements serve as firm boundaries so operational solutions do not mistakingly transfer decision authority to AI.

The progression is deliberate:

Map the work → **Classify the tasks** → **Establish the guardrails** → **Define the strategy** → **Design and test AI solutions**

AI tools and capabilities will continue to change. This framework provides something more durable; a way for teams to decide how AI should participate in their work while protecting the judgment, trust, quality, and accountability on which that work depends.

IN THIS FRAMEWORK

Contents

01 **Purpose**
What the framework helps a team do

02 **The three delegation levels**
Automate, assist, never delegate

03 **The classification method**
Six dimensions and routing rules

04 **Guardrails are permissions plus conditions**
Statement templates and examples

05 **Team workshop**
Three activities, suggested agendas, and follow-up

06 **Required outputs**
Four working artifacts

07 **Cross-functional examples**
Starting points by function

08 **Reclassification**
When and how a task moves between levels

09 **Framework principles**
Ten rules to carry forward

CHAPTER 01

Purpose

This framework helps a team:

- Adopt AI without defaulting to either blanket resistance or blind adoption
- Make decisions about specific tasks rather than entire jobs or broad capabilities
- Accelerate low risk work without weakening quality, trust, ethics, or accountability
- Turn abstract AI principles into guardrails that can be embedded in everyday workflows
- Maintain a shared, defensible position as AI capabilities evolve

THE FRAMEWORK ANSWERS ONE QUESTION

What can AI automate, where can it assist, and what must people never delegate?

CHAPTER 02

The three delegation levels

LEVEL 1

Automate

AI may perform most or all of the task with routine oversight.

Appropriate when the work is

- Repetitive or mechanical
- Governed by stable rules
- Low consequence if an error occurs
- Easy to verify
- Reversible or recoverable
- Primarily organizing, transforming, or transporting information rather than interpreting its meaning

Human responsibility

- Define the task and acceptable result
- Establish data and tool restrictions
- Specify exceptions that must be escalated
- Periodically check performance and error rates

KEY QUESTION

Would it be acceptable for AI to handle most of this task as long as exceptions were escalated?

LEVEL 2

Assist

AI may create a draft, surface options, organize evidence, or identify possible patterns. A person reviews the output and makes the decision.

Appropriate when the work

- Benefits from exploration or a faster first draft
- Requires context in order to validate
- Contains ambiguity or multiple reasonable answers
- Can affect downstream decisions
- Requires comparison with source evidence
- Involves interpretation but can be separated into an AI contribution and a human judgment

Human responsibility

- Review the actual substance, not merely approve the output
- Validate important claims against authoritative sources
- Preserve contradictions, uncertainty, and relevant context
- Revise or reject the output when necessary
- Own the conclusion and its consequences

KEY QUESTION

Can AI provide useful output without being allowed to decide whether it is correct?

LEVEL 3

Never Delegate

A person must make and own the decision. AI may provide information or possibilities, but it may not make the final call.

Appropriate when the task determines

- What evidence means
- Whether evidence is sufficient
- What level of risk is acceptable
- Whether a person should be included, excluded, approved, denied, evaluated, or disciplined
- Whether an outcome is ethical, fair, valid, or aligned with organizational values
- Which strategic direction the organization should pursue
- How to respond in a sensitive human interaction
- What the organization is willing to defend publicly

Human responsibility

- Exercise judgment
- Consider context and competing interests
- Be identifiable as the accountable decision maker
- Explain and defend the decision
- Accept responsibility for its consequences

KEY QUESTION

Would delegating this decision weaken trust, quality, fairness, or accountability?

"Never delegate" does not mean that AI can provide no input. It means AI cannot hold decision authority or validate its own work.

CHAPTER 03

The classification method

Classify individual tasks, not job titles, departments, platforms, or broad activities.

For example, “use AI for product management” is too broad. Break it into tasks such as:

Format meeting notes **Summarize customer feedback**

Draft a product requirements document **Prioritize opportunities** **Decide product strategy**

Evaluate each task across six dimensions.

Dimension	Lower delegation risk	Higher delegation risk
Consequence	Errors are minor and recoverable	Errors materially affect people, customers, strategy, or trust
Judgment	Stable rules determine the answer	Context, interpretation, or trade-offs determine the answer
Verifiability	Accuracy can be checked quickly	Correctness is subjective, delayed, or difficult to establish
Reversibility	The result can be easily corrected	The result creates lasting or difficult to reverse effects
Sensitivity	Uses public or low risk information	Uses confidential, personal, regulated, or sensitive information
Accountability	Output supports routine execution	Someone must explain and defend the decision

Routing rules

Automate

A task is a candidate for Automate only when:

- Judgment requirements are low
- Consequences are limited
- Results are readily verifiable
- Errors are reversible
- Data can be handled safely
- Exceptions can be detected and escalated

Never Delegate

A task is Never Delegate when any of the following is true:

- It assigns responsibility, opportunity, access, or consequences to a person
- It makes a final determination about validity, fairness, ethics, evidence sufficiency, or acceptable risk
- It establishes organizational strategy or an official position
- It requires accountable interpretation of ambiguous evidence
- It depends on empathy, trust, rapport, or real-time human judgment
- No qualified person can meaningfully verify the result

Assist Tasks between these boundaries default to **Assist**.

When uncertain, begin at the more restrictive level. Move a task only after evidence demonstrates that its risks are controlled.

CHAPTER 04

Guardrails are permissions plus conditions

A classification alone is incomplete. Every permitted use must include the conditions that make it acceptable.

Automate statement

AI may automatically [perform task], as long as [required controls, boundaries, and escalation conditions].

EXAMPLE

AI may automatically format and categorize approved project documents, as long as the original content is preserved, the taxonomy is predefined, and ambiguous classifications are escalated.

Assist statement

AI may help [perform task], as long as [accountable person] [reviews, validates, or decides specific things].

EXAMPLE

AI may help identify possible themes in customer feedback, as long as a product manager validates them against the source material and examines contradictions and outliers.

Never delegate statement

AI must not [make decision or exercise authority], because [human judgment, ethical responsibility, or accountability required].

EXAMPLE

AI must not decide which customer problem the organization should prioritize, because that decision requires accountable judgment across evidence, strategy, feasibility, risk, and organizational commitments.

Use precise review language

“Human in the loop” and “subject to review” are too vague. State:

- Who reviews the output
- What they must check
- What source they must use
- What decision remains theirs
- What evidence of review must be retained

CHAPTER 05

Team workshop

The workshop moves through three connected activities:

01

Map

Map the team's workflow and identify its tasks.

02

Sort

Sort each task into Automate, Assist, Never Delegate, or Gray Area.

03

Draft

Begin writing a guardrail statement for every classified task.

The output is a task level AI delegation library supported by explicit operating conditions, not a general opinion about whether the function should use AI.

Workshop preparation

Before the session, the facilitator should determine whether the team already has a shared workflow map.

If a workflow already exists

- Bring its major stages into the workshop
- Review the stages briefly with participants
- Use them as the structure for identifying tasks

If a workflow has not been mapped

- Include workflow mapping as the first workshop activity
- Ask the team to identify the major stages through which its work moves
- Keep the stages broad enough to describe the full workflow without turning every activity into its own stage

Prepare a shared workspace with:

- An area for each major workflow stage
- Individual task cards within each stage, each written as **Workflow stage: Task**
- Four sorting areas: Automate, Assist, Never Delegate, and Gray Area
- Guardrail statement templates for the three final categories

Automate Assist Never delegate Gray area

ACTIVITY 1 Map the workflow and identify its tasks

OBJECTIVE

Create a shared map of what the team actually does, organized by the major stages of its workflow.

Process

1. Identify the beginning and end of the team's workflow
2. Divide the workflow into its major stages
3. Ask participants to identify the tasks performed during each stage
4. Write each task on a separate card in the format **Workflow stage:Task**
5. Review the cards for missing, duplicate, or overly broad tasks
6. Break broad activities into smaller tasks when different parts could reasonably receive different AI classifications

For example, **"create a product strategy"** is too broad. It might include:

- **Strategy:** Compile existing customer evidence
- **Strategy:** Summarize market information
- **Strategy:** Generate possible strategic options
- **Strategy:** Evaluate trade-offs
- **Strategy:** Select the product direction

Task card format - Workflow stage:Task

Every card names the stage it belongs to, followed by the task. The stage label stays attached to the card through the sorting activity and carries into each guardrail statement, so the final library can be organized and reviewed by workflow stage.

INTAKE	DISCOVERY	STRATEGY
Log incoming requests	**Summarize customer feedback**	**Select the product direction**

ACTIVITY OUTPUT

A workflow map containing a reasonably complete inventory of task level activities.

ACTIVITY 2

Sort tasks by AI delegation level

OBJECTIVE

Classify each task according to the level of responsibility that may be delegated to AI.

Process

1. Review the definitions of Automate, Assist, and Never Delegate
2. Sort every task card into one of the three categories
3. Place any task without clear agreement into Gray Area
4. Discuss the Gray Area tasks after the initial sort rather than allowing them to stall the exercise
5. Move a Gray Area task into a final category when the team agrees on both its classification and the conditions required for that classification
6. Keep genuinely unresolved tasks in a Gray Area list for follow-up

Sort the task, not the tool

Participants should sort the task itself, not the tool currently used to perform it or the general capability of AI.

Keep the label attached

Keep the **Workflow stage:Task** label intact as cards move between sorting areas. The stage shows where each classification applies and keeps related tasks traceable when they are later grouped into guardrail statements.

Sorting questions

Automate

- Is the task repetitive, mechanical, or governed by stable rules?
- Can the output be checked quickly?
- Would an error be limited, detectable, and reversible?
- Can exceptions be identified and sent to a person?
- Would it be acceptable for AI to handle most of the task?

Assist

- Can AI provide a useful draft, option, summary, or starting point?
- Does a person need to validate the output against source information?
- Does the task contain ambiguity or require contextual interpretation?
- Can AI contribute without being allowed to make the final judgment?
- What specifically must the reviewer examine or decide?

Never Delegate

- Does the task determine what evidence means?
- Does it establish whether evidence is sufficient or trustworthy?
- Does it require ethical, strategic, relational, or professional judgment?
- Could the decision materially affect a person, customer, employee, or user group?
- Must someone be able to explain and defend the decision?
- Would delegating it weaken trust, quality, fairness, or accountability?

Gray area

Using Gray Area

Gray Area is a temporary holding category, not a fourth delegation level. It captures tasks whose classification depends on unresolved conditions or context.

For each Gray Area task, discuss:

- **Why is the classification debatable?**
- **What could go wrong if AI performed the task?**
- **Does the risk change depending on the data involved?**
- **Does the risk change depending on who is affected?**
- **Would dividing the task into smaller tasks resolve the disagreement?**
- **What human review would make AI assistance acceptable?**
- **What exception would require escalation?**
- **Does the task change categories in a higher stakes context?**
- **What additional evidence or policy is needed to classify it?**

A task may be classified differently under different conditions. Record those conditions rather than forcing a universal answer.

ACTIVITY OUTPUT

A first pass task classification organized into:

Automate | Assist | Never delegate | Gray area

ACTIVITY 3

Draft task level guardrails

OBJECTIVE

Turn each classification into an actionable statement that defines what AI may do and what conditions must remain in place.

Begin with the clearest or highest priority tasks. The workshop does not need to perfect every statement. It should establish the first working set and create a repeatable method for completing the remainder.

Head each statement with its task card, Workflow stage:Task, so guardrails can be grouped, reviewed, and maintained by workflow stage.

Automate statement

AI may automatically [perform task], as long as [boundaries, controls, and escalation conditions].

EXAMPLE · REPORTING: TAG AND FORMAT APPROVED DOCUMENTS

AI may automatically apply established tags and formatting to approved documents, tickets, and status updates, as long as it does not change their meaning and flags uncertain classifications for review.

Assist statement

AI may help [perform task], as long as [accountable person] [reviews, validates, or decides specific things].

EXAMPLE · DISCOVERY: IDENTIFY THEMES IN CUSTOMER FEEDBACK

AI may help identify possible themes in customer feedback, as long as a product manager validates them against the source material and examines contradictions and outliers.

Never Delegate statement

AI must not [make decision or exercise authority], because [human judgment, responsibility, or accountability required].

EXAMPLE · STRATEGY: PRIORITIZE CUSTOMER PROBLEMS

AI must not decide which customer problem the organization should prioritize, because that decision requires accountable judgment across customer evidence, strategy, feasibility, risk, and organizational commitments.

Guardrail quality check

Before accepting a statement, ask:

- [x] Does it describe one recognizable task?
- [x] Is the amount of authority given to AI clear?
- [x] Does it identify the condition that makes the use acceptable?
- [x] For Automate, does it define what must be escalated?
- [x] For Assist, does it say who reviews what and against which evidence?
- [x] For Never Delegate, does it explain why the decision must remain human?
- [x] Could someone use this statement to guide an actual workflow?
- [x] Would the team be comfortable defending this statement to leadership or an affected stakeholder?

Avoid vague conditions such as

- ~~"With human oversight."~~
- ~~"When appropriate."~~
- ~~"If used responsibly."~~
- ~~"Subject to review."~~

Replace them with observable responsibilities and controls.

ACTIVITY OUTPUT

An initial guardrail library containing:

- A guardrail statement for each task, stating what AI may do (Automate, Assist) or must not do (Never Delegate), organized by workflow stage
- The conditions attached to each permitted use
- The accountable role for human review or decision making
- Unresolved questions requiring follow-up

Suggested agenda

The time required depends on whether the team already has a workflow map.

OPTION A Existing workflow map · 60 minutes

Time	Activity	Output
5 min	Establish the purpose and define Automate, Assist, Never Delegate, and Gray Area	Shared language
10 min	Review workflow stages and complete the task inventory	Task map
15 min	Sort tasks into the four working areas	Initial classification
15 min	Discuss Gray Area tasks	Conditions and unresolved questions
12 min	Begin drafting guardrail statements	Initial guardrail library
3 min	Assign owners and follow-up work	Completion plan

OPTION B Workflow mapping included · 90 minutes

Time	Activity	Output
5 min	Establish the purpose	Shared objective
10 min	Define Automate, Assist, Never Delegate, and Gray Area	Shared language
20 min	Map workflow stages	End-to-end workflow
15 min	Identify tasks within each stage	Task inventory
15 min	Sort task cards	Initial classification
12 min	Discuss Gray Area tasks	Conditions and unresolved questions
10 min	Begin drafting guardrail statements	Initial guardrail library
3 min	Assign owners and next steps	Completion plan

After the workshop

The facilitator or designated owners should:

1. Consolidate duplicate task cards, keeping each in Workflow stage:Task format
2. Preserve the team's intended classifications and conditions
3. Complete and refine unfinished guardrail statements
4. Circulate the draft for team review
5. Record unresolved Gray Area tasks separately
6. Assign an owner and resolution date to each open question
7. Build the guardrails into the team's templates, checklists, approvals, and tools they already use
8. Establish a review cadence for framework updates as workflows, risks, and AI capabilities change

CHAPTER 06

Required outputs

Each team should leave the process with four working artifacts.

A

Workflow task map

A list of workflow stages and the tasks within them

B

AI delegation inventory

Every task, classified by AI delegation level

C

Guardrail library

The approved Automate, Assist, and Never Delegate statements written in clear, actionable terms

D

Gray area list

Unresolved tasks, each with an owner and a date for resolution

CHAPTER 07

Cross-functional examples

Function	Automate	Assist	Never Delegate
Product	Format status updates from approved roadmap and ticket data	Draft requirements or surface possible opportunity themes	Set product strategy or decide which customer need deserves priority
Design	Resize approved assets and check established design system rules	Generate layout alternatives or draft interface copy	Make the final judgment about whether a design is appropriate and fit for its intended users
UX Research	Clean transcript formatting and organize source material	Draft plans, cluster observations, or suggest themes	Judge whether evidence is valid, interpret what users need, or decide what the research concludes
Engineering	Apply standard code formatting and generate starter files from approved templates	Draft code, tests, or implementation options for review	Accept security risk or approve a consequential production change without accountable human review
Marketing	Reformat approved content for established channels	Draft campaign concepts or audience variations	Decide the brand's public position on a sensitive issue or approve messaging for a crisis response
People/HR	Coordinate routine scheduling and reminders	Draft communications or surface patterns from approved aggregate data	Make hiring, promotion, performance, compensation, or disciplinary decisions

These are starting points, not universal classifications. The same task may move between levels depending on its context, data, affected population, and consequences.

CHAPTER 08

Reclassification

AI capability may expand what can be automated or assisted, but improved capability alone does not remove human accountability.

↗ Toward greater delegation, only when

- Quality has been demonstrated in the actual operating context
- Errors are observable and controllable
- Review requirements remain meaningful
- Data handling is approved
- A responsible human still owns the outcome

↘ Toward less delegation, when

- Its consequences increase
- More sensitive information is introduced
- Verification becomes weaker
- People begin treating AI output as authority
- Important context, differing views, or uncertainty is repeatedly lost
- The accountable person can no longer meaningfully review the work

CHAPTER 09

Framework principles

01 Classify tasks, not jobs

02 AI may organize evidence more readily than it may decide what evidence means

03 A permission without conditions is not a guardrail

04 AI output is a draft until an accountable person validates it.

05 Human review must specify who reviews what against which source

06 AI must not validate its own analysis

07 The higher the consequence, the stronger the human control

08 Measure quality, trust, and error, not efficiency alone

09 AI capability will change, human accountability does not

10 If no one is clearly accountable for the outcome, the AI workflow is not ready