

The bot stops here

Human judgment architecture
for AI-assisted due diligence in grantmaking

Josh Davidoff

the Hollyhock
foundation

AI disclosure: Claude Design drafted this deck based on my outline.
All concepts are mine.

Agentic AI tools can process any grant diligence materials

(mostly agnostic to format)

Budget	Audit	Form 990
Proposal narrative	Prior report	Internal notes

Agents can one-shot a diligence report...

Figures pulled Cross-document extraction	Sources compared Triangulating budget, 990, audit	Conflicts surfaced Discrepancies flagged inline
Risks compressed A short, scannable summary	Questions drafted Follow-ups for the grantee	Memo language started A confident-sounding first draft

...producing a *full draft* before anyone has formed a view.

EFFORTFUL HUMAN JUDGMENT



REDUCED TO REVIEW OF AI OUTPUT

What are we trying to answer?

Which evidence governs?

What facts are material to the grant decision?

Where to trust and where to verify?

"Does this draft look right?"

Review is not the same as judgment.

Writing is thinking. *

**DUAL-PROCESS
REASONING**

**AUTOMATION
BIAS**

**OUT-OF-THE-
LOOP RISK**

**MEANINGFUL
OVERSIGHT?**

Lee et al, "[The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers](#)," *Microsoft Research* (2025)

Jussupow et al., "[Trust and reliance on AI: An experimental study on the extent and costs of overreliance on AI](#)," *Computers in Human Behavior* (2024)

Skitka, Mosier & Burdick, "[Accountability and automation bias](#)," *Int. J. Human-Computer Studies* (2000)

*For some. See recent popular articles on the pros and cons of AI-assisted drafting and brainstorming by [Ethan Mollick](#) and [Rebecca Winthrop](#)

Cognitive checkpoints (forcing functions)

Stop the workflow

Buçinca, Malaya & Gajos, “[To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making](#),” *ACM HCI* (2021)

Require human judgment

Ghosh et al., “[An Experimental Comparison of Cognitive Forcing Functions for Execution Plans in AI-Assisted Writing](#),” (2026) [Preprint]

Incorporate judgment into AI context

Continue to the next checkpoint

Cognitive checkpoints

Gates should mandate the act of judgment

DOES NOT PASS THE CHECKPOINT

Gestures of involvement.

~~"Please review carefully."~~

~~"Looks good." "Go ahead."~~

~~The model's own explanation of itself.~~

~~A bare approval click.~~

REQUIRED TO PASS — HUMAN ANALYSIS

A first view, written by a person.

Choose the source of record.

Name the material risk.

Write the working assessment.

Mark what remains unsupported.

Scaffolded Skill.MD flow

Gate 1	Packet completeness and scope	<i>"What are we reviewing, and what is missing?"</i>
Gate 2	Extracted fact table and unresolved conflicts	<i>"Which facts are settled, disputed, or decision-relevant?"</i>
Gate 3a	Required user analytical input	<i>"What is your working assessment before the memo exists?"</i>
Gate 3b	Independent AI first pass <i>(Use subagent)</i>	<i>"What does a fresh reviewer see using only the approved facts?"</i> <i>Mitigate confirmation bias.</i>
Gate 4	Reconciliation, final memo, and acceptance	<i>"What changes after comparing the human view with the independent AI read?"</i>

Cognitive checkpoints

Implementation

AI policy

“If employing AI for grant diligence, mandatory to use xyz tools/skills. Must check all AI work.”

Staff training

Necessary, *but* decays under time pressure.

Basic Skill.MD deployed
in ChatGPT / Claude Chat

Easy to distribute, *but* lower fidelity. Users can skip steps. Less processing power.

Robust Skill.MD packages in desktop
coding harnesses (Claude Code/Codex)

High fidelity to sequential instructions, powerful file handling, scripted code execution *but* risky/impractical for broad deployment to nontechnical users.

Production web app

Best for UX and workflow enforcement, *but* expensive, rigid, and technical debt.

What if AI's greatest value is not speed?

Responsible AI for due diligence may not save much time,
but it can improve analysis.

Activation
energy

Attention
to detail

Consistent
rubric

Thought
partner