



Co-Intelligence Readiness Framework

What If Love — a lab for benevolent AI

May 2026

whatiflove.org

Introduction to the Framework

Readiness is not what you know about AI. It is knowing what you are becoming as you use it.

There is no standard for the quality of human judgment inside AI collaboration. We assess what people know about the tools and how well they operate them, but those skills expire as quickly as the tools themselves—what counts as fluency now will look quaint a year later. As the technology advances, the demands placed on human judgment grow with it: to recognize new possibilities, to anticipate new risks, and to pursue the leaps forward that become possible when both are held in view at once. What does not turn over is the question beneath the tool: not what the system can do, but what is worth doing at all.

The **Co-Intelligence Readiness Framework (CIRF)** measures a human's capacity to direct increasingly capable intelligence systems toward purposes those systems cannot define for themselves. It scores that quality of human judgment across five domains: whether there's moral intent behind the work, how you plan and scope it, how you steer the model turn by turn, what you actually produce, and how you elevate the partnership. The companion assessment tool scores each of the 17 sub-skills within the five domains against collaboration that has already happened. **It is conducted entirely within the user's own model, grounded in evidence from prior collaboration rather than self-reported claims, and designed to balance maximum insight with strict privacy preservation.**

In *Co-Intelligence*, Ethan Mollick offers four rules for working with AI: invite the model, stay human, treat it like a person, and assume it will become more capable over time. CIRF extends that insight from individual practice to human development at scale.

While models update, humans practice. Each session leaves the next one standing in a larger room: more moves visible, more worth pursuing, more that the model never raised. What compounds is not the model's capability, which arrives the same for everyone, but the range of possibilities the co-intelligent can see and steer toward.

Manufacturing possibility out of concern and capability is what it has always meant to be human. The question CIRF poses, then, is: *"Does the human grow more human with every turn, and does the collaboration improve with it?"*

A snapshot of capacity cannot answer that question. The choices we leave behind do.

The Co-Intelligence Readiness Assessment is available at whatiflove.org/co-intelligence-readiness-assessment

THE FIVE DOMAINS

- 1 **Benevolent Intent and Moral Anchoring**
- 2 **Planning and Scope**
- 3 **Agency, Judgment, and Steering**
- 4 **Follow-Through and Impact**
- 5 **Adaptive Intelligence and Co-Evolution**

THE IDEH SCALE

I Introductory	D Developing	E Effective	H Highly Effective
----------------	--------------	-------------	--------------------

FRAMEWORK NOTES

Scale	I and H are narrow bands describing category changes. D and E are wide bands where most growth occurs. The D-to-E progression contains the most granular indicators because it represents the core developmental journey—from instinct to reliable practice.
NR	Not Rated applies when evidence is genuinely insufficient—not because the user is weak, but because the chat record lacks the relevant activity. Capped at two subdomains per assessment.
No gating or ceiling	Each domain is rated on its own evidence independently. One domain's score does not adjust another's.
What this does not measure	Technical skill with AI tools, speed of production, volume of output, or comfort with technology. A slow, careful user who produces one well-governed output scores higher than a fast user who produces ten ungoverned ones.

ABOUT THE ASSESSMENT

Practice, not intent	CIRA scores what is visible in the conversation record—patterns in what was actually done, not stated values or intentions. Vocabulary is not evidence. The report describes patterns, not instances; specific projects inform pattern recognition but do not appear as citations.
How to run it	Paste the CIRA prompt into your AI model (Claude or ChatGPT). The model reviews your existing conversation history and produces a scored report. No conversation content leaves your model; the assessment runs entirely within it.
Sampling weight	Production sessions carry more diagnostic weight than knowledge sessions. The instrument is most revealing when you are building, deciding, writing, or steering—not studying. One extended production session with visible steering is more diagnostic than ten short exchanges.
Session requirements	Domains 1–4 can be assessed from a single extended session. Domain 5 requires evidence across multiple sessions and cannot be rated otherwise.
Calibration	Scores are calibrated to be honest. Not every user should score well; many score Introductory across most domains. That is the assessment working, not failing. Inflated scores destroy the tool's credibility.

DOMAIN 1

Benevolent Intent and Moral Anchoring

Why the user wants to use AI, and whether moral intent shapes what gets built.

The core risk of the AI age is capability without concern; this framework measures concern before we assess ability. Neutrality does not exist in morality. Domain 1 treats relational intelligence as infrastructure—not sentiment—by designing work that strengthens the conditions for care, trust, and collaboration to flourish. Domain 1 constraints are system-level moral commitments: non-negotiables that shape vision, design, strategy, and what should never be built or automated.

1a

What You're Creating, and Who It's For

Whether the user's purpose is legible—who it serves, what good it creates, and what harm it refuses. Not what they won't do, but who they are choosing to serve and why.

- Names intended beneficiaries in context—not just 'users' but people in specific situations with specific needs.
- States purpose in terms that constrain action.
- Names explicit refusals.
- Shows a bias toward outcomes that reduce suffering rather than maximize status or control at others' expense.

I Introductory Purpose is vague or mainly self-oriented. Beneficiaries are unnamed or generic.	D Developing States intent to serve others but inconsistently connects that intent to decisions.	E Effective Purpose is clear and consistently shapes choices. Beneficiaries are specific. Refusals are explicit. Intent is durable.	H Highly Effective Moral commitments challenge norms or industry-standard practice when those norms permit avoidable harm. Positions are defensible, not merely idealistic.
--	--	---	---

1b

Seeing Consequences, Now and Later

Whether the user anticipates who could be harmed, how misuse could happen, and how power and inequality might be amplified—not just now, but at scale and over time, in hands they don't control.

- Names plausible misuse cases and harm pathways.
- Asks who benefits, who bears cost, who loses agency.
- Thinks forward in time and outward in scale.
- Traces causal chains through lived experience—how a design choice reaches individuals and communities.

I Introductory Harm and power are rarely considered unless someone else raises them.	D Developing Notices obvious harms; misses less visible or second-order impacts.	E Effective Proactively maps harms and misuse. Adjusts plans and constraints accordingly.	H Highly Effective Maps unintended consequences across time horizons, fields, and contexts. When tradeoffs are made, they are genuinely intentional—not accidental omissions disguised as decisions.
--	--	---	--

1c

Build Trust, Bring Together

Whether the user orients toward human connection and recognizes that production decisions carry moral weight—because someone bears the cost of every tradeoff between speed and accuracy, shortcuts and rigor.

- Designs for connection.
- Seeks consent for sensitive material.
- Preserves voice and context when representing people.
- Recognizes when a production shortcut carries a hidden cost for someone else, and names that cost before deciding.

I Introductory	D Developing	E Effective	H Highly Effective
Avoids obvious violations but is casual with sensitive details. Production decisions made on efficiency grounds.	Applies protections sometimes, but inconsistently—especially under urgency.	Consistent practices around consent, dignity, and agency. Production tradeoffs made with the moral dimension visible.	Work is implicitly designed to promote trust and human connection. Builds mechanisms that actively strengthen relational bonds as a structural feature of the work itself.

1d

Translating Commitments Into Features and Constraints

Whether moral intent actually shows up in the work as requirements, defaults, constraints, and boundaries. Values stated but not operationalized are not yet real.

- Converts values into operational constraints (“must,” “must not,” “requires human review”).
- Builds safety and agency features into the design.
- Detects proxy traps and redesigns success criteria rather than optimizing a flawed metric.

I Introductory	D Developing	E Effective	H Highly Effective
Values are stated, but features and constraints are missing or optional.	Some constraints exist, but they are incomplete, inconsistent, or easy to bypass.	Moral commitments reliably become constraints and design choices. Tradeoffs are made explicitly.	Constraints and features serve multiple purposes simultaneously—one design choice addresses safety, usability, and values alignment at once.

DOMAIN 2

Planning and Scope

How the user plans and scopes AI projects, and whether that planning is grounded in what AI can actually do.

Without that foundation, scope is set against a fantasy or an arbitrary limit set in the past. Domain 2 measures whether you can lead the work as structural authority: setting a clear goal, defining what “done” means, and designing scope, sequence, roles, and stopping rules before production begins. Good planning protects intent: converting capability into aligned progress and preventing drift, overreach, and the erosion of moral boundaries under scale.

2a

Knowing What Done Looks Like

Whether the user defines purpose, audience, scope, and success criteria before producing, and maintains orientation when the work expands.

- States purpose, audience, and intended use.
- Sets scope boundaries.
- Defines success criteria that can actually be checked.
- Maintains orientation when the model expands scope or introduces attractive tangents.

I Introductory A goal exists, but scope and success are unclear. Starts producing before defining what good looks like.	D Developing Scope and criteria exist but shift during production or remain vague enough to accommodate almost any output.	E Effective Clear scope and success criteria. Keeps work oriented. Can explain what is in, what is out, and why.	H Highly Effective Internalized planning system that carries across projects without reinvention. Attention goes entirely to the work itself, not to maintaining the plan.
---	--	--	--

2b

Setting Boundaries That Help

Whether the user sets operational boundaries and pause points that improve outcomes—including anticipating how the plan could break at the structural level.

- Sets stopping rules around time, quality, confidence, or risk.
- Adds pause points before high-stakes actions.
- Names when AI should not be used for a given task.
- Anticipates which decisions would compound a mistake versus contain it.

I Introductory Few boundaries. Relies on instinct to know when to stop.	D Developing Adds boundaries after problems appear. Stopping rules are reactive.	E Effective Boundaries are present from the start. Stopping rules are consistent and purposeful.	H Highly Effective Boundary-setting is internalized as infrastructure, not overhead. Stopping rules deploy automatically with minimal setup cost.
---	--	--	---

2c

Who Does What, When

Whether the user separates phases, assigns responsibilities, and maintains clarity about what must remain human work.

- Separates phases: plan, draft, test, revise, package, deliver.
- Assigns human-required responsibilities.
- Uses templates, checklists, or standard processes to reduce drift.
- Keeps the workflow valid even if tools change.

I Introductory Process is ad hoc. Steps blur. The user and the AI trade turns without clear structure.	D Developing Some structure exists but execution is inconsistent. Roles blur under time pressure.	E Effective Repeatable workflow. Clear roles. The user knows which decisions are theirs.	H Highly Effective Role clarity designed to protect bandwidth for the unexpected. Workflow accommodates in-the-moment pivots because structure was built with slack in it.
---	---	--	--

DOMAIN 3

Agency, Judgment, and Steering

The user's skill in steering AI, during a session, move by move.

Domain 3 measures the “third turn”: your ability to reframe, challenge, verify, simplify, deepen, or stop without losing coherence or voice. Strong steering is not endless iteration—it is choosing the right move at the right moment.

Human hallucination is every bit as real as AI hallucination. The model drifts through fabrication and overconfidence. The human drifts through excitement, fatigue, deference, and the gradual replacement of their own thinking with the model's framing. Domain 3 measures whether the human catches both.

3a

The Third Turn

The user requests, the model delivers, and what the user does next sets the ceiling for the entire collaboration. The third turn is where co-intelligence is won or lost.

- Uses purposeful moves—reframe, narrow, broaden, challenge, simplify, stop, ship.
- Detects drift and resets direction quickly.
- Seeks opposing viewpoints and alternative framings as a deliberate practice.
- Generates alternatives from outside their own framework.

I Introductory Accepts the model's framing. Default move is to ask for another revision of the same thing.	D Developing Tries different moves but falls back to “revise again” under pressure.	E Effective Chooses moves intentionally and can explain why this move now. Regularly seeks perspectives outside their own framework.	H Highly Effective Steering moves consistently improve the work beyond its original trajectory—pushing to greater depth, sharper framing, or higher impact than originally envisioned.
--	---	--	--

3b

Knowing When the Work Is Real

Whether the user maintains contact with reality throughout the collaboration—not just factual accuracy, but alignment of intent, groundedness of quality, and resistance to losing purpose turn by turn.

- Verification: checks claims that carry real-world consequences. Catches internal contradictions.
- Intent alignment: recognizes when the AI's interpretation has drifted from original purpose.
- Self-monitoring: catches their own drift. Distinguishes 'satisfied because this is good' from 'satisfied because I'm tired.'

I Introductory Takes AI output largely at face value. Does not recognize when intent has drifted or quality is performed rather than genuine.	D Developing Verifies some claims and catches some drift, but inconsistently.	E Effective Maintains a reliable practice of checking whether the work is real—across facts, intent, quality, and their own engagement.	H Highly Effective Reality-testing is rigorous and habitual. Self-monitoring is continuous rather than triggered by obvious signals.
---	---	---	--

3c

Identity and Autonomy

Whether the user maintains their own thinking, voice, and analytical framework. The challenge is not whether writing still sounds like you—it is whether you still think like you.

- Maintains stable intent, tone, and identity across iterations.
- Resists model-shaped language that conflicts with purpose or voice.
- Uses anchors—principles, style notes, decision logs—to hold coherence.
- Can name the model's framing before accepting or rejecting it.

I Introductory Voice shifts easily. Adopts the model's framing, vocabulary, and conclusions without visible independent evaluation.	D Developing Has preferences and positions but struggles to preserve them under the model's influence.	E Effective Stable voice and framework. AI supports the user's identity rather than replacing it.	H Highly Effective Vision, voice, quality bar, and tolerance for ambiguity are so consistently displayed that collaborators could predict standards and feedback before receiving them.
---	--	---	---

3d

Inventing a Way Forward

When the user hits genuine gridlock—between feedback that stings and shortcuts that skip, between competing priorities that both matter—can they find a creative resolution that is neither capitulation nor stubbornness?

- Recognizes genuine gridlock rather than forcing a premature resolution.
- Produces creative resolutions that integrate competing demands rather than choosing between them.
- Reaches for analogies, reframings, or lateral connections when the direct path is blocked.

I Introductory When stuck, either capitulates or digs in. Does not generate new options.	D Developing Occasionally finds creative resolutions, but momentum or clarity often drops during the pivot.	E Effective Consistently turns gridlock into forward motion. Produces resolutions that are genuinely new—not compromises but integrations.	H Highly Effective Produces resolutions the AI genuinely could not have generated. Sometimes invents options that did not exist before the moment demanded them.
--	---	--	--

3e

Understanding What AI Can and Can't Do

Whether the user has a working sense of what AI systems can and cannot do—not as a memorized list, but as calibrated trust that shapes how they verify, delegate, choose tools, and override.

- Anticipates common failure modes—confident wrongness, plausible fabrication, over-agreement.
- Treats “sounds good” as a weak signal.
- Matches tool to task based on speed, accuracy, sensitivity, and stakes.
- Avoids premature automation.

I Introductory Treats AI as generally reliable or generally unreliable. Cannot translate “AI hallucinates” into different behavior.	D Developing Names some limitations but cannot consistently translate them into steering behavior or tool choices.	E Effective Has a usable mental model that improves verification, steering, and tool selection.	H Highly Effective Actively tracks how AI capabilities and limitations are evolving—experimenting at edges, probing new behaviors, and revising their working understanding as the tools change.
---	--	---	--

DOMAIN 4

Follow-Through and Impact

What the user has produced with AI, and whether it survives contact with reality.

A draft that only the model sees is not delivery. Delivery means an output that other people can use, trust, and act on without you in the room—concrete outputs that hold up under scrutiny, serve their intended audience, and persist or compound rather than expire.

4a

Shipping

Whether the user produces concrete, complete deliverables—not just drafts, ideas, or interesting conversations.

- Produces concrete deliverables—plans, memos, curricula, policy drafts, analyses, prototypes, toolkits.
- Chooses format that fits the audience.
- Output stands alone beyond the chat.
- Moves work from draft to deployment.

I Introductory Work stays at the level of drafts and ideas. The chat is productive but nothing ships.	D Developing Can finish simple deliverables with support. Quality is fragile.	E Effective Produces complete, usable artifacts reliably.	H Highly Effective Deliverables are finished with a degree of polish and completeness that would hold up in any professional context—not just the intended one.
---	---	---	---

4b

Craft and Handoff

Whether the user's outward-facing work holds up, travels well enough for others to act on, and persists or compounds over time.

- Quality: checks accuracy, internal consistency, missing pieces, and harmful ambiguity.
- Handoff: provides summary, assumptions, risks, and next steps.
- Prompt-level causation: traces a drifted output back to the specific instruction that produced it.

I Introductory Shares first-pass output that reads generic or performative. Others cannot act without asking the user to explain.	D Developing Notices quality problems and requests revisions, but standards are inconsistent. Sometimes mistakes polish for depth.	E Effective “Done” means no material errors, audience fit, clear and credible voice, and conceptual integrity. Outputs travel well.	H Highly Effective Work does not just meet the existing quality bar—it raises it. Output sets a new standard. Creates value that compounds.
---	--	---	---

DOMAIN 5

Adaptive Intelligence and Co-Evolution

What the user brings as a cognitive partner—and whether the partnership co-evolves or stalls.

Domains 1 through 4 describe the conditions for co-intelligence: intent, planning, steering, delivery. Domain 5 is the co-intelligence itself—the emergent state where the partnership operates at a level neither party reaches alone.

What the user signals through depth, breadth, and recursive practice sets the ceiling of what the collaboration can produce. A user can be Highly Effective on Domains 1–4 and Introductory here—operating AI well but bringing nothing the partnership couldn't replace with a different user. Domain 5 is observable only across sessions, because what it measures is whether the partnership compounds.

5a

Depth That Resists Displacement

Vertical knowledge in a specific field. Whether the user's expertise is visible in session—not claimed, but demonstrated through resistance when the model is confident and wrong.

- Corrects outputs using domain reasoning and real constraints, not just preference.
- Adds what the AI missed: real-world constraints and lived experience that change the work's conclusions.
- Builds and reuses knowledge over time so expertise compounds rather than resets.
- Can name not just when the model is wrong but why its reasoning structure produced the error.

I Introductory The absence of expertise is a signal. The model receives no pushback, reads no depth markers, and defaults to lower effort. Output quality drops not from model limitation but from user signal.	D Developing Has knowledge but hesitates to assert it against confident AI output. Some knowledge builds; much is lost.	E Effective Corrects and contextualizes reliably. Sets standards the AI must meet. The model reads the practitioner as an expert and works accordingly.	H Highly Effective Domain knowledge deep enough to diagnose not just when the AI is wrong, but why its reasoning structure produced that specific error—the shape of the model's blind spots as seen from the user's field.
---	---	---	---

5b

Breadth That Reshapes Possibility

Horizontal knowledge across fields. Whether imported frames change what question is being asked—not analogies that decorate the work, but structural connections that produce insight neither field generates alone.

- Bridges disciplines in ways that change strategy or constraints.
- Imports concepts responsibly—understanding what transfers and what doesn't.
- Brings connections the AI would not make: cross-domain links grounded in the user's unique combination of experience.

I Introductory Stays in one domain. Does not draw on knowledge from other fields.	D Developing Occasional connections across fields. Usefulness is uneven.	E Effective Regularly generates cross-domain connections that improve the work. Bridges are structural, not decorative.	H Highly Effective Cross-domain connections do not just add insight—they reframe what question is being asked. Imports a lens from another field that makes visible something the original framing could not see.
---	--	---	---

5c

Recursive Growth Through Practice

The closed loop of reflection and action, structured to improve how the practitioner improves—not just the work at hand. Agency is what makes the loop possible. The record reveals whether it runs.

- Reuses prompt structures and approaches that worked previously rather than rebuilding from scratch.
- Names what did not work and adjusts before re-attempting.
- Generalizes from single sessions rather than treating each as isolated.
- Later sessions show structural improvement in ways not explained by model improvements.
- Designs and improves the reflection process itself, not just the work.

I Introductory Reflects occasionally, usually when prompted. Rarely changes behavior as a result. Each session is treated as a fresh start.	D Developing Thinks about how sessions went but inconsistently acts on reflections. Adjustments tend to be one-off rather than carried forward.	E Effective Regularly captures what worked and uses those observations to change practice in the next session. The learning cycle closes.	H Highly Effective Treats the learning process itself as something to design and improve. Builds repeatable patterns that reduce the setup cost of future sessions.
---	---	---	---