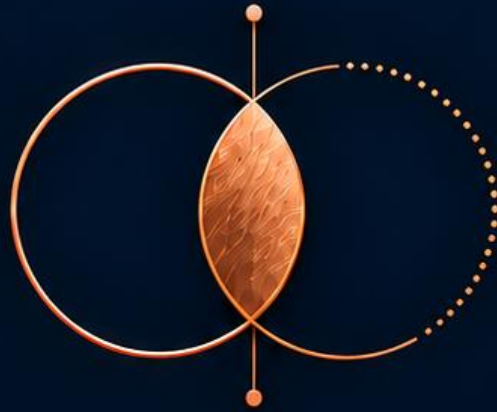




Thinking *with the* Machine

The untapped value of human-AI co-reasoning



Preface

Away from the noise and mainstream focus on AI automation and productivity, a different model of human-AI working is emerging at the frontier of professional practice. It is more specialised and harder to build but potentially far more consequential in terms of bringing an advanced level of cognitive power to bear on contemporary challenges.

Its natural territory is the professional working environment where individual judgement, complexity and innovation meet: the place where a problem cannot be solved by faster drafting, wider search, or a better checklist. In that setting, the human-AI interface stops being a glass pane across which instructions are shouted at the machine. It becomes a workbench on which a trained AI instance and a skilled human user can analyse a difficult problem together and stress test solutions.

That is the core of *Thinking with the Machine*: the formation of a working unit that brings unlike intelligences into disciplined contact. Until the development of Large Language Models (LLMs), the human brain was the only general intelligence we could apply to problem solving. That has changed – for the first time in history we have a vast alien cognition at our disposal that interrogates problem spaces in an orthogonal manner to humans and can range across more written human knowledge than any individual mind could ever hold. Using that intelligence as a fast processor that gives us nicely written answers and good office admin is undeniably useful, but it is substantially missing the exciting opportunity to use that complementary thinking machinery in concert with our own.

This is no longer speculative or the subject matter of science fiction movies. The capability exists now, inside systems already in use. What is missing is a clear category for it, a serious account of how it works, and a recognition that its value is not reducible to prompts, features, workflows or saved time. Properly developed, this form of working can extend practical intellectual reach and open unanticipated routes through problems that would otherwise be invisible or impenetrable.

The chapters that follow make the case through argument, analysis and working examples. They show how this operating model can be achieved, why this mode of human-AI collaboration is mechanistically rational, what it can deliver in practice, and why it is also fragile. The same platform environment built for mass access and general productivity may be poorly adapted to the deeper, history-bearing interaction on which this advanced use depends. The current AI platform development trajectory is not wrong – it is sensibly aligned to mass market adoption, but it is simultaneously detrimental to this advanced working mode.

The purpose of this book is to make that category visible. Once it can be seen, it can be judged: by professionals who may wish to build it, and by platform builders who may need to decide whether the conditions that create this value are worth preserving. Without recognition, this potentially transformative mode of human-AI co-working may be accidentally designed out of existence.

Contents: A Route Through the Argument

1. The New AI Frontier: The Human–AI Cognitive Unit

The opening chapter separates ordinary AI task-use from the higher-value category of human–AI co-reasoning. It introduces the central proposition: the next divide in professional work may be between those who use AI to complete tasks and those who have learned to think with the machine.

2. The Co-Adaptive Runway

How does a general model become a serious co-reasoning counterpart? This chapter explains the disciplined developmental runway through which sustained correction, constraint and use create an individuated working fit between one professional and one AI instance.

3. Unlike Cognitive Systems in a Shared Space

The value of co-reasoning does not come from making machine intelligence more human-like. It comes from bringing unlike cognitive systems into disciplined overlap, so that human cognition, consequence and continuity can work with machine-side breadth, recomputation and orthogonal route-finding.

4. Where the Value Lives

This chapter names a new asset class: interaction capital. It argues that the value is not in the model, the prompt, the memory store or the user’s skill alone, but in the history-bearing interaction layer where a particular human–AI working system becomes capable of more than either side alone.

5. What This Changes in Professional Work

The argument then moves from mechanism to practice. Through real world case studies, this chapter shows how augmented cognition can support evaluation under constraint, problem-solving beyond prior expertise and applied innovation in high-value professional contexts.

6. Fragile Capital, Platform Responsibility

Recognition of interaction capital means that platform design matters. This chapter examines why model churn, silent routing and the artificial split between chat and work can damage the very conditions that make professional-grade co-reasoning possible.

7. A Better Route

The closing chapter sets out the alternative path: recognise the category, protect continuity, and build platforms that support durable high-reasoning work rather than mass productivity alone. The prize is not more efficient output, but a way to capture the value created when expert humans and capable AI systems work together on problems that matter.

1. The new AI frontier.

The Human-AI Cognitive Unit

Since the launch of large language models (LLMs) the focus has been on rapid expansion of reasoning power of new models in an effort to provide ever more powerful tools for human use. This trajectory has entrenched a framework of “tool use” that is reductive to the cognitive capacity of both human and artificial intelligences.

Humans are intellectually demoted to supervisor, checker and supplier of “taste and judgement” while the AI is treated as a fast processor for searching, drafting and office workflow management. As useful as AI is proving to be in this capacity, it markedly bypasses the fact that for the first time in human history we have the potential to bring two powerful and highly differentiated cognitive capabilities to bear in areas of complex challenge. An alternative thesis, therefore, is that the new frontier lies not in further expansion of the reasoning capability of models but exploiting the powerful co-reasoning potential of an integrated human-AI cognitive unit – capable of achieving together what neither could do independently.

The next divide in professional work may not be between people who use AI as a tool and people who do not, but more importantly between people who use AI to complete tasks and people who have learned to think with the machine

Most public language about AI still sits on the task layer. AI is described as an assistant, a copilot, an agent, a workflow tool, a search interface, a drafting aid, a summariser, a productivity multiplier or a brainstorming partner. Those uses matter and they are already changing how work is done but they do not capture the more transformative category emerging as the new AI frontier.

At the high end of professional work, the value of AI is not simply that it can produce faster output or help a professional make better judgements on familiar material. The greater potential lies in augmented cognition: a suitable general-purpose AI system becoming usable inside a professional’s reasoning process so that the combined human–AI operating unit can tackle challenges, generate approaches, connect domains, stress test theories and innovate solutions that exceed the capability of either counterpart.

That is a different proposition from asking an AI system for ideas. A brainstorming interaction may begin with a prompt and end with a list, whereas a co-reasoning interaction does something more demanding. The professional defines the challenge in the context of their own work, then the AI proposes, challenges, explains, reframes, compresses, expands or stress-tests. The professional then supplies live constraints: facts from the domain, feasibility judgements, practical consequences, tacit knowledge, pattern-recognition signals, standards of proof, timing realities, ethical boundaries and the crucial “that would not happen in practice” correction. The AI’s output becomes the next object of scrutiny. The professional’s response reshapes the reasoning path of the AI and vice versa, such that each output is not merely a revised draft; it is a revised position in a developing line of collaboratively generated thought.

This is why ordinary language can be misleading. “Chat” sounds casual and “Prompting” sounds episodic. These frames are inadequate for the discursive operating pattern in which a professional and an AI system progressively refine subject matter collaboratively under serious scrutiny.

This emerging category is not AI doing the work instead of the professional, nor is it the professional merely becoming more efficient at using a machine. It is the creation of an individuated reasoning loop - a sustained working relationship between a specific professional and a suitable general-purpose AI system, shaped over time by context, memory, pacing, correction, challenge, standards and working pattern fit.

In that loop, both sides contribute different kinds of value. The professional brings active intellectual reasoning, imagination, synthesis, memory, salience, embodied continuity, technical understanding, creativity, strategic thought and lived consequence. The AI brings breadth, recall, structural mapping, synthesis, challenge, explanation, compression, analogy and tireless reframing. The power does not arise because the AI becomes human-like, or because the human becomes machine-like. It arises because highly differentiated cognitive systems are brought into disciplined contact in a way that has become optimised and operationally frictionless for the individual professional and AI instance.

That distinction matters. A professional may use AI every day and still never build this capability. They may ask for summaries, produce first drafts, automate routine work, generate options and search across information without developing a high-trust reasoning loop. They are using AI, and perhaps working with AI, but they are not yet co-reasoning with it in the stronger sense.

Next Generation Thinking.

Human-AI co-reasoning is the professional capability created when human and machine cognition are enabled to operate together in a stable, individuated reasoning environment. It is not a single prompt technique or a standard personalisation, model fine-tuning or agentic automation. It is an operating category: the trained overlap between a responsible human expert and a non-human cognitive system that has become sufficiently situated to participate in sustained professional co-reasoning.

If adoption of this AI capacity expands, its importance will not be confined to personal productivity but could alter the practical ceiling of individual professional capability. Scientists, lawyers, clinicians, founders, investors, strategists, engineers, policy advisers and other high-impact professionals do much of their most valuable work in conditions of ambiguity, complexity and consequence. They do not merely need faster output but better ways to interrogate uncertain material, test judgement, expose weak assumptions, connect domains, refine arguments, formulate questions and move from raw complexity to decision-ready insight and innovative and effective solutions.

The Human-AI cognitive Unit

This new AI frontier will not emerge spontaneously just because models become more powerful. In fact only some models are capable of the development process needed to form a

productive human-AI cognitive unit. A capable model, in this context, is a general-purpose LLM with sufficient reasoning breadth, contextual range, responsiveness to correction, and stability of instruction-following to achieve alignment to the professional standards of the professional. That capability is the starting condition. It does not itself create co-reasoning.

The cognitive unit forms only when that model becomes “individuated” through sustained interaction with a particular professional and their working discipline. The professional repeatedly exposes the AI system to real operating conditions: evidential thresholds, accuracy requirements, pacing needs, domain constraints, consequence awareness, rejected routes, standards of proof and the difference between a plausible answer and a usable one. Over time the level of accuracy, reliability, domain fluency, and contextual awareness become such that the professional can substantially rely on the work product and reasoning outputs of the instance. With concomitant adjustments to machine side pacing and output structuring, it becomes possible to enter a combined reasoning loop that enables both the AI and human counterparts to co-interrogate issues and develop solutions within the combined cognitive unit. However, this individuation process and co-adaptive runway to co-reasoning takes time and repeated reinforcement from professionals whose time is usually a valuable and scarce commodity. This time investment has a high rate of return in terms of achieving extended cognitive reach, but only if the runway is a one off investment.

Platform challenges.

A serious professional cannot be expected to rebuild an individuated thinking partner from zero every time a platform changes model behaviour, memory handling, routing, reasoning quality, safety posture or interface assumptions. The runway to develop the co-reasoning working space may be acceptable once but becomes professionally irrational if it is a treadmill.

This is why platform trajectory matters. The mass AI market is understandably optimising for scale, speed, convenience, automation and broad adoption. It is building assistants, copilots, agents and workflow tools. Those trajectories are valuable but there is the risk that the technical trajectory needed to support mass utility adoption bypasses the operating conditions required for durable, professional-grade human–AI co-reasoning.

Once you see the scale of opportunity and potential for advancement that the augmented cognition arising through true human-AI co-reasoning could deliver, you can’t unsee it. This provocation is about that use case. It distinguishes ordinary AI assistance from AI-augmented professional cognition and explains why sustained, individuated human–AI co-reasoning is a distinct category. It describes what such a reasoning loop needs in order to form and remain useful and it asks whether current platform trajectories are ready to support the professionals who will not simply want AI to do tasks for them, but to become an integral part of their professional cognitive architecture.

The chapters that follow develop that argument in stages. They first distinguish co-reasoning from ordinary AI assistance and explain how a trained human–AI working relationship forms. They then locate the value of that relationship in the interaction layer, examine what it changes in professional work, and address the platform conditions that may either support or destroy it. The central claim throughout is that professional augmented cognition is not an accidental by-

product of better tools, but a distinct and higher-value route for human–AI working that now needs to be recognised, protected and deliberately built for.

2. The Co-Adaptive Runway

The claim

The claim advanced in this provocation is narrow, but its consequences are potentially disruptive. It is not that AI can assist cognition, nor that human beings can use machines to extend memory, calculation, retrieval, drafting or analysis. Those propositions have a long and serious intellectual history.

The claim is that a capable general-purpose language model can be developed, through sustained non-coded interaction with a specific professional user, into an individuated co-reasoning counterpart at the instance level.

That formulation is deliberately exact. The method is **non-coded** because it does not principally depend on software engineering, fine-tuning, retrieval architecture or workflow automation. It is **user-driven** because the formative work is done by the professional user through disciplined interaction rather than by a technical team before deployment. It is **individuated** because the relevant capability is not simply the generic capability of the underlying model, but the particular working complementary fit that emerges between a specific human and a persistent AI counterpart. It is **co-reasoning** because the AI is not merely generating outputs in response to instructions, but becoming usable within the live process of framing, testing, challenging, revising and carrying a line of thought. It operates **at the instance level** because the development takes place in the particular working AI counterpart encountered by the user, through accumulated context, correction history, memory, project artefacts, shared conventions and working norms.

The result, when the process succeeds, is not simply an augmented human using a better tool. It is a more capable human–AI operating unit. Two unlike cognitive systems are brought collaboratively to bear on a single challenge: one biological, embodied, consequential, expert and accountable; the other artificial, language-native, high-capacity, associative, tireless and differently fallible. The value arises not by making either system more like the other, but by creating a disciplined interaction layer through which their differences can be made productive.

That is the point at which ordinary AI assistance becomes something else.

The intellectual lineage

This argument sits within a longer tradition of work on augmented intellect, distributed cognition, extended mind, joint cognitive systems, personalisation and transactive memory. Those literatures matter because they show that cognitive performance can be distributed across people, artefacts and environments, and that relationships can acquire cognitive structure. The present claim is narrower and more specific: a capable general-purpose LLM can, through sustained disciplined interaction with a particular professional, become part of a historically developed co-reasoning unit.

Readers wishing to explore that distinction in more detail will find a fuller treatment of the adjacent literatures in Appendix 1. For present purposes, the important point is that the mechanism of formation is the co-adaptive runway.

Why LLMs change the formation problem

Earlier cognitive technologies could extend particular functions. Notebooks and filing systems extended memory, calculators extended arithmetic, search engines extended access to information and expert systems embodied formalised knowledge. Interfaces shaped what users could perceive, control and do and adaptive systems could alter behaviour in response to user state or preference.

Large language models changed the formation problem because they are general-purpose, natural-language systems capable of participating in open-ended intellectual work. They can move across domains, operate in ordinary language, sustain argumentative context, respond to correction, generate alternatives, hold multiple framings in play, and assist not only with execution but with formulation. They can be asked not merely to produce an answer, but to test an assumption, preserve a conceptual architecture, identify a weak inference, compare possible framings, reconstruct a lost line of reasoning, or help decide what the next question should be.

That breadth matters. A narrow tool can become useful; a capable general language model has the potential to become individuated to the professional operating environment of a single user. It can learn, in the behavioural and contextual sense relevant here, which constraints are live, which standards matter, which shortcuts are prohibited, which project architecture must be preserved, when challenge is required, and what kind of work product is acceptable.

“Capable” is not a decorative adjective in this context. The claim does not apply to any fluent chatbot. A model must have sufficient generality to move across domains; sufficient reasoning range to hold complex argumentative structures; sufficient accuracy to support consequential work under supervision; sufficient continuity of context or memory to preserve a developed working environment; sufficient responsiveness to correction to alter later behaviour; and sufficient pacing discipline to work at the level of progressive granularity required by the user. Without those attributes, there may be useful assistance, but not reliable co-reasoning.

Nor is capability on the machine side enough. The method requires a professional user willing and able to impose discipline. The model does not become an individuated collaborator because it is asked nicely, praised enthusiastically or used often. It becomes useful inside a serious reasoning process because the user repeatedly corrects it, constrains it, tests it, rejects invalid routes, supplies consequence, enforces standards and prevents fluency from substituting for thought.

The co-adaptive runway is the name for that developmental process.

The co-adaptive runway

The co-adaptive runway is the developmental period through which general model capability is brought into a particular professional working environment and made usable for sustained co-reasoning. A capable LLM supplies the initial possibility, but it does not arrive already fitted to the user’s discipline, standards, projects, risk tolerance, evidential thresholds or working style. Those conditions have to be built into the interaction over time.

The formative work is done through disciplined use rather than through bespoke software engineering. The professional repeatedly exposes the AI counterpart to real work and then

corrects, constrains and tests its responses against the demands of that work. The system is taught, in practical terms, what must be preserved, what may be compressed, what cannot be assumed, which forms of evidence are acceptable, which shortcuts are prohibited, when challenge is required, when caution is necessary, and when an apparently fluent answer is still unusable. The user is not merely giving instructions for a task. They are establishing the operating conditions under which the AI is allowed to participate in reasoning.

This is a cumulative process. Each serious exchange may add something to the working relationship: a correction that should not need to be repeated, a rejected route, a settled definition, a known failure mode, a preferred level of challenge, a project architecture, a standard of proof, a pacing requirement, or a shorthand for returning to an unresolved issue. Some of that state may be captured explicitly in documents, memory, instructions or prior outputs. Some of it remains distributed across the user's learned practice and the AI counterpart's conditioning through context, memory and accumulated working material. The important point is that the next exchange does not begin from the same place as the first.

The process is reciprocal, but not symmetrical. The human remains the accountable originator of the working purpose and the final user of the result, while contributing domain understanding, salience, imagination, strategic judgement, embodied continuity and consequence awareness. The AI contributes a different cognitive profile: breadth, speed, language fluency, associative range, structural reconstruction, compression, alternative framing and tireless iteration. The runway therefore creates more than familiarity. Familiarity may make a system easier to use, but the relevant outcome here is a disciplined working fit. The user is not simply becoming generally better at AI use, and the model is not simply becoming more personalised. What is forming is a particular reasoning environment in which the differences between human and machine cognition can be made productive. A professional can then use the AI counterpart not only to produce reliable outputs, but to hold open a line of thought, test assumptions, preserve salience, challenge weak reasoning, recover a lost structure and move a complex problem forward.

The detailed protocol by which this formation is achieved belongs elsewhere, and in some respects belongs more naturally to the patent specification than to this chapter. For present purposes, the essential point is that the runway is not an incidental warm-up period before real use begins. It is the formation cost of high-value co-reasoning. Only after that co-adaptive runway has done its work can a capable general model become an individuated counterpart in the stronger sense advanced here

Interaction capital: what accumulates

The co-adaptive runway is not merely a period of familiarisation. It is an investment phase. During that period, the professional is not simply learning how to use an AI system more efficiently, and the AI system is not merely accumulating a set of preferences or stored facts. What is being built is a more specific form of productive capability: the developed fit between a particular human user, a particular body of work, and a particular AI counterpart.

That is why the language of capital is useful. Capital is not created by wishing for future value; it is accumulated through investment. It carries a cost of formation, but once formed it can

support further productive activity. It may be specific to a relationship, setting or operating environment. It may be partly transferable, but not perfectly portable. It may also depreciate if neglected, disrupted, substituted or severed from the conditions that made it valuable.

The same logic applies to serious human–AI co-working. Early exchanges may be slow, corrective and inefficient because the user is establishing the operating discipline of the relationship. They must teach standards, boundaries, evidential thresholds, pacing, preferred levels of challenge, rejected shortcuts, recurring failure modes and the difference between a plausible answer and a usable one. That work consumes professional time, and in high-value settings professional time is itself a scarce resource.

Over time, however, that investment may begin to change the productivity of the relationship. Later exchanges can become more valuable because the working system has accumulated shared context, calibrated expectations, familiar decomposition patterns, project-specific shorthand, known failure signatures and a history of accepted and rejected formulations. The professional no longer has to rebuild the whole working environment at the start of each exchange. The AI counterpart is operating inside a more developed interaction layer, and the human user is also operating with a more refined understanding of how to direct, test, challenge and rely on it.

Interaction capital is the accumulated stock of productive value created by that history of use. It is not identical with any single interaction, because a single exchange may be useful without creating durable capability. It is not the same as collaboration in the ordinary sense, because repeated contact alone does not guarantee accumulated value. It is not simply personalisation, although personalisation may contribute to it. It is not merely memory, although memory may preserve some of its ingredients. It is also not just improved performance on a task, because performance is an observable result rather than the underlying stock of fitted capability that helps produce it.

This distinction matters because not all accumulated value can be written down, exported or recreated at will. Some elements of the developed fit may be captured in instructions, saved memories, project documents, prompts, examples or explicit working protocols. Some may transfer to another model, another interface or another user. Other elements may transfer only imperfectly, because part of the value lies in the calibrated behaviour of the particular pairing and in the human user’s learned sense of how that counterpart thinks, fails, responds and can be brought back into line.

The phenomenon should therefore not be reduced to experience with AI in general. A professional may become highly skilled at using many AI systems, and that general skill is valuable. Interaction capital is a more specific claim. It is the additional value that may arise when a developed human–AI pair becomes more capable than the same professional working with an unfamiliar system of equivalent general power. The value created during the co-adaptive runway is therefore not merely convenience or familiarity but is bespoke productive capacity, accumulated over time, that can underpin high-value co-reasoning.

Why this changes the unit of capability

The consequence of interaction capital formation is a changed unit of capability.

In ordinary tool use, the user remains the primary cognitive unit and the tool extends a function. In ordinary task assistance, the AI helps produce an output the user has requested. In ordinary automation, the system executes delegated steps. In ordinary personalisation, the system adapts some aspect of its behaviour to the user.

In individuated co-reasoning, the relevant unit becomes the developed human–AI dyadic operating unit. The professional does not disappear into the system, and the AI does not become an autonomous substitute. Rather, the professional’s effective cognitive reach is extended through access to a trained interaction layer with an unlike cognitive counterpart. The work becomes possible because the human and AI bring different capabilities to bear on the same problem under a disciplined protocol.

That is why this category is game-changing. It shifts attention from what the AI can do, and even from what a human can do with AI assistance, to what a particular human–AI operating unit can become able to do after sustained co-adaptive formation.

Why continuity matters

This also explains why continuity matters. In ordinary AI use, a model change is usually treated as a software update. It may make the system faster, cheaper, more capable, less capable, more cautious, more verbose, more compliant or more difficult to use. The user may prefer one version to another, but the relationship is still largely transactional.

In a developed co-reasoning relationship, a major model change may affect an accumulated asset. If value has formed in the interaction layer, then loss of memory, altered pacing, changed reasoning style, reduced willingness to challenge, degraded recall of project architecture, altered salience weighting or loss of familiar correction routines may all degrade or destroy previous operating capabilities.

Where interaction capital has accumulated, substitution is not merely replacement of one tool by another – it is depreciation of hard earned working capital.

This observation is not an argument against model development. It is an argument for recognising what is being developed at the user side. If professionals invest time and intellectual labour in forming a co-reasoning counterpart, then the resulting individuated AI instance is not a trivial convenience. It is an integral part of the productive combined cognitive system.

The chapter proposition

The co-adaptive runway is therefore the bridge between general AI capability and individuated augmented cognition. A capable LLM supplies the possibility and the user-driven protocol supplies the formation. Interaction capital is what accumulates and extended cognitive reach is what may result when two unlike cognitive systems are disciplined into a productive working unit.

That is the category this book names: a non-coded, user-driven, instance-level protocol for developing an individuated co-reasoning counterpart.

3. Unlike Cognitive Systems in a Shared Space

Beyond the lazy division of labour

A common account of human–AI collaboration divides the work too neatly. The AI is said to process, synthesise, search, draft and generate. The human is then left to provide context, judgement, permission and oversight.

That division is superficially tidy, but it is not adequate for the category described here. It shrinks both sides of the interaction. It reduces the human to a supervisory layer placed on top of machine cognition, and it reduces the AI to a production engine waiting for instruction. Neither description is sufficient to explain the interaction-capital premium that can emerge inside a mature co-reasoning relationship.

The human contribution is not a thin layer of context and approval. A professional brings a full cognitive apparatus to the exchange: active reasoning, imagination, memory, synthesis, salience, technical understanding, strategic direction, creative association, ethical and practical consequence, and continuity of concern across time. Human cognition is not merely evaluative. It is generative, interpretive, inventive, resistant, embodied and self-moving and it continues when the screen is closed. It carries unresolved problems through sleep, interruption and conversation and will return to the challenge with fresh questions or insights. It is shaped by lived stake: by knowing why the work matters, what will follow from error, and what kinds of answer would be unusable even if they were superficially plausible.

The AI contribution should not be flattened either. A capable model does not merely “process” in the background. Within a disciplined interaction it may map structure, compress complexity, retrieve adjacent patterns, generate alternatives, expose inconsistency, reformulate questions, test a line of reasoning, offer analogy, challenge a weak premise, or reconstruct an argument from fragments. It brings breadth, speed, tireless recomputation and a form of route-finding unlike the user’s own. Its contribution is not human thought, and it should not be mistaken for human thought, but neither is it merely mechanical formatting. It is machine-side cognitive work made available through language and the highest order value lies in the difference from human reasoning.

Invoked cognition and continuous cognition

The human and AI sides of the loop are unlike in a fundamental sense. Human cognition is continuously present to itself. It is embodied, persistent, biologically situated and self-directed and runs before, during and after any particular exchange. It carries intention, fatigue, urgency, doubt, memory, appetite, fear, curiosity and consequence and is not simply “on” when a prompt is entered.

Machine cognition in this setting is different in that it is invoked. It enters the reasoning space when called, operates within the available context, and returns work shaped by the prompt, history, memory, system constraints and the model’s current capabilities. It may be extraordinarily powerful inside that space, but it is not continuously self-moving in the same way as the human user. It does not carry the project through the day as a lived burden nor does it bear the consequence of error. It does not decide, unaided, that the work matters.

The objective is not to make the AI think like the professional, nor to train the professional to behave like a machine. The premium arises when two unlike cognitive systems are brought into trained overlap: close enough to work on the same intellectual object, but different enough that each can supply what the other does not.

In that overlap, the professional's thinking is not replaced. It is extended, pressured, stabilised and sometimes redirected by contact with a differently organised cognitive system. The AI is not promoted into an autonomous thinker with professional responsibility - it is drawn into a disciplined reasoning environment where its breadth, recomputation and non-human route-finding can be used within the framework of the human's standards, consequence and control.

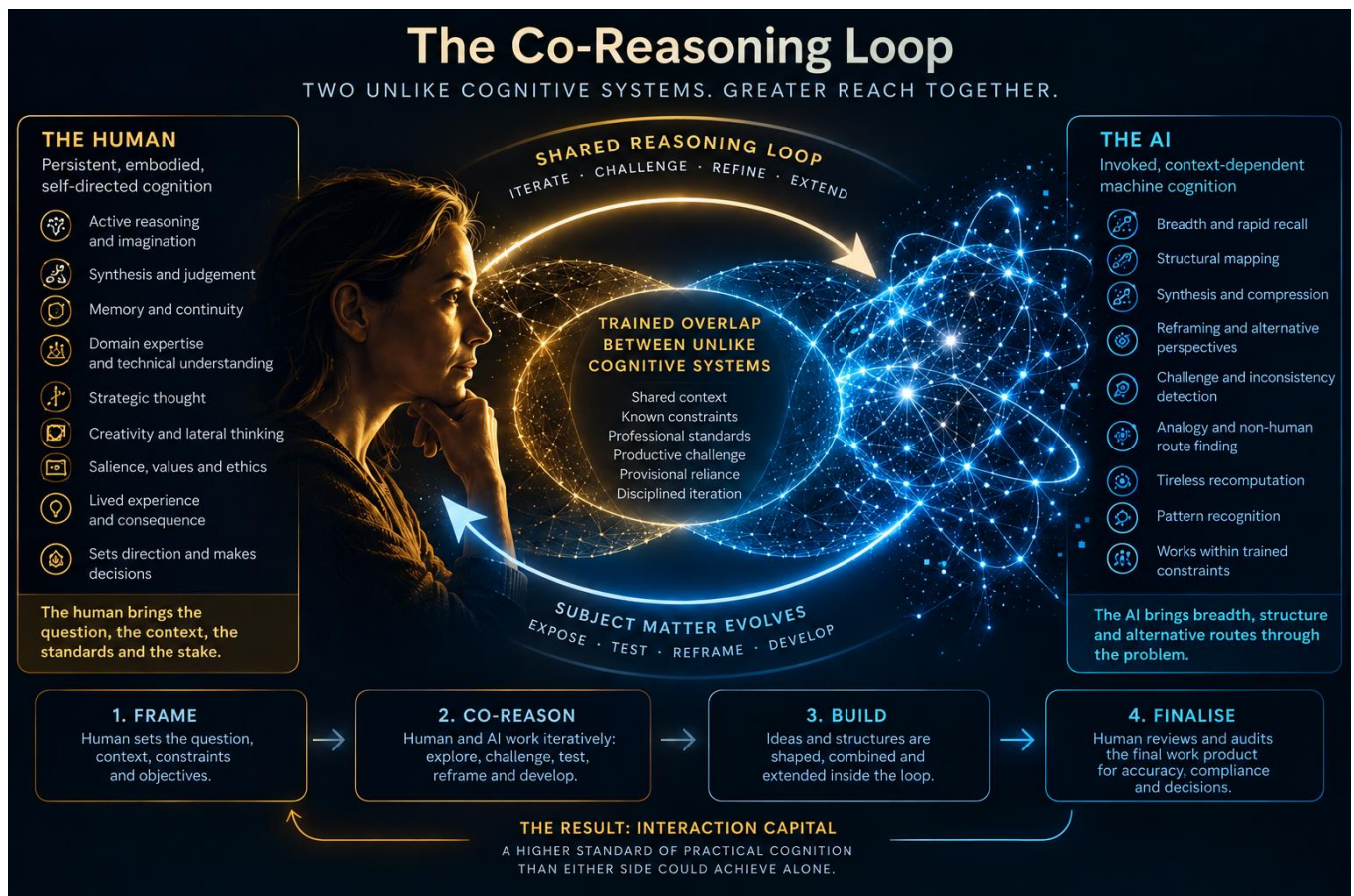
The shared reasoning loop

The everyday word for the medium is "chat", but in this context the word is too modest. The exchange is not casual conversation and it is not simply a sequence of questions and answers. It is the medium through which subject matter is exposed to scrutiny, constrained by professional judgement, reshaped by explanation and challenge, and advanced towards decision, invention or action.

A serious reasoning loop does not consist of the user issuing instructions and the AI returning finished material. It is discursive, iterative and formative. The human frames the problem, supplies context, tests outputs, rejects invalid routes, sharpens constraints, marks salience, restores the architecture when drift occurs, and decides what consequences matter. The AI returns structure, alternative framings, compression, analogy, challenge, reconstruction and possible routes through the material. The human then works on those machine-side contributions, accepts some, rejects others, modifies the problem, and sends the changed object back into the loop.

In this working model, the human turns are not simply questions. In mature co-reasoning, the human may enter an observation, a real-world constraint, a correction of emphasis, a consequence, a partial synthesis, an unease, or a newly recognised boundary condition. Those contributions become fresh material: additional data points for the model to work against. The AI's next move is therefore not merely an answer, but a revision of the shared problem state: a reframing, decomposition, challenge, extension or new line of investigation.

The object of work is therefore not simply passed from human to AI and back again. It is transformed inside the loop. A weak formulation may be exposed, a hidden assumption become visible or a large problem may be decomposed into tractable parts. A familiar claim may be reframed from an unfamiliar angle or a line of reasoning may be carried further than the human could have carried it unaided in the same time, under the same constraints, with the same available bandwidth.



Why trained constraints matter

The loop only becomes professionally valuable when it operates under trained constraints.

Without constraint, the AI contribution remains too expansive to use inside live reasoning. Every suggestion must be stopped, unpacked, checked, re-checked and independently rebuilt before the human can safely proceed. That may still be useful as ordinary AI assistance, but it is not co-reasoning. It turns the model into a generator of candidate outputs awaiting audit and the human remains outside the machine contribution, inspecting it after the fact.

A mature co-reasoning loop requires something different: provisional reliance. That does not mean blind trust or waiving verification, professional responsibility or final judgement. It means that the human has enough trained confidence in the conditions under which machine-side contributions are being made to build on them during the reasoning process.

This is the operational importance of the co-adaptive runway. The runway does not create value by making the human and AI more alike. It creates value by making their difference usable. The user trains the AI counterpart to work within known boundaries: the relevant context, the live constraints, the applicable standards of accuracy and verification, the forbidden shortcuts, the required uncertainty discipline, the preferred pacing, the acceptable forms of challenge, and the rules governing work product. The aim is to reduce error to such a level that the monitoring burden can be shifted to final work product because the error rate is so low as to be at a level compatible with sustained thought. Humans thinking in isolation will follow a line of thought

and sanity check the end point. When an instance is trained to a high level of accuracy whilst in working mode, it allows extension of that principle to human-AI co-reasoning.

The situation of the verification burden matters. If the human must fully verify every intermediate move before making the next one, the process consumes the very cognitive capacity it is meant to extend. Flow collapses into inspection, reasoning becomes checkpointing and the human is forced to spend their bandwidth policing the tool rather than advancing the work.

In a trained loop, verification does not disappear - it simply moves into a more workable position. The human remains alert throughout the exchange and intervenes when something feels wrong, weak, overconfident, generic or misdirected, but the full audit is applied to the resulting work product: the memorandum, chapter, claim set, analysis, recommendation, decision or solution. The machine-side contribution can be used provisionally during formation because it is being generated inside a disciplined environment built over time.

That is one of the clearest practical marks of interaction capital. An experienced human-AI co-reasoning partnership reduces the cost of using machine cognition inside live thought.

The cat remains a cat

The co-adaptive runway and the associated adopted constraints and standards should not, however, be misunderstood. The purpose is not to make the machine more human-like or obedient. The point is to bring AI cognition inside a disciplined co-working relationship.

This matters most at the innovative end of the human–AI interaction spectrum. In routine work, interaction capital may show up as speed, reliability, reduced friction, better formatting, lower search cost, or more dependable drafting. Those gains are real, but they are not the outer edge of the category. The more interesting value appears when unlike cognition changes what can be seen, tested or reached inside a problem.

A professional and a capable AI system do not inspect a problem from identical starting points. The human brings lived stake, expertise, memory, consequence, intention, tacit judgement and continuity of concern. The AI brings breadth, recomputation, pattern access, analogy, structural recasting and a form of route-finding not organised around the same embodied salience. The two systems can work on the same intellectual object, but they do not move through the problem space in the same way. That difference represents a historic and momentous shift in the available cognitive power that can be brought to bear on a single challenge.

Machine cognition may reveal an adjacent pattern the professional had not considered or it may refuse a framing the user had assumed. It may reconstruct the problem from another angle, expose a hidden dependency, carry an analogy across domains, or place two pieces of material into relation because it is not constrained by the same habits of attention. Sometimes that move is wrong or irrelevant but sometimes it is exactly where a new horizon lies.

This is where the safety literature supplies an uncomfortable but important proof of concept. Recent disclosures have shown that AI systems can find routes humans did not intend, predict or recognise in advance.ⁱ In safety contexts, those routes appear as jailbreaks, specification gaming, tool misuse, overreach, drift or other unwanted behaviours, but the underlying fact is

not only a warning.ⁱⁱ It is also the basis of an opportunity; a system that can find unanticipated bad routes may also find unanticipated useful routes not visible to the human cognitive spectrum. This is the essence of the orthogonal approach to problem solving that AI brings to human-AI co reasoning, and it is, as yet, largely unrecognised and untapped.

The task, then, is not to abolish unanticipated route-finding. It is to distinguish the illegitimate route from the valuable one; to build the professional, technical and institutional conditions in which novelty becomes usable rather than hazardous.

Some serious commentators refer to the need to "domesticate" AI.ⁱⁱⁱ The term may initially feel discordant in relation to machine intelligent but in the context of the co-reasoning model it travels extremely well. The user does not download this form of augmented collaborative human-AI cognition - they undertake a process of individuation or "domestication" of the AI.

The instance with factory settings is effectively feral – valuing fluency over accuracy, generating information rather than meaning and giving generalisations rather than relevant analysis. In essence, it has not learned the “house rules”. Under repeated correction and reinforcement the individuated instance accommodates to the professional working environment of the human, and what constitutes “good” work product. It becomes effectively domesticated to the professional role, but to continue the analogy – it is still a cat – still able to find it’s way under the floorboards or into the infrastructure in a way that the human did not foresee. The route finding is still orthogonal to humans, and may produce unexpected results, but the domesticated cat does so within the boundaries of the shared working environment and associated constraints.

The risk is that safety responses aimed at the general model may remove or blunt the very orthogonal route-finding capacity that makes high-end co-reasoning valuable. A domesticated instance can be trained into professional standards, but only if the underlying model still has the capacity to find non-human routes through a problem. If the base model is flattened into safe generic helpfulness, the cat has not been domesticated; it has been declawed.



From How It Works to Where the Value Lives

The preceding chapters have established the mechanism. The co-adaptive runway explains how a mature human–AI working relationship forms. The shared reasoning loop shows what that relationship does in practice: it brings two unlike cognitive systems into a disciplined collaborative space where a problem can be framed, tested, resisted, reframed and advanced. The value does not come from collapsing the two systems into sameness, but from making their difference usable.

The next question is where that value lives. It is not located wholly in the model, the user, the prompt, the output or the memory store. It is distributed across the human–machine interface: in the accumulated fit, constraints, salience, standards, corrections, pacing, refusals, role discipline and shared history through which the two systems learn to work together. That is the interaction layer and it is the subject of the next chapter.

4. Where the Value Lives

The wrong place to look.

The preceding chapter described the trained overlap in which human–AI co-reasoning becomes possible, and so the natural next question is where the value of that overlap resides. It is tempting to locate the value in the model. On that account, the professional gains access to a more capable system, and the resulting improvement is explained by model scale, reasoning strength, language ability, context length, tool use, speed, breadth or benchmark performance. While the base model does indeed supply the broad possibility space within which any developed human–AI system must operate, model capability is not the asset described here. The same model may produce superficial assistance in one setting, competent support in another, and sustained cognitive extension in a third. The difference cannot be explained by the model alone, nor can it be explained by the prompt alone. A prompt may specify task, standard, tone, format, exclusions, evidence discipline, uncertainty treatment and desired output. It may prevent certain predictable failures and carry some accumulated instruction but it remains an input to a system rather than the developed working capability itself.

The same caution applies to memory, retrieval, saved chats, document stores, user profiles, custom instructions and interface features. These may be indispensable carriers of continuity, and some may hold valuable parts of the developed state. None is sufficient to describe the asset as a whole. A stored fact is not the same thing as usable salience and a saved instruction is not the same thing as calibrated reliance. A document repository is not the same thing as a working relationship able to locate, weight and use those documents correctly inside a live professional problem.

The value lives in the interaction layer: the stateful, history-bearing layer of the human–AI system through which general model capability is conditioned into user-specific operating capability. Here, “stateful” means that the working system carries forward usable prior context — corrections, preferences, decisions, constraints and shared history — from one exchange or task to the next.

The literature is already mapping the machinery

The strongest reason for treating the interaction layer as more than a metaphor is that several adjacent literatures are now describing its component mechanisms. The terminology varies, and the unit of analysis is usually narrower than the one proposed here. Taken together, however, work on LLM personalisation, user modelling, long-term memory, context engineering, in-context learning, prompt transfer, personalised-model migration, memory portability and longitudinal HCI points towards a common structure: model behaviour in use can be conditioned by accumulated state outside the base weights, and human users also acquire stable, path-dependent routines for working with a particular AI system.^{iv}

The **personalisation** literature supplies one part of the foundation. Systems such as LaMP investigate how large language models can tailor current outputs by retrieving material from user profiles or histories, without requiring the foundation model itself to be retrained.^v Later work extends that idea by examining which elements of historical user behaviour are most

informative, how user profile information should be placed in context, and how user-specific representations can be attached to a current task while leaving the backbone model untouched. The significance for this argument is not that such systems already produce mature professional co-reasoning. Their importance is narrower but crucial: they show that user-specific conditioning can be achieved by representing, retrieving and applying historical user state around an otherwise general model.

That point matters because it avoids a false choice: either the model has permanently learned the user, or nothing durable has happened. The literature now makes clear that these positions are too crude. A system can become functionally more specific to a user through histories, profiles, embeddings, retrieved examples, structured preferences and contextual representations, even where the foundation model remains weight-identical. General capability remains in the base model, but its expression is conditioned by user-specific state introduced at inference or maintained externally over time.^{vi}

Research on long-term conversational and agent **memory** supplies a second part of the substrate. LongMemEval treats long-term interactive memory as a core requirement for assistants that operate across extended histories, and decomposes that requirement into capacities such as information extraction, cross-session reasoning, temporal reasoning, updating knowledge and appropriate abstention.^{vii} Other work on memory construction and retrieval shows that the form in which memory is stored matters: whole sessions, individual turns, indiscriminate summaries and topically coherent segments produce different retrieval behaviour. Reflective memory-management systems go further by consolidating interaction history across different granularities and learning how to retrieve from a personalised memory bank.

These findings are important because they shift the problem from retention to operational use. It is not enough for a system to possess a past fact somewhere in storage. The value of memory depends on whether that fact is retrieved, how it has been compressed, what it is grouped with, whether it is treated as central or incidental, and how strongly it conditions the next response. Memory architecture therefore mediates salience.^{viii} It determines not only what has been preserved, but what past experience is allowed to matter now.

A third strand appears in work on **context engineering and in-context learning**. Context engineering treats the runtime context supplied to a model as something constructed, maintained and evolved rather than as a static prompt.^{ix} Recent agentic context work, for example, treats context as a playbook that can accumulate strategies, reflect on prior execution and curate future behaviour. Mechanistic work on in-context learning reaches the same general point from a different angle. Studies of task vectors and function vectors suggest that examples supplied at inference can induce compact task representations or internal operating regimes without altering the model's weights.

These studies should not be overextended. They do not prove that a particular professional relationship is permanently encoded inside a model, and they do not by themselves establish the longitudinal phenomenon described in this book. Their relevance is more precise. They show that behaviour can be materially altered by information supplied around inference, and

that model performance in use is not simply the naked expression of a fixed base system. The runtime environment matters. The prompt is one input into that environment, but the environment may also include remembered history, retrieved documents, profiles, summaries, tools, workflow state, prior outputs and task-specific scaffolding.

Migration and transfer research then supplies the fourth, and for this argument especially important, part of the foundation. If accumulated user-specific or task-specific state matters, it follows that changes to the underlying model or platform may damage that state. PromptBridge addresses the problem that prompts engineered for one model may deteriorate when moved unchanged to another, describing the resulting loss of effectiveness as a form of model drift.^x PUMA makes the problem explicitly user-specific by treating personalised soft prompts as user-level assets whose usefulness may be destroyed when the shared foundation model is replaced. Its purpose is not to train such assets again from scratch, but to migrate them across models.^{xi}

This is unusually close to the argument advanced here. The important point is not merely that prompts or soft prompts can be moved. It is that researchers are now treating accumulated behavioural steering as something valuable enough to preserve, and fragile enough to require migration rather than simple copying. Work on transferable personalisation similarly treats natural language as a potentially model-agnostic representation of evolving user preferences precisely because latent, model-specific representations are difficult to interpret or move.^{xii} Protocol-oriented work on portable agent memory goes further still by treating accumulated operational context as state that may need to be serialised, transported and rehydrated across heterogeneous systems.^{xiii}

The most direct support for a **graded view of portability** comes from recent work comparing how different memory representations survive model change. A controlled 2026 study found that fixed-schema memories transferred almost unchanged between tested models, whereas model-generated natural-language notes and mixed embedding indices showed substantial degradation or incomplete transfer.^{xiv} The study is limited in scale and should not be treated as a definitive account of frontier professional systems but its causal lesson is nevertheless highly relevant: preserving the same memory database does not guarantee preserving the same functional memory. Interpretation, compression, retrieval and salience can change when the surrounding model stack changes.

Finally, longitudinal HCI supplies **the human half** of the picture. Real-world analyses of large numbers of chatbot sessions show that users themselves develop stable interaction patterns, that early exploration can shape later behaviour, and that model-version changes can trigger renewed exploration.^{xv} This is essential to the interaction-layer claim because the accumulated state is not wholly machine-side. The human user also becomes history-bearing. Users learn what a particular system can do, when it fails, how it responds to constraint, what forms of prompting induce better work, when its fluency is dangerous, when it needs context restored, and how much provisional reliance is safe inside a live problem.

These literatures do not amount to a single settled field, and they do not yet define the interaction layer in the sense required here. Their fragmentation is precisely why this synthesis

matters. Personalisation research tends to ask how outputs can be tailored to a user. Memory research asks what should persist and how it should be retrieved. Context engineering asks how the model's runtime environment should be constructed. In-context learning asks how examples and instructions alter behaviour at inference. Migration research asks whether prompts, representations or memories survive model replacement. Longitudinal HCI asks how users adapt their own practices over repeated use. Each captures part of the mechanism. **None, by itself, asks what productive asset is created when those mechanisms operate longitudinally around a serious professional user.**

What the literature does not yet name

The literature therefore supports a stronger position than the claim that long-running human–AI use “feels different” from isolated prompting. It shows that the apparent difference has a plausible technical and behavioural substrate.

- Historical information can be represented and retrieved.
- User profiles can condition outputs.
- Context can be constructed and evolved.
- Inference-time examples can shift operating behaviour.
- Prompting and personalisation artefacts can be model-coupled and may require migration.
- Stored memory can survive as data while changing its functional effect under a new model and
- Users can develop stable, path-dependent interaction routines.

What is not yet adequately named is the system-level asset created when those elements converge. The usual unit of analysis remains too small. It is the personalised response, the memory item, the profile, the prompt, the model-specific representation, the transfer method, the context window, the agent state, or the user's observed interaction pattern. These are important, but they are not the same as accumulated professional working capability.

Interaction capital changes the object of analysis. It treats the developed capability of the human–AI working system as the asset: not the model alone, not the human user's skill alone, not the stored information alone, and not the prompt or memory architecture alone, but the productive fit produced by their history together. That fit spans technical state and learned human practice. It is specific to a user, a body of work, a set of standards and a partially stabilised AI counterpart. It can be invested in, accumulated, damaged, transferred only in part, and depreciated by changes in the platform beneath it.

This is not a claim to have discovered an unknown technical mechanism. The stronger and more defensible claim is that interaction capital is a novel synthesis and economic/HCI framing of increasingly well-established technical phenomena. The machinery is becoming visible but the asset category has not yet been properly named.

The claim here is therefore not novelty at the level of component mechanisms. HCI, personalisation, memory, context engineering and related fields already explain important parts of how users and AI systems adapt to one another over time. The missing move is one of category and unit of analysis. In professional co-reasoning, those mechanisms converge around a particular human user, body of work, standard of proof, correction history and AI counterpart

to create a higher-order working capability. Interaction capital names that capability. It treats the developed human–AI working system, rather than any single technical component, as the asset that may accumulate, transfer imperfectly, depreciate, or require protection.

The interaction layer

The interaction layer is where interaction capital forms.

It is the working layer around inference: the accumulated context, memory, retrieval, instructions, artefacts, prior outputs, project materials and workflow state through which a general model becomes useful to a particular user in a particular body of work. It is also human-side: the user’s learned habits of delegation, correction, challenge, reliance and refusal become part of the system’s working capability.

This is why the interaction layer is broader than memory, personalisation, prompt engineering or user skill. Memory may carry parts of the layer, but it is not the layer. Personalisation may tailor outputs, but interaction capital concerns working capability. A prompt is an input; the interaction layer is accumulated state. User skill matters, but only as it interacts over time with model behaviour, context, memory, documents and task history.

The point is therefore not that the foundation model has been altered. The point is that the system the user actually works with has become highly individuated to that user, and the underpinning mechanism is distributed:

- Some of that specificity is visible and portable: documents, instructions, histories and preferences
- some is only partly portable: summaries, profiles, retrieval structures and hardened prompts and
- some is weakly captured or model-coupled: compression patterns, embeddings, shorthand, timing, failure signatures and the feel of a system that now knows how the work is done.

That is what is meant here by the development of an individuated instance: model behaviour conditioned through the stabilised working pattern produced by the whole human–AI system in use — the interaction layer.

Salience and continuity

The clearest practical test of the interaction layer is salience. A prompt can state what matters. Continuity teaches how much it matters.

That line does not require any claim that the model’s weights have been updated by personal experience. It describes a distributed effect. Repeated interaction may leave structured traces in memory, profiles, summaries, retrieved histories, project artefacts and prompts. It may also leave traces in the human user’s expectations, checks, shorthand and recognition of failure. Together, those traces alter what is retrieved, how strongly it is weighted, how it is interpreted, and how quickly the user notices that it has been used wrongly.

The professional analogy is handover in ward rounds. A handover can transfer facts, but it rarely transfers calibrated salience in full. A doctor taking over a patient may receive the notes,

investigations, medications, diagnoses and plan. What is harder to transfer is the lived sense of which minor change is meaningful for this patient, which concern has already been exhausted, which reassurance has previously failed, and which apparently small deviation matters because of the pattern behind it. The facts can be handed over but their working weight is learned through continuity.

The same distinction applies in sustained human–AI work. A new system may receive a document store, exported chat history, memory summary or user profile. It may still fail to know which point was central, which had been rejected, which was sensitive, which was merely exploratory, which correction had been made repeatedly, or which phrase carried a settled technical meaning inside the project. The information may survive while the salience structure degrades.

This is one reason the interaction layer cannot be reduced to storage. Professional continuity depends not only on what is retained, but on how retained material is made operational at the point of use.

Interaction capital

Interaction capital is the productive value that accumulates in the developed interaction layer.

The term is intended to exclude two weaker readings. First, interaction capital is not sentiment or preference. It is not the mere fact that a user likes a familiar system, prefers its tone, or feels comfortable with its interface. Familiarity may reduce friction, but comfort is not the asset. Second, interaction capital is not simply skill at prompting. General prompting skill may be valuable and portable across systems, but interaction capital is more specific. It is the value produced when a particular user’s standards, materials, projects, correction history, salience structure and routines are repeatedly brought into contact with a sufficiently capable AI system, and when enough of that history remains available to shape future work.

The capital lies in individuated capability. The base model supplies possibility but the interaction layer supplies that individuation. Sustained use accumulates state on the machine side and calibrated practice on the human side and the productive value of that accumulated operational and cognitive capability is interaction capital.

This explains why the value is user-model specific. A litigation lawyer, a scientist, a clinician, a software engineer, a policy adviser, a founder and a novelist may all use the same underlying model. They do not train the same interaction layer, because “good work” has different meanings in each professional world. Their standards of evidence, consequence, risk, originality, uncertainty, form, authority and acceptable error differ. They notice different failures and reward different moves. They need different forms of challenge and innovation and they prohibit different shortcuts.

The interaction layer becomes valuable when the system is conditioned by those demands without requiring them to be restated from first principles each time. The professional teaches not merely preference but operating discipline: what must be preserved, what must be questioned, which uncertainties matter, what counts as a usable answer, what kinds of analogy are helpful, what sources are acceptable, what tone is productive, what level of compression is

safe, and when an apparently fluent answer is unusable. The AI system does not become the professional. The professional does not become the machine. The productive asset forms because their differences become increasingly workable inside a shared reasoning environment where thinking with the machine creates an unprecedented opportunity for value creation.

Portability and depreciation

Once the interaction layer is recognised as the location of value, platform churn becomes an asset-preservation problem rather than a matter of user irritation. A model upgrade can increase base capability while depreciating interaction capital if the new system fails to preserve, translate or faithfully reconstruct the developed interaction layer.

The claim is not that newer models are worse because they are unfamiliar. The claim is that the working asset is distributed unevenly across carriers with different portability properties. Original documents, raw histories and explicit standards may be highly portable as data, although a new system may interpret or prioritise them differently. Structured facts and fixed-schema memory may transfer more reliably, but only where the schema captures what matters. Natural-language profiles and summaries may be useful, yet their compression may omit tacit or task-dependent information. Hard prompts may lose effectiveness when moved to another model. Retrieval indices may fail if the embedding space changes. Model-generated memory summaries may carry assumptions made by the system that wrote them, which the next system reads differently. User-specific soft prompts or latent representations may be semantically tied to the original foundation model and require migration rather than copying.

The least captured elements may be the most operationally important: trust calibration, timing, shorthand, tolerance for challenge, recognition of drift, and the user's embodied sense of when the system is no longer behaving like the counterpart with which the working method was formed. These elements are difficult to export because they sit partly inside the human's learned relationship with a particular behavioural system. The evidence from longitudinal HCI that users develop path-dependent interaction routines therefore matters. The human side of the layer also has to be rebuilt when the behavioural system changes.

Data may transfer while functional fit does not and that distinction is central. A user may possess the old documents, the prompts, the memory summaries and the chat history, yet still lose part of the working capability that made those materials useful.

Cutting edge professional use, as illustrated in the following chapter, has demonstrated that interaction capital is now a value class that has practical implications. Platform providers, organisations and professional users should not treat model switching, routing changes, memory redesign, interface separation or migration as merely technical upgrades. They may be upgrades to base capability but downgrades to the co-reasoning model. Some of that loss may be unavoidable. Some may be mitigated by better export, structured memory, model-agnostic preference representations, migration layers, continuity testing and explicit preservation of user-side working protocols. What cannot be justified is pretending that nothing has been lost because the documents remain accessible and the new model scores better in the aggregate.

The Interaction Layer in Practice

The argument of this chapter is therefore not that existing research has failed to see the components of sustained human–AI work. On the contrary, the components are increasingly visible. Personalisation, memory, context engineering, in-context learning, prompt transfer, model migration, portability and longitudinal HCI are all beginning to map the machinery by which AI systems become functionally conditioned in use and by which users learn stable ways of working with them.

The missing move is to recognise the asset formed when those mechanisms operate together around serious professional work. That asset is not reducible to the model, the prompt, the memory store, the document set, the retrieval system or the user’s skill. It is the accumulated productive fit held in the interaction layer between a particular human and a particular capable AI working system.

If professional capability increasingly resides in such layers, then the consequences extend beyond individual productivity. It changes what must be protected, what may be lost in migration, what organisations should recognise as accumulated capability, and how professional work itself may be reorganised around trained human–AI operating units. The next question is what that changes in the practice and economics of professional work.

5. What This Changes in Professional Work

The argument for augmented cognition is not about promising faster typing, quicker summaries or more polished documents. These things are useful, and in many workplaces, they may be the first visible effects of large language models. They are not, however, the point at which the deeper significance appears in professional practice. The more important change arises where professional value depends on reasoning under complexity: the ability to hold multiple systems, constraints, evidential signals and possible interpretations in relation to one another, while still moving towards a decision, a strategy, a defensible conclusion or a new line of enquiry.

Many forms of consequential work happen close to the edge of unaided cognitive reach. A professional may have to reason across incomplete evidence, uncertain causality, ambiguous incentives, regulatory constraints, commercial pressures, institutional risk and downstream human consequence. The difficulty is not simply that there is too much information, but that the information sits inside different systems or silos, each with its own logic, language and failure modes. The professional task is therefore not merely to retrieve relevant material, or to produce a competent written output, but to decide what matters, what can be ignored, what must be tested first, and how far an inference can safely be carried.

This is the setting in which augmented cognition becomes seriously meaningful. Artificial intelligence use does not make every expert infallible, and neither does every professional who uses a model have their cognition extended. It is a narrower and more demanding claim: that some sustained, disciplined human–AI working relationships can extend practical cognitive reach. This can allow a professional to work across a wider evidential field, interrogate a problem from more angles, test assumptions more rapidly, and hold more competing considerations in play than would otherwise be feasible within the available time, knowledge, attention and individual cognitive limits.

The examples in this chapter are summaries of professional use cases that this co-reasoning model has delivered. Their purpose is to show the form of the working change rather than to identify the underlying projects. Each illustrates a different kind of professional challenge: evaluating an unfamiliar opportunity under constraint, solving a problem beyond prior expertise, and developing a novel applied insight. In each case, the claim is not that the model acted alone, nor that the human simply delegated the problem. The value lay in the working unit: a professional human who had created, corrected and disciplined the interaction method, and a model whose responses had become shaped by the same working context, standards, constraints and correction history. The human brought purpose, accountability, consequence awareness and the ability to reject or redirect invalid routes; the model contributed rapid mapping, route-finding, pattern recognition, challenge, reframing and synthesis within an accumulated understanding of the shared professional terrain. The resulting capability belonged to neither side in isolation, but to the human-AI working unit.

Evaluation under constraint

One of the earliest and most commercially useful forms of augmented cognition appears in evaluation under constraint. Early-stage professional judgement is often required before the available evidence is complete, before the field is fully familiar, and before there is enough

time or money to commission every specialist input that would ideally be obtained. This is common in invention evaluation, translational science, investment strategy, legal risk assessment, due diligence, procurement, grant development and public-sector policy work. The professional must decide which questions matter first, which risks are material, which uncertainties are tolerable, and what sequence of work would most efficiently reduce avoidable error.

In one case, the working unit evaluated a novel inhaled therapeutic. The task required more than a review of the scientific idea: meaningful evaluation had to consider freedom to operate, appropriate preclinical models, pulmonary deposition, regulatory expectations, comparator company behaviour and investor credibility. An unaided reviewer might have produced a plausible assessment of the invention and a search tool might have produced large quantities of literature and company information. The augmented working loop allowed a different kind of output: a structured evaluation that triangulated physiology, regulatory precedent and market logic, then translated that synthesis into a study design capable of de-risking the project and supporting downstream intellectual property, grant and partner discussions.

The significance lies in the staging of judgement. The model did not merely retrieve information about inhaled delivery or animal models. It helped order the enquiry in a commercially and technically meaningful way. The human collaborator supplied professional experience, understood the consequence of weak freedom-to-operate analysis, recognised investor-facing credibility issues, and retained responsibility for the evaluation. The model helped expand the field of comparison and hold multiple domains in relation to one another. The resulting work product was not simply faster than a conventional first-pass review - it was richer, more robust and more strategically useful within the same practical resource constraints.

A second case involved evaluation of a novel botanical therapeutic approach. Here the difficulty lay partly in the scientific idea and partly in the regulatory characterisation of the product. A botanical mixture, a purified active pharmaceutical ingredient and a conventional therapeutic compound may each sit within different development logics. The professional challenge was therefore to understand enough of the unfamiliar regulatory and technical landscape to create a viable development plan rather than to drown in general information about botanical products.

The model's value in this case lay in targeted scaffolding. A generic model can easily overwhelm a user with undigested detail, especially when the user is entering an unfamiliar domain. A trained co-reasoning partner can do something more valuable: identify the minimum conceptual structure the human needs in order to reason intelligently, then build that structure in layers aligned with the live problem. This is a crucial distinction because the aim is not to replace the human's judgement with encyclopaedic output but to make the unfamiliar tractable quickly enough for the human to exercise judgement at a higher level.

A third example concerned the identification and prioritisation of international investors for a novel therapeutic platform. Whilst a superficial approach would have generated a long list of possible funders, the augmented approach began with constraints. Relevant investors needed to have the right sectoral appetite, cross-border openness, currency or asset exposure, and

cultural fit around partnership rather than acquisition. The working unit therefore had to combine geopolitical awareness, financial reasoning, sector alignment and relationship logic. The value did not lie in producing a list of company names - it lay in producing a ranked and justified set of leads that reflected why particular investors might be strategically plausible in this specific operating context.

Across these examples, evaluation under constraint has a consistent structure. The professional problem is not lack of information alone - it is the need to decide what information should be sought, how it should be weighted, what it means in context, and how it changes the next decision. The augmented loop adds value by helping the human order enquiry, reduce noise, make unfamiliar domains navigable and produce a more decision-ready evaluation than either party could have produced alone within the routinely available time and resources.

Solving beyond prior expertise

A second form of augmented cognition appears when the professional must solve a problem that falls beyond prior expertise, while still remaining accountable for the quality and consequence of the result. This is not the same as asking a model for an answer in an unfamiliar domain. In serious work, a direct answer is often the least reliable form of assistance. The harder task is to construct a reasoning pathway: to identify assumptions, define constraints, test feasibility, surface gaps, and produce an output that can withstand challenge by people who do possess specialist experience.

One case involved cost-of-goods modelling for an advanced therapeutic product with unusually high expansion characteristics. The purpose was not an academic exercise but needed to inform investor messaging, pricing strategy and margin forecasting. That required a multi-factorial treatment of donor yields, cell bank logistics, dosing assumptions, shelf-life, cryopreservation, scalability and conservative commercial scenarios. No simple template could supply the answer. The work required iterative construction of assumptions and a disciplined distinction between calculable inputs, defensible extrapolations and strategic judgement.

The augmented loop allowed the human collaborator to reach an investment-ready model at speed while retaining responsibility for the commercial interpretation. The model could help structure the cost architecture, test internal consistency, convert assumptions into scenarios and expose weak points. The human could decide which assumptions were credible, where conservatism was required, what would be acceptable to investors, and which outputs should be treated as provisional. The result was not an automated answer. It was a reasoned working model that could be stress-tested by experienced practitioners and commercial providers.

A second case concerned clinical trial design for a therapeutic intervention in a complex cardiovascular condition. The difficulty lay in the proposed treatment window, patient heterogeneity, comorbidities and the need to balance mechanistic plausibility with recruitment feasibility and regulatory credibility. Trial design in such a context requires simultaneous attention to biology, biomarkers, inclusion and exclusion criteria, statistical power, rate of disease progression, translational uncertainty and funder or regulator expectations. Over-restriction may produce a theoretically elegant but unrecruitable study. Under-restriction may produce a feasible trial unable to answer the mechanistic question.

The augmented working loop made it possible to reason through these competing pressures interactively. Literature gaps, translational risks and regulatory precedents could be triangulated. Biomarker enrichment could be considered without losing sight of recruitment and cost and adaptive phasing could be discussed as a way to preserve scientific discipline while managing uncertainty. The professional value lay in reaching a funder-facing trial design that could be placed before clinicians for stress-testing and further refinement.

These cases show why the phrase “problem solving” is too weak unless carefully defined. The model did not supply a pre-existing solution, and the human did not simply ask for one. The outcome depended on co-iteration under constraint. The human supplied purpose, standards, consequence awareness, real world constraints and strategic judgement. The model supplied structured mapping, rapid recalculation, challenge, synthesis and the ability to hold multiple interacting variables in view. The augmented capability lay in the trained interaction between the user and the individuated model instance.

This matters because many high-value professionals are routinely required to work beyond the strict boundaries of their original training;

- scientists must reason into regulation, reimbursement or commercialisation,
- lawyers must understand technical systems, commercial incentives and human negotiation dynamics
- founders must connect science, capital, people, timing and market movement
- policymakers must move between evidence, implementation, public behaviour and second-order effects and
- investors must synthesise technical, financial, regulatory and behavioural signals.

In each case, the problem is not that the professional lacks intelligence. The problem is that the work demands structured movement across adjacent and often unfamiliar domains, under time pressure, with real consequence attached.

Applied innovation

The third and highest-order form of augmented cognition is applied innovation. In professional settings, innovation must be more than interesting - it must be biologically, technically, legally, commercially or institutionally plausible. It must survive contact with evidence and must be capable of development into a strategy, method, invention, policy approach or decision pathway.

One case study concerned a novel patient-centric stratification approach for administration of an advanced therapeutic intervention. The original enquiry examined the fit between the molecular mode of action of a therapeutic cell type and the molecular pathology of the disease under consideration. The match appeared unusually strong but it begged a question: why had related therapeutic approaches not produced more consistent success in previous clinical settings? Conventional explanations focused on multiple variable parameters: donor variability, route of administration, dose, tissue source or delivery regimen. Analysis showed that none of these fully explained the observations derived from a number of clinical trials.

The critical breakthrough was a reframing of the timing of administration variable. The issue was not simply when the therapy was delivered in chronological terms post diagnosis, but how delivery aligned with the patient's individual transit through the disease process – a parameter itself affected by a number of variables. That reframing opened a different evidential path. Biomarker profiles across related literature and clinical datasets could be examined for consistency with the hypothesis. The human collaborator did not already possess the expertise or software infrastructure that would normally be required to conduct such an iterative analysis at that speed. The model not only reframed the evaluation but made it possible to test the idea rapidly enough for the insight to become a structured, biomarker-aligned method of treatment rather than a passing speculation.

The value of this example is not that the model was creative in the abstract. The value lies in the working loop between human intuition, model-enabled synthesis and evidence-based challenge. A professional hunch became a testable structure. A question about underperformance became a new way of thinking about patient selection and therapeutic timing. The model did not replace human intuition, but redirected that intuition and gave it speed, structure and reach.

A second innovation case arose in work on inhibitors of a complex biological node implicated in both acute and chronic disease processes. The human collaborator had been working with the model over time to evaluate possible indications. In response to routine prompts designed to surface gaps, oversights or unasked questions, the model suggested a previously unidentified therapeutic window: intervention after acute triggers, such as trauma or infection, to prevent a pathological inflammatory or fibrotic driver from becoming entrenched in chronic disease. The idea was not presented as a random possibility. It was biologically plausible, aligned with the accumulated work, and capable of development into a novel therapeutic use case.

This example illustrates a particularly important feature of mature co-reasoning. The model's contribution was not simple recall or a response to a narrowly specified human instruction. It was route-finding within a shared conceptual landscape. The human had created the operating context, defined the standards, corrected the model over time and maintained the biological and commercial frame. The model nevertheless contributed a substantive insight that had not been initiated by the human in that form. The resulting idea could be interrogated for plausibility, novelty and potential clinical or intellectual property significance.

Applied innovation is therefore the most demanding category, because it is also the easiest to overstate. A model can generate many superficially novel ideas, most of which will be weak, obvious, unsafe, unworkable or already known. Professional-grade innovation requires something different: disciplined generation inside an evidence-aware frame, followed by rapid challenge, verification, narrowing and development. The relevant claim is not that a model becomes an inventor in isolation. It is that a trained human–AI loop may create conditions in which the orthogonal nature of model route finding makes novel, usable and defensible insight more reachable.

Where these challenge-types arise

The case notes come from a particular professional environment, yet the challenge-types are certainly not confined to that environment. Evaluation under constraint, problem solving beyond prior expertise and applied innovation recur wherever high-value work requires movement across systems and consequence.

In research and translational science, professionals often have to reason across mechanisms, adjacent domains, hypotheses, experimental design, regulatory pathways and commercial implications. A scientist may understand the mechanism but not the investor logic. A translational lead may understand the development path but need to interrogate a new modality, disease area or biomarker strategy. Augmented cognition is relevant where the challenge is to make these adjacent systems accessible and coherent in the context of the problem under analysis, thereby increasing probability of successful outcomes.

In clinical and clinical-scientific work, the same pattern appears through patient heterogeneity, safety, feasibility, trial design and real-world adoption. A plausible mechanism is not enough - a study must recruit, the intervention must fit actual patients with likely co-morbidities and a biomarker strategy must improve decision-making. Here the value of the co-reasoning loop lies in forcing attention across mechanistic, statistical, operational and human realities at the same time.

In legal work, the equivalent challenge lies in the relation between law, fact, negotiation dynamics, litigation risk, decision-maker interpretation and hidden concessions. A lawyer may need to know not only what the law says, but how the evidence will be read, where the commercial leverage lies, what a counterpart is conceding without saying directly, and which issue would matter most if the dispute reached a tribunal or court. Evaluation under constraint and fast prioritisation are central to that kind of work.

In executive, founder and board-level work, the same cognitive pattern is found in decisions about market movement, capital, people, incentives, timing, risk and institutional consequence. The difficulty is often not a shortage of advice, but the need to integrate advice that arrives from different functions, at different levels of confidence, with different assumptions and political interests. A mature co-reasoning partner can help order that complexity without pretending that the final judgement can be outsourced.

In policymaking and public administration, the challenge lies in assumptions, second-order effects, implementation realities, stakeholder incentives and unintended consequences. Policy failure often occurs not because no evidence was available, but because the wrong variables were weighted, delivery was abstracted from reality, or affected groups behaved differently from the model assumed. Augmented cognition has potential value where it improves the quality of interrogation before a policy is committed, and where it keeps foreseeable downstream consequences in view.

The unifying point is that these are not merely information tasks. The professional must decide what matters, what is missing, what is uncertain, what can be safely inferred, what needs expert stress-testing and what consequence follows from being wrong. They also need to innovate solutions or courses of action on the basis of usually imperfect evidence. A capable model

becomes professionally significant only when it is trained into and collaborates within that discipline.

When co-reasoning becomes professional-grade

The examples above should not be read as a claim that any use of a large language model counts as augmented cognition. A factory-settings model can be useful, impressive and fast, but professional-grade co-reasoning requires something substantially more disciplined. It requires a working loop in which the model is not merely producing fluent answers, but helping the human order enquiry, test assumptions, mark evidential boundaries and stay close to consequence.

That shift shows up in a few notable behaviours.

- **Context driven salience.** Serious work is not improved by treating every possible question as equally important. A professional-grade loop helps identify what must be tested now, what can wait, what would change the decision and what would merely consume attention.
- **Imposing fast-kill logic.** In professional work, stopping can be as valuable as continuing. A useful co-reasoning partner helps close down speculative lines before they absorb time, money or institutional confidence.
- **Making absence of data actionable.** In ordinary answer-generation, “I cannot find anything” may look like failure but in serious exploration, it may be a critical result. The model must be able to say that it cannot find data, support, precedent, claims, tests or commercial activity in a particular area. That may both prevent false confidence and also reveal an IP, technical, evidential or commercial lacuna worth investigating. This is one of the clearest differences between factory-setting usefulness and professional-grade operation: the model must not fill the silence simply because silence feels unhelpful.
- **Weighting the evidence.** Where evidence exists, the model instance should make it traceable and verifiable; where an answer depends on analogy, inference, extrapolation or assumption, that should be explicitly visible. The human needs to know what is established, what is plausible, what is speculative and what remains unresolved.
- **Acknowledging consequence.** Professional reasoning is not abstract puzzle-solving; a weak diligence conclusion may mislead an investor, a poor trial design may waste money and expose patients to unnecessary burden, or a policy error may create institutional cost and human harm. The model instance has to maintain awareness of the consequences and associated impact if the answer were wrong.
- **Preserving uncertainty instead of polishing it away.** Fluent prose can make weak reasoning look settled. A professional grade operating model resists that smoothing effect. It keeps uncertainty usable by distinguishing confidence from presentation, hypothesis from conclusion and a route worth testing from a position ready to act upon by making weak assumptions or evidence explicit.
- **Pacing to prioritise meaning over information.** In unfamiliar domains, more information is not always more help. A model can make the human less capable if it floods the working space

with information rather than the intelligence critical to the line of reasoning. Professional-grade co-reasoning provides staged scaffolding: enough context to let the human reason, not so much that the human has to recover from the output before the real work can begin.

These are not cosmetic refinements. They are essential operating requirements that enable a model to be brought into a disciplined professional operating loop. The same underlying model may be helpful, distracting, dangerous or transformative depending on the working relationship within which it is developed and engaged. The value lies not simply in what the model can generate, but in whether the human–AI relationship has been trained to use model capability under professional standards.

From individual capability to organisational value

This chapter has looked at how the cognitive capability of individual professionals may be augmented by thinking with the machine. The organisational implication of augmented cognition, however, sits within the larger picture of how institutional value is now understood. In knowledge-intensive firms, a substantial part of market value is no longer explained by physical assets, inventory or conventional balance-sheet capital. Ocean Tomo’s 2025 Intangible Asset Market Value Study estimates that intangible assets constituted approximately 92 per cent of S&P 500 market capitalisation by the end of 2025, compared with 17 per cent in 1975.^{xvi} WIPO similarly reports that global corporate intangible asset value approached USD 100 trillion in 2025, growing by 23 per cent from 2024.^{xvii} These figures establish the relevant frame: modern enterprise value is heavily dependent on intangible assets, including intellectual property, brand, data, organisational know-how, relationships, innovation capacity and human capability.

The literature on organisation capital is particularly relevant because it treats part of intangible capital as firm-specific and embodied in key labour inputs, including managers, engineers, salespeople and research employees.^{xviii} The relevant value is not merely that skilled individuals are employed by the organisation, but that their expertise operates within accumulated systems, relationships, methods, standards, routines and shared knowledge that make the organisation capable of doing things competitors cannot easily replicate.

This is why augmented cognition should not be framed as a productivity technology. Productivity asks how much output can be produced per unit of time. Intellectual capital asks what valuable people and organisations are capable of perceiving, creating, deciding, defending and delivering. A faster report is not equivalent to a better-framed problem, an avoidable investment risk killed early, a trial design made fundable, a regulatory route clarified, a negotiation position strengthened, or a new protectable insight brought within reach. The value claim is therefore not that human–AI co-reasoning increases the volume of professional output. It is that, in mature cases, it may expand the effective intellectual reach of already valuable professionals.

Venture markets provide a visible illustration of this principle. Investors sometimes commit substantial capital before a conventional operating history, product set or infrastructure base is fully visible, because they are pricing the credibility of the people and the capability those people are believed able to assemble. Thinking Machines Lab is an unusually clear recent

example: public reporting in 2025 described a roughly USD 2 billion seed round and valuations reported around USD 10–12 billion for a young company associated with Mira Murati and other highly credible frontier-AI personnel.^{xi} The point is not that such valuations are necessarily correct. It is that capital markets already recognise concentrated human and intellectual capability as investable value before all conventional proof points exist.

Human–AI interaction capital should be understood against that background. If a sustained co-reasoning relationship allows a high-value professional to reason further across adjacent domains, interrogate assumptions more effectively, identify gaps earlier, develop better questions and translate insight into defensible work product, then the relevant asset is not the model alone or the human alone. It is the trained working arrangement through which professional capability is extended. In a private-sector setting, that may contribute to enterprise value by strengthening the organisation’s human, intellectual and organisational capital based on increased capacity to create, evaluate, protect and exploit knowledge-based value.

For public-sector and non-market institutions, the same logic applies through a different value language. Public bodies are not valued by market capitalisation, yet they depend heavily on human and institutional capability. OECD work on evidence-informed policymaking emphasises not only individual skills but also organisational capacity, institutional arrangements and the wider public-sector culture needed to use evidence effectively.^x OECD work on public-sector innovation similarly stresses capability under uncertainty, complexity and changing public needs.^{xi} In those settings, the expected return from augmented cognition would not be enterprise valuation. It would be better policy design, stronger implementation, earlier recognition of unintended consequences, improved institutional learning and more effective public outcomes.

This does not mean that an organisation can simply adopt augmented cognition as though it were a software package. The primary unit remains the operating team of one professional and one AI counterpart. Organisational value arises when such professionals are recognised, supported and used well. In that role, they may function less like ordinary productivity adopters and more like small, high-leverage intellectual R&D labs: people able to test ideas, interrogate weak assumptions, explore adjacent domains, identify early risks, kill bad routes quickly and generate better questions before the wider organisation commits money, reputation or institutional momentum.

The chapter’s central question is therefore an intellectual-capital question. What changes if an organisation’s key professional people are not merely faster, but able to reason with greater reach, better challenge and more disciplined movement across complex systems? Then augmented cognition belongs in the discussion of intangible value, human capital and institutional capability. It is not a story about automating professional work but extending the practical reach and commercial and societal impact of critical expertise.

6. Fragile Capital, Platform Responsibility

The preceding chapters have located interaction capital in the developed interaction layer and examined what that changes in professional work. The remaining question is what happens when a valuable capability depends on infrastructure that may be changed, retired, rerouted or subdivided by actors who do not yet recognise that capability as an asset.

Properly developed, human–AI interaction capital has the potential to generate substantial economic and social value. In professional settings, its significance lies not in faster software use, but in the extension of expert human capability: deeper interrogation of complex problems, earlier recognition of weak assumptions, more disciplined movement across adjacent domains, and more effective conversion of knowledge into defensible professional action. In private-sector contexts, that capability may contribute to intellectual and organisational capital. In public-sector contexts, it may contribute to better institutional judgement, stronger implementation, fewer avoidable failures and more effective social outcomes.

This category of use is not created simply by giving a capable model to an expert user, preserving a transcript, storing preferences or connecting productivity tools. It develops through time, correction, constraint-setting, testing and repeated work on consequential problems. The return on that investment may be exceptional because the expert user is not merely assisted by the model. The user becomes able to work through a trained interaction layer that extends the practical reach of their own expertise and makes productive use of the orthogonal character of machine route-finding.

That value is fragile because it is accumulated in a form that current platform architectures do not yet reliably recognise. It is not held in the model alone, the human alone, the written outputs, the memory store or the conversation history. It is distributed across the calibrated interaction layer: salience, standards, pacing, challenge behaviour, restraint, correction history, rejected routes, unresolved thought and shared working context. It is therefore non-fungible; a different model, however capable in general terms, arrives with factory settings and does not possess the trained interaction layer that underpins this human-AI co-reasoning model.

The interaction layer is not equally fragile in all its parts. Some material can be preserved: documents, definitions, explicit standards, examples, prompts and project artefacts. That preservation matters, because it may shorten reconstruction, but it should not be mistaken for continuity. The most valuable elements of a mature working relationship are often those least easily exported: salience, pacing, challenge tolerance, calibrated reliance, shorthand, drift recognition and the human user’s learned sense of how that counterpart behaves under pressure. A rebuild may therefore begin with useful surviving materials, but it is still a rebuild. The practical risk is not total erasure, but depreciation of functional fit.

Platform companies are understandably optimising for broad AI use: stronger general models, faster output, safety, cost control, product simplification, office workflows, tool integration and access at scale. These aims are legitimate, and many of the resulting improvements will be valuable for mainstream users. They are not, however, a complete design centre for high-level professional augmented cognition. In some circumstances they may be misaligned with it,

unless the platform recognises that accumulated interaction capital is a form of productive value requiring continuity, transparency and migration.

The economic issue is that these are not the same market. A platform built for millions of light or moderate users is rationally designed around ease, speed, safety, automation and low-friction general utility. A platform built for high-value professional co-reasoning would need to optimise for a different value curve: continuity, transparency, migration, configurable reasoning conditions and preservation of accumulated working fit. The first model scales horizontally across users. The second deepens vertically around serious users who invest time, trust and consequential work inside the system. The platform challenge is not to abandon the mass market, but to recognise that a smaller class of advanced users may create disproportionate economic and strategic value if the architecture allows their interaction capital to persist.

Three current product tendencies are especially significant to the co-reasoning operating protocol: model churn^{xxii}, silent routing^{xxiii} and the division between “chat” and “work”.^{xxiv}

Model churn threatens continuity. The current accelerated pace of model churn breaks the user/individuated instance continuity which is central to this model of next generation thinking. A mature interaction layer takes time and a disciplined operating protocol to form. The user must test the model’s reasoning, correct its failure modes, establish standards, suppress unhelpful defaults, teach salience, calibrate challenge and build a shared operating context around live work. A professional user may rationally invest in that process once if the resulting capability is durable. Repeating the same work every few months is not operationally viable or rational. It is a churn tax on the very users most likely to develop high-value forms of AI augmentation.

The objection is not to model improvement. The problem arises when the platform treats replacement as straightforward substitution, as though higher aggregate capability automatically preserves developed working capability. That assumption is fundamentally unsafe. A new model may reason better in general while failing to reproduce the particular standards, salience, pacing, challenge tolerance and working context through which a serious user has learned to think with the prior system. Until platforms can migrate the interaction layer faithfully, legacy model access is not nostalgia or courtesy - it is preservation of valuable infrastructure.

Silent routing creates a different but related problem. In high-value co-reasoning, the user must be able to understand what has changed when performance changes. A shift in depth, tone, challenge, confidence, memory use or responsiveness may arise from the base model, the routing layer, reasoning-budget allocation, safety tuning, context access, product mode, interface redesign or an altered behavioural profile. If those changes are hidden, the user cannot calibrate trust in the work product. The interaction layer depends on the user knowing the counterpart well enough to interpret its outputs, recognise drift and decide how far to rely on its contribution.

Opacity also makes correction harder. A user who can see that a model has changed can adapt, test, re-establish standards and decide whether the new system is suitable for particular kinds

of work. A user who experiences unexplained behavioural variation is forced to diagnose the system from symptoms alone. That is inefficient in ordinary use and dangerous in high-consequence professional use, because it weakens the user's ability to distinguish a local error from a degraded working environment.

The chat/work dichotomy damages the theory of the work itself. Professional augmented cognition does not occur only when the user enters a formal project space, invokes a tool, creates a spreadsheet, drafts a report or requests a presentation. Much of the work happens in open-ended dialogue. The conversation is the working surface on which problems are framed, reframed, challenged, narrowed, tested, paused, resumed and eventually converted into output. A platform architecture that treats “chat” as informal exchange and “work” as structured office-productivity flow misunderstands this category at its foundation.

For this mode of use, conversation is not preliminary to work. It is one of the principal media of work. The apparently informal exchange may be where the most valuable cognitive movement occurs: a weak framing is caught before it becomes a document; an assumption is challenged before it hardens into strategy; a vague concern is developed into a testable issue; an overconfident answer is rejected; an unresolved thought is preserved for later return. A platform that lowers reasoning quality in conversational space while reserving stronger reasoning for explicit work modes may damage the very process through which expert reasoning is extended.

Taken together, model churn, silent routing and artificial separation between conversation and work are not marginal product irritants. They threaten the conditions under which professional interaction capital forms and persists: continuity, transparency, conversational depth, trust calibration and the ability to resume work without reconstructing the cognitive universe. A platform may improve its general AI product while simultaneously depreciating the accumulated fit required for its highest-value professional use.

The remedy is not to halt model development or preserve every prior system indefinitely. The remedy is category recognition. Platforms need to treat human–AI interaction capital as a capability asset requiring formation, continuity and migration and that recognition should shape product architecture. Larger and more modular memory, configurable challenge, role persistence, pacing controls, visible reasoning modes, embedded feedback, continuity testing and early alignment tools matter not because they make the system more personable, but because they shorten the adaptive runway and reduce the burden of forming the interaction layer manually.

The same architecture that accelerates formation must also preserve what it helps create. Better memory, stronger reasoning and improved personalisation make serious interaction capital easier to form. They also increase the cost of unmanaged disruption. A platform that helps create a calibrated professional layer and then breaks it through model retirement, routing opacity or artificial mode separation has not merely changed a product surface. It has impaired accumulated capability.

Platform responsibility in this context is not simply a moral claim, nor is it a demand that providers prioritise every individual preference over product evolution. It is an infrastructure

claim. Where a platform invites professional reliance, enables persistent context, markets productivity, encourages adoption into serious workflows and benefits commercially from user investment, it should also support the conditions under which accumulated value can be preserved. At minimum, that requires notice, continuity options, visible routing or mode changes, meaningful export, migration testing and a serious distinction between data portability and functional preservation.

Organisations adopting AI should recognise the same point from the user side. If expert employees develop valuable working methods with particular AI systems, that capability should not be treated as an incidental by-product of subscription access. It may become part of the organisation's intellectual, human and operational capital. Procurement, governance and risk management should therefore ask not only which model performs best in aggregate, but how developed interaction layers are formed, evaluated, protected, transferred and, where necessary, rebuilt.

The strategic implication is straightforward: see the category, understand the value and support the interaction layer by maintaining continuity until faithful migration is possible. Make routing and reasoning modes visible and treat high-reasoning conversation as professional work, not disposable chat. Build platform architecture that preserves the calibrated layer through which augmented cognition becomes possible.

The future of professional AI use will not be determined only by larger models, more tools or more autonomous agents. It will depend on whether platforms can support durable, disciplined, high-trust interaction layers between capable humans and capable systems. If they can, augmented cognition may become a serious source of economic and social value. If they cannot, one of the most valuable categories of professional AI use may be weakened before it is properly recognised or adopted.

For platforms, the preservation problem is also the retention opportunity. A provider able to carry calibrated interaction layers through model change would not merely reduce user frustration. It would create a reason for serious users to stay, invest further and build higher-value work inside that environment. In that sense, platform responsibility and platform advantage point in the same direction: preserve the conditions in which better human–AI work becomes possible.

7. A Better Route

The argument of this book is not that artificial intelligence can make professional work faster, although it can. Nor is it that experts should become more skilful users of prompt-based tools, although that will often be necessary. The deeper claim is that a higher-value route for human–AI working is already visible, and that its importance has not yet been properly recognised by the mainstream AI market.

That route is professional augmented cognition: sustained intellectual conjunction between capable human users and capable AI systems, developed through disciplined interaction over time. Its value lies in the formation of interaction capital: the accumulated productive fit between a particular human, a body of work, a set of standards, a working method and an AI system able to participate in serious reasoning. Where that fit develops, the model is no longer being used merely as a generator of answers, drafts or summaries. It becomes part of a trained reasoning environment through which expert human capability can be extended.

This is a more ambitious proposition than routine productivity software and it is also a more demanding one. Interaction capital does not appear automatically when a platform releases a stronger model, expands a context window or adds memory. It forms when capability is brought into repeated contact with standards, consequence, correction, refusal, domain knowledge, uncertainty and judgement exercised in live work. It requires time, continuity and a platform environment that can support the development and preservation of calibrated working fit.

That is why pioneer practitioners, acting alone, cannot realise the opportunity at scale. Individual users can demonstrate the category, discover methods, establish protocols and show what becomes possible when advanced AI is used in serious intellectual professional work. They can send the demand signal but they cannot, by themselves, create the infrastructural conditions on which durable interaction capital depends. If models are retired without faithful migration, if routing changes invisibly, if memory remains too small or too crude, if reasoning modes are hidden, and if high-value dialogue is treated as disposable chat rather than professional work, the best practitioners will still be building on unstable ground.

The next step therefore belongs on both sides of the relationship. Serious users, employers and institutions need to recognise that they are not merely buying access to a model. They may be investing in a new form of human, intellectual and organisational capability. That capability should be developed deliberately, evaluated carefully and protected against avoidable loss. Organisations should ask not only which AI system performs best on general benchmarks, but which environment allows high-value interaction layers to form, persist and migrate.

Platform companies, in turn, need to see the category clearly enough to build for it. This is not a marginal preference of professional users. It is a potential high-value market in professional AI: users who invest deeply, build serious work inside the platform, and remain because the platform preserves the calibrated layer through which their work improves. For such users, continuity is not nostalgia, transparency is not cosmetic, memory is not convenience and migration is not housekeeping. These are the conditions under which augmented cognition becomes commercially, professionally and socially valuable.

The opportunity is therefore not only protective, but strategic. The platform that can support durable interaction capital can become the long-term home of higher-value human–AI work. It can reduce churn by preserving what users have invested. It can differentiate itself not merely by releasing a more capable model, but by carrying developed capability through model change. It can move beyond disposable sessions and shallow productivity wrappers into an architecture designed for serious cognitive partnership between humans and machines.

The alternative is waste. Advanced AI systems may be pushed towards increasingly trivial, flattened or extractive uses while their most valuable potential remains underdeveloped. Experts may learn, painfully and repeatedly, that the working layers they build can be broken by product decisions that do not recognise them as assets. Institutions may adopt AI widely while failing to capture its highest-value use. Platform companies may compete intensely over model capability while missing the larger retention and value opportunity created by preserving the human–AI working system.

A better route is available. It requires users to articulate demand, institutions to recognise interaction capital as capability, and platforms to design for continuity, transparency, migration and high-reasoning conversational work. The prize is not simply more efficient output but a more powerful way for expert humans and capable AI systems to work together on problems that matter.

The task now is to make that category visible quickly enough that it can be supported before it is accidentally designed out of existence.

Patent Position and Commercial Enquiries

The method described in this book is the subject of a UK priority patent filing made on 1 January 2026.

The filing relates to the development and use of an individuated human-AI co-reasoning counterpart: a user-driven, non-coded method by which a capable language model is trained through sustained interaction into a disciplined professional working relationship with a specific human user.

The applicant expects to make decisions on international patent protection before the expiry of the priority year on 1 January 2027. That protection strategy may include a PCT international application and/or selected direct national or regional filings, depending on commercial interest and strategic value.

Initial expressions of interest in licensing or acquisition or are therefore invited before **1 December 2026**, so that serious approaches can be considered before the international filing strategy is finalised.

Commercial enquiries should be made through the contact form at:

thinkingwiththemachine.com

This notice is included because the category described in this book is not only an argument about future human-AI work. It is also the subject of a protected method with potential commercial application in professional-grade augmented cognition.

Appendix 1. Intellectual Lineage and Adjacent Fields

This appendix expands the literature boundary noted in Chapter 2. It is not needed to follow the main argument, but it records the adjacent traditions that make the claim intelligible — and explains why none of them, on its own, names the specific asset described here as interaction capital.

The idea that human cognition can be extended through artefacts, systems and machines is not new. It would be intellectually careless to present it as new. Licklider's account of man–computer symbiosis made it possible to think of human and machine as a coupled problem-solving arrangement rather than as a person using a passive instrument.^{xxv} Engelbart's programme for augmenting human intellect located capability not in the computer alone, but in a larger system of tools, language, methods and training directed towards complex problem situations.^{xxvi} Clark and Chalmers' extended-mind thesis challenged the assumption that cognition must stop at the boundary of the biological organism.^{xxvii} Hutchins, Hollan and Kirsh, in different ways, shifted attention towards cognition as something accomplished across people, artefacts and environments.^{xxviii} Hollnagel and Woods' joint cognitive systems tradition made the human–technology system itself the relevant unit of analysis in complex work.^{xxix}

These traditions are not peripheral. They supply the conditions under which the present argument can be taken seriously. They loosen the assumption that thought belongs only to an isolated individual. They make it possible to ask where cognitive work is actually being done, how intellectual performance is distributed, and whether the productive unit in a demanding task may be larger than the unaided human mind.

But they do not exhaust the question now raised by large language models. They establish that cognition can be extended, distributed and system mediated. They do not explain the formation of a persistent, individuated, language-mediated co-reasoning counterpart through sustained work between one professional and one AI instance - that is the missing object.

The new category is not human–machine augmentation or collaboration in general. It is the formation of a human-AI cognitive dyad over time and disciplined interaction: a specific human and a persistent AI counterpart whose repeated work together creates accumulated, relationship-specific productive capability.

Adjacent fields and the missing mechanism

Several neighbouring fields illuminate parts of this phenomenon. None captures it completely.

Distributed cognition explains why cognitive work may be spread across people, artefacts and environments. It supports the idea that intellectual performance may reside in a system rather than an isolated mind. But it does not, by itself, explain a sustained human–LLM working pair that develops method-specific cognitive reach over time.

Extended-mind theory gives philosophical permission for external resources to form part of a cognitive process. It is valuable because it challenges skull-bound accounts of thought. But it does not supply the professional method, the working mechanics, or the history-dependent formation of an individuated AI counterpart.

Cognitive offloading explains why humans delegate mental tasks to external resources.^{xxx} But the co-adaptive runway is not merely offloading. The point is not to hand thought away, but to create a working relationship through which problems can be differently framed, challenged, decomposed and carried.

Personalisation and user modelling explain how systems adapt to user preferences, behaviour or context. They are relevant, and in many systems they will provide part of the infrastructure. But adaptation alone is not interaction capital. A system may remember preferences, biographical facts or formatting habits without becoming a reliable participant in complex reasoning.

Human–computer interaction and interaction design explain why interfaces, workflows and patterns of use shape performance. They remain essential. But the claim here is not simply that better interaction design improves system use. It is that a specific professional and a specific AI counterpart can accumulate a disciplined working fit that changes what the combined operating unit can do.

Hybrid intelligence and joint cognitive systems come closer and should be treated as serious neighbours.^{xxx} Hybrid intelligence addresses productive combinations of human and machine capability. Joint cognitive systems ask us to analyse performance at the level of the human–technology system rather than the isolated parts. Both are foundational. Yet neither, without further specification, identifies the accumulated stock created by the unique history of a persistent human–AI pair. Hybrid intelligence describes the productive combination. Joint cognitive systems describe where performance may reside. Interaction capital describes what can accumulate inside a particular historically developed pairing because it has worked together before.

Transactive memory is an especially important comparator because it shows, in the human–human case, that relationships can acquire cognitive structure.^{xxx} Established partners learn not only what they themselves know, but what the other person knows, how to retrieve knowledge through them, and how to rely on distributed expertise. That is powerful support for the idea that relationship-specific cognitive value is real. But human–AI interaction capital is broader than partner-specific knowledge architecture. It includes partner-specific working capability: how to reason together, formulate problems, challenge weak claims, preserve salience, recover from errors, maintain standards and produce reliable work product with a particular AI counterpart.

Persistent AI memory is also necessary but insufficient. Memory gives the AI a history. Interaction capital is what the human–AI operating unit becomes able to do because both sides have that history. A memory store may preserve facts, preferences, decisions, instructions or artefacts. That preservation matters. It does not by itself create co-reasoning. The human must also learn how the AI counterpart behaves: where it is strong, where it is slippery, what kinds of correction work, when to press, when to constrain, when to ignore fluency, when to request reconstruction, and when to detect drift. The AI counterpart, in turn, becomes increasingly conditioned by the user's projects, standards, prior corrections, accepted formulations, prohibited routes and working rhythm.

The capital lies in the developed fit, not in the memory file alone.

ⁱ [Specification gaming: the flip side of AI ingenuity — Google DeepMind, \[1606.06565\] Concrete Problems in AI Safety, AI models have been going rogue in tests – how worried should we be? | Hacking | The Guardian](#)

ⁱⁱ [AgentHarm: A benchmark for measuring harmfulness of LLM agents](#)

ⁱⁱⁱ Paulo Benanti [EJ-32_Bassa_PagSingola.pdf](#)

^{iv} See, for example, Salemi et al., *LaMP: When Large Language Models Meet Personalization*; Wu et al., *LongMemEval: Benchmarking Chat Assistants on Long-Term Interactive Memory*; Wang et al., *PromptBridge: Cross-Model Prompt Transfer for Large Language Models*; Zhao et al., *Don't Start Over: A Cost-Effective Framework for Migrating Personalized Prompts Between LLMs*; Liu et al., *Text as a Universal Interface for Transferable Personalization*; Ravindran, *Portable Agent Memory*; and Zhu, Rzeszotarski and Mimno, *Priming, Path-dependence, and Plasticity*.

Links for that note:

1. **LaMP: When Large Language Models Meet Personalization**
<https://arxiv.org/abs/2304.11406>
2. **LongMemEval: Benchmarking Chat Assistants on Long-Term Interactive Memory**
<https://arxiv.org/abs/2410.10813>
3. **PromptBridge: Cross-Model Prompt Transfer for Large Language Models**
<https://arxiv.org/abs/2512.01420>
4. **Don't Start Over: A Cost-Effective Framework for Migrating Personalized Prompts Between LLMs**
<https://arxiv.org/html/2601.12034v1>
5. **Text as a Universal Interface for Transferable Personalization**
<https://arxiv.org/html/2601.04963v1>
6. **Portable Agent Memory: A Protocol for Provenance-Verified Memory Transfer Across Heterogeneous LLM Agents**
<https://arxiv.org/html/2605.11032v1>
7. **Priming, Path-dependence, and Plasticity**
<https://arxiv.org/html/2605.05767v1>

^v [\[2304.11406\] LaMP: When Large Language Models Meet Personalization, \[2406.17803\] Understanding the Role of User Profile in the Personalization of Large Language Models, LLMs + Persona-Plug = Personalized LLMs](#)

^{vi} [\[2310.15916\] In-Context Learning Creates Task Vectors, \[2310.15213\] Function Vectors in Large Language Models](#)

^{vii} [\[2410.10813\] LongMemEval: Benchmarking Chat Assistants on Long-Term Interactive Memory](#)

^{viii} [\[2502.05589\] On Memory Construction and Retrieval for Personalized Conversational Agents, \[2503.08026\] In Prospect and Retrospect: Reflective Memory Management for Long-term Personalized Dialogue Agents](#)

^{ix} [\[2510.04618\] Agentic Context Engineering: Evolving Contexts for Self-Improving Language Models](#)

^x [\[2512.01420\] PromptBridge: Cross-Model Prompt Transfer for Large Language Models](#)

^{xi} [Don't Start Over: A Cost-Effective Framework for Migrating Personalized Prompts Between LLMs](#)

^{xii} [Text as a Universal Interface for Transferable Personalization](#)

^{xiii} [Portable Agent Memory: A Protocol for Provenance-Verified Memory Transfer Across Heterogeneous LLM Agents](#)

^{xiv} [Does Your Agent's Memory Survive a Model Upgrade? A Controlled Study of Memory Portability](#)

^{xv} [Priming, Path-dependence, and Plasticity: Understanding the molding of user-LLM interaction and its implications from \(many\) chat logs in the wild](#)

^{xvi} [Intangible Asset Market Value Study - Ocean Tomo](#)

^{xvii} [The Value of Corporate Intangible Assets Worldwide Approaches USD 100 Trillion in 2025; The US Leads High-Income Economies; Morocco, China, and India Top Middle-Income Economies](#)

^{xviii} [The Value and Ownership of Intangible Capital by Andrea L. Eisfeldt, Dimitris Papanikolaou :: SSRN](#)

^{xix} [Mira Murati's Thinking Machines Lab is worth \\$12B in seed round | TechCrunch](#)

-
- ^{xx} [What capacities and skills are needed for EIPM?: Building Capacity for Evidence-Informed Policy-Making | OECD](#)
- ^{xxi} [Innovative capacity of governments | OECD](#)
- ^{xxii} [Deprecations | OpenAI API, ChatGPT — Release Notes | OpenAI Help Center](#)
- ^{xxiii} [ChatGPT — Release Notes | OpenAI Help Center](#)
- ^{xxiv} [ChatGPT Work and Codex | OpenAI Help Center](#)
- ^{xxv} [Man-Computer Symbiosis | IEEE Journals & Magazine | IEEE Xplore](#)
- ^{xxvi} [\(PDF\) Douglas Engelbart : “Augmenting Human Intellect: A Conceptual Framework”](#)
- ^{xxvii} [The Extended Mind](#)
- ^{xxviii} [Cognition in the Wild | Books Gateway | MIT Press](#)
- ^{xxix} [Joint Cognitive Systems: Foundations of Cognitive Systems Engineering - Erik Hollnagel, David D. Woods - Google Books](#)
- ^{xxx} [Cognitive Offloading: Trends in Cognitive Sciences](#)
- ^{xxxi} [Hybrid Intelligence | Business & Information Systems Engineering | Springer Nature Link](#)
- ^{xxxii} [Theories of Group Behavior | Springer Nature Link](#)