

Human Capability Safeguards for AI-Assisted Work

A design worksheet for deciding when AI should answer directly, provide a hint, ask for a first attempt, require verification, or route the task to a human.

Design record

Design owner: _____ Evidence or rationale: _____ Review date: _____

1. Task consequence

Record: _____

Suggested decision:

- What decision or action can the AI output influence?
- Who or what is affected if it is wrong, late, incomplete, or unauthorized?
- Is the outcome reversible?
- How quickly can an error be detected and repaired?
- Does the task involve financial, legal, safety, privacy, eligibility, access, or service consequences?
- Which cases must refuse or escalate rather than answer?
- Low-consequence reference or clerical support
- Moderate-consequence professional assistance
- High-consequence reviewed recommendation
- Prohibited for AI proposal or execution

Record: _____

2. User expertise

Record: _____

Decide whether the interface should explain more, require a first attempt, provide direct assistance, or route to a qualified reviewer.

- Novice, developing, experienced, or expert user
- Domain knowledge required
- Ability to detect unsupported AI output
- Familiarity with sources and policy
- Need for training or calibration
- Risk of over-reliance or automation bias

Record: _____

3. Learning and transfer goal

Select the primary goal:

If learning or independent judgment matters, define how the system will test transfer beyond the assisted case.

- Complete the task quickly
- Teach a concept or procedure
- Build independent judgment
- Preserve a rarely used critical skill
- Support reflection and review
- Produce a governed operational result

Record: _____

4. Time and safety constraints

Record:

Do not introduce staged assistance when delay is itself unsafe or operationally irresponsible.

- Response deadline
- Safety or incident urgency
- Service-level expectation
- Cost of delay
- Availability of a human expert
- Whether direct guidance is necessary to prevent harm

Record: _____

5. Accessibility accommodations

Record:

Accessibility is not optional friction. The assistance design must remain perceivable, operable, understandable, and robust.

- Keyboard and assistive-technology requirements
- Cognitive-load and memory demands
- Language or literacy needs
- Visual, auditory, motor, or speech accommodations
- Need for reduced steps, direct answers, or alternative formats
- Whether a required first attempt creates an unnecessary barrier

Record: _____

6. Select an assistance level

Level 0 - Observe

The system provides a workspace, captures context, or waits without generating an answer.

Use when: the human should perform the full reasoning or when evidence is not yet sufficient.

Level 1 - Ask

The system asks the user to state the goal, assumptions, or first interpretation.

Use when: a first attempt supports learning, judgment, or clearer intent.

Level 2 - Hint

The system identifies a relevant concept, source, or area of concern.

Use when: the user should continue the reasoning but may need orientation.

Level 3 - Scaffold

The system provides a checklist, template, staged questions, or partial structure.

Use when: the process matters and the user should retain ownership of the conclusion.

Level 4 - Draft

The system creates proposed work with sources, limitations, and required verification.

Use when: acceleration is valuable and a qualified person can review the result.

Level 5 - Execute under authority

The system performs a bounded action after explicit approval and with audit evidence.

Use when: the action, authority, rollback, and failure behavior are defined and tested.

7. Reflection and verification

Choose at least one appropriate control:

- User identifies the supporting source
- User explains what changed from the AI draft
- User states the remaining limitation
- User compares alternatives
- User confirms the downstream consequence
- User performs a second independent check
- Another reviewer approves
- The system runs a deterministic test or reconciliation
- The system refuses when evidence is insufficient

Record: _____

8. Independent-performance check

When human capability matters, define one or more measures:

Do not use immediate time-on-task as the only outcome.

- Similar task completed without AI
- Transfer to a new case
- Ability to identify an incorrect AI response
- Confidence compared with correctness
- Persistence after a difficult case
- User-authored reasoning quality
- Need for escalation
- Retention after time has passed

Record: _____

9. Human-review ownership

Name:

Define who may change the assistance level and who approves a higher-authority execution path.

- Product or workflow owner
- Domain reviewer
- Approver
- Executor
- Escalation owner
- Accessibility owner
- Learning or capability owner
- Evidence owner

Record: _____

10. Measurement plan

Track only what is necessary and approved. Candidate measures include:

Do not add surveillance merely because it is technically possible.

- Assisted and independent task quality

- Override and refusal rate
- Time and cost
- Accessibility defects
- Blind acceptance indicators
- User-reported usefulness
- Escalation patterns
- Operational incidents

Record: _____

Usage boundary: This worksheet is a design aid, not a certification, clinical assessment, training standard, or universal rule that more friction is always better.

Last reviewed 2026-07-23 | HTML version: <https://longtermcapabilities.com/resources/human-capability-safeguards-ai-work/> | Contact: 1 (464) 274-1476 / consulting@longtermcapabilities.com