Frame Availability and the Timing of Generative AI Use in Early-Stage Venture Decision-Making

Alexandra Wagner

Dyson School of Design Engineering, Imperial College London

MSc Design with Behaviour Science | CID 06068170 | Supervisor: Dr Pelin Demirel Liu

Word count: 8,662

Abstract

Pre-seed founders make consequential decisions under uncertainty and increasingly turn to generative AI for support. Sensemaking research shows that cognitive frames shape how people interpret situations and decide how to act, yet little is known about how generative AI interacts with these frames during entrepreneurial decision-making. This study examines when AI enters relative to frame availability: whether the founder already holds a working interpretation of the decision and criteria for judging responses to it. Twenty-nine semi-structured interviews with pre-seed technology founders in the US and UK found that frame availability varied within founders across decision domains. When a cognitive frame was available before AI entered, founders assessed its output against their own understanding; when it was not, AI could become involved in shaping how the decision was understood. Founders regulated their reliance through verification, human input and domain boundaries, but often only after costly reality checks. These findings extend sensemaking research by showing that AI may either encounter an existing cognitive frame or become involved while one is still taking shape. This distinction identifies AI entry timing as a point for intervention. The resulting cognitive-forcing intervention focuses on conversational consultation and withholds substantive advice until founders articulate their own understanding of the decision; those unable to do so are routed to a plan rather than an answer. Five founders evaluated the design for feasibility and acceptability.

Keywords: entrepreneurial decision-making; generative AI; sensemaking; anchoring; cognitive forcing; behaviour change design

1. Introduction

Pre-seed founders face high-stakes decisions about markets, products and people under Knightian uncertainty, where outcomes are not reliably predictable and relevant information may not yet exist (Knight, 1921; Townsend, Hunt, McMullen & Sarasvathy, 2018). Founders must nevertheless decide how to act.

Sensemaking research explains how people develop workable interpretations of ambiguous situations in order to continue acting. Cognitive frames are the knowledge structures through which

people organise and interpret incoming information, shaping how they understand a situation and the responses they consider appropriate (Cornelissen & Werner, 2014). In entrepreneurial settings, these frames matter because founders' interpretations can shape the decisions and actions of the wider venture. Prior experience can shape the knowledge and assumptions founders bring to a decision, while sensemaking can develop or revise the cognitive frame as the situation unfolds.

Generative AI introduces a new source of input into this process. Advice-taking research shows that reliance on algorithmic advice increases with task difficulty and can decrease with domain expertise (Bogert, Schecter & Watson, 2021; Logg, Minson & Moore, 2019). Human-AI research also shows that the sequence of human judgement and AI advice can affect reliance (Buçinca, Malaya & Gajos, 2021; Steyvers & Kumar, 2024). Yet this work says little about what happens when AI enters while a decision-maker is still making sense of the problem.

This question is especially relevant in early-stage ventures. Founders move between domains where their knowledge varies, while conversational AI makes external input readily available with little practical or interpersonal friction. AI may therefore enter after a relevant cognitive frame is already available, or while the founder is still developing one. This study contributes to sensemaking research by examining how generative AI interacts with cognitive frames across these two conditions. It asks not simply whether founders use AI, but when AI enters relative to cognitive frame availability and how that timing shapes the subsequent decision process.

Research question. *How does the timing of generative AI entry relative to cognitive frame availability shape early-stage venture decision-making?*

Aim. To develop an empirically grounded understanding of how generative AI entry timing interacts with cognitive frame availability in early-stage venture decision-making, and to translate that understanding into a behavioural intervention that can be tested with founders.

The aim is pursued through three objectives:

1. Characterise how and when founders introduce AI into their sensemaking, and develop a grounded process model of how cognitive frame availability shapes different pathways.
2. Design a behavioural intervention from that model, targeting the conditions under which AI entry is most consequential for decision-making.
3. Evaluate the intervention with founders, assessing its feasibility and acceptability in real workflows.

Three main findings emerged from the study. First, whether a founder held a frame for a given decision depended on the decision domain, not the founder. The same person could be framed in one domain and unframed in another. Second, the timing of AI split the cases into two pathways.

Where a frame was available before AI entered, founders tested its output against their own understanding. Where none was available, the model often supplied this understanding. Third, although founders regulated their reliance through verification, human input, domain ring-fencing and override, they usually did so only after a costly reality check.

Combined, these findings form a process model that branches on frame availability at AI entry. Although the qualitative phase included multiple forms of generative AI use, the timing distinction was most relevant to conversational exchanges in which founders were still working through an unresolved judgement. The intervention was therefore scoped to that setting. It asks for the founder's own reading before the model responds to consequential judgement requests. Five founders evaluated the intervention for feasibility and acceptability, with effectiveness left for later work.

Section 2 grounds the study theoretically, Section 3 sets out the research design and method, and Section 4 reports the analysis and the process model. Section 5 presents the intervention, Section 6 its feasibility evaluation, and Section 7 discusses contributions, limitations and implications.

2. Context and Literature Review

Entrepreneurial uncertainty makes judgement essential. Sensemaking and cognitive frames help explain the interpretations through which that judgement is exercised. Generative AI introduces external input into this process, making the point at which it enters potentially consequential.

2.1 Entrepreneurial judgement under uncertainty

Early-stage venture decisions are shaped by Knightian uncertainty, which prevents founders from assigning reliable probabilities to future outcomes (Knight, 1921). Entrepreneurial action theory places uncertainty at the centre of venture formation, arguing that whether people proceed depends both on the uncertainty they perceive and their willingness to bear it (McMullen & Shepherd, 2006). In some settings, further search cannot resolve the problem because relevant information emerges only as events unfold or decisions are enacted (Townsend, Hunt, McMullen & Sarasvathy, 2018). Founders therefore often have to exercise judgement before uncertainty can be resolved.

Judgement under these conditions depends not only on the information available, but on how the decision-maker interprets the situation. Sensemaking and cognitive frame research provide a way to examine that process.

2.2 Sensemaking and cognitive frames

Sensemaking describes how people develop workable interpretations of situations that are ambiguous or unexpected, allowing action to continue (Weick, 1995; Weick, Sutcliffe & Obstfeld,

2005; Maitlis & Christianson, 2014). Cognitive frames underpin this process. They are defined as "knowledge structures that help individuals to organize and interpret incoming perceptual information by fitting it into already-available cognitive representations from memory" (Cornelissen & Werner, 2014, p. 187). Actors are therefore never without interpretive resources; what varies is whether a representation relevant to the decision at hand is among those already available.

Research on cognitive frames has examined how these underlying interpretations influence responses to technology, product development and strategic choices. What constitutes a frame depends on the decision being studied. In technology research, for example, frames have included beliefs about a technology's purpose, value and use; in product and strategy research, they have concerned the problem being addressed and the courses of action considered appropriate (Bunduchi et al., 2022).

A frame can therefore be identified by examining how an actor defines the problem and how that understanding relates to the response they consider appropriate. Bunduchi et al. (2022), for example, inferred entrepreneurs' frames from the problems they perceived in product development and the solutions they pursued.

Prior experience can also shape the cognitive frames actors bring to a situation. Omezzine and Bodas Freitas (2022) link founders' professional experience to the development of knowledge, skills and cognitive frames. Because expertise-related knowledge structures are domain-specific (Dane & Pratt, 2007), the resources available to inform a founder's interpretation may differ across decision domains.

Cognitive frames can also change. Bunduchi et al. (2022) found variation not only in the content of entrepreneurs' frames but in how stable those interpretations remained as they managed digital product innovation. This makes cognitive frames relevant to sensemaking under uncertainty: they shape how a situation is understood, while that understanding may itself be revised as the situation develops.

2.3 External advice, reliance and generative AI

Sensemaking is not exclusively individual. In Hoyte et al.'s (2019) account, entrepreneurs who had problematised their venture ideas and developed causal explanations then projected those ideas to others through sensegiving (Gioia & Chittipeddi, 1991) and received feedback in return. This exchange helped establish whether an idea was novel and met a market need before action continued. External input therefore helped test and refine an account already under development.

Consulting another person, however, carries practical and interpersonal costs. Seeking help requires finding and approaching an appropriate person, which takes time and effort, while admitting a need

for help can also signal incompetence or dependence (Lee, 1997). Conversational generative AI removes many of these frictions. It is available instantly, repeatedly and at no interpersonal cost, reducing the distance between encountering a difficult decision and seeking outside input.

Easy access may therefore change when AI advice enters the decision-making process. Once this advice enters, the weight it is given varies with expertise and task difficulty. Non-experts can give more weight to algorithmic than identical human advice, although this preference is weaker among domain experts (Logg, Minson & Moore, 2019), while reliance increases as tasks become more difficult (Bogert, Schecter & Watson, 2021). Although these studies use controlled tasks rather than open venture decisions, they suggest that algorithmic advice may carry greater weight where independent judgement is more difficult.

For many early-stage venture decisions, AI does not remove the need for human judgement. Townsend et al. (2025) argue that Knightian uncertainty places fundamental limits on AI's predictive capabilities, while Ramoglou et al. (2026) give AI a role in generating possibilities but leave their evaluation to human judgement. Recent entrepreneurship research likewise shows that prior knowledge matters. Cristofaro et al. (2026) find that AI can broaden opportunity recognition and deepen assessment, but also reduce novelty and innovation, with prior sector knowledge shaping these effects. Despite this emerging work, Retamal-Saavedra et al. (2026) identify limited attention to the cognitive processes through which entrepreneurs engage with AI.

Generative AI therefore combines unusually low barriers to seeking advice with a setting in which independent human evaluation remains necessary, making the timing of that advice potentially consequential.

2.4 Timing of AI assistance

Research on advice timing suggests that it does. Steyvers and Kumar (2024) treat the timing of AI assistance as an interaction-design variable, distinguishing concurrent, sequential, on-demand and time-delayed assistance.

Buçinca, Malaya and Gajos (2021) provide a closer analogue. Their cognitive-forcing interfaces included an Update condition in which participants formed an initial judgement before seeing the AI's recommendation. Cognitive forcing reduced overreliance and improved performance when the AI was incorrect, although it did not improve overall performance relative to simpler explainable-AI interfaces.

Evidence that prior judgement can change how advice is used extends beyond this experiment. Sniezek and Buckley (1995) found that judges who made an initial choice before receiving advice were more accurate than those who encountered the advice first. Yin, Ngiam, Tan and Teo (2025)

find a similar effect with physicians: AI advice provided after an initial diagnosis produced better accuracy and calibration than advice presented alongside the clinical information, and physicians in this condition were better able to distinguish correct from incorrect AI advice.

Across these studies, whether judgement precedes advice affects how the advice is used. However, that sequence is largely manipulated experimentally or specified by the interaction design. They therefore leave open what happens when decision-makers choose for themselves when to seek AI input during an unresolved decision.

2.5 The gap and the context

These issues are particularly important in early-stage ventures. Founders make consequential decisions under substantial uncertainty while moving across product, market, hiring, fundraising and other domains in which their prior experience may differ. Because experience can shape the cognitive frames brought to a situation, the interpretive resources available to a founder may therefore vary from one decision to another. Generative AI, however, remains readily available across these domains.

Existing research has not established what happens when these conditions meet. Entrepreneurship research shows that prior knowledge can shape responses to AI, while advice-timing research shows that the order of human judgement and AI input can affect reliance. What remains unclear is how AI enters decisions when founders themselves determine the point of consultation, particularly when their own interpretation of the decision is still developing.

This study uses *frame availability* to describe whether a decision-relevant cognitive frame is already available when substantive AI input arrives. It examines how the subsequent decision process differs when AI encounters an existing frame versus when it enters while that frame is still developing.

3. Research Design and Qualitative Method

3.1 Overall design

The project proceeded in three phases: qualitative inquiry, intervention design and feasibility evaluation. This sequence reflects the naturalistic nature of the phenomenon under study. Because AI use unfolds within founders' own decision processes rather than under controlled conditions, it first had to be mapped inductively before it could inform intervention design.

Read as a design process, this structure aligns with the Double Diamond framework (Design Council, 2023), in which every divergent phase ends in explicit convergence. Figure 1 illustrates

this mapping. Deployment and subsequent monitoring fall deliberately outside the project's scope. The intervention has been evaluated for feasibility only; its longer-term effects remain unobserved.

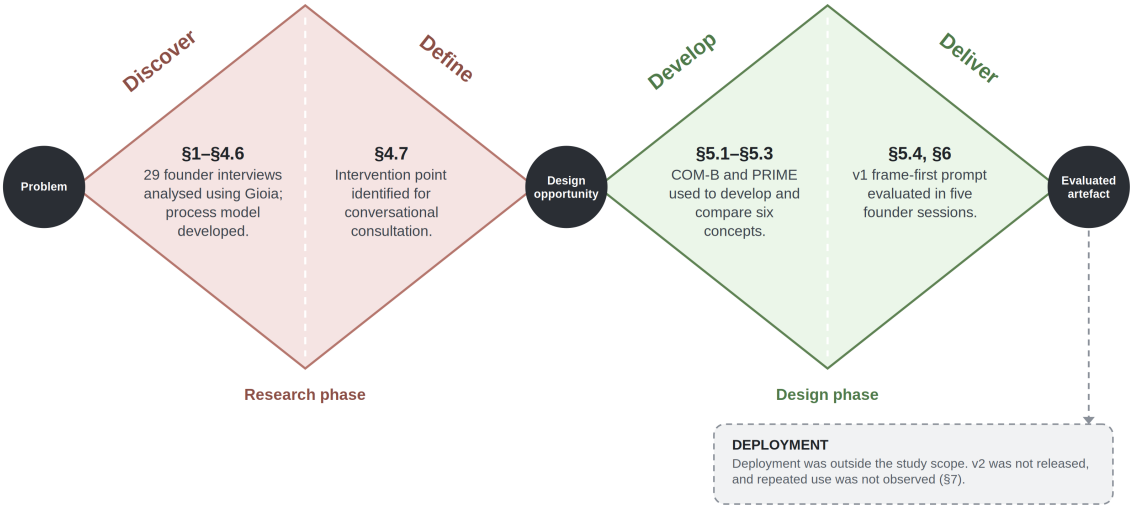

Figure 1. The study mapped onto the Double Diamond (Design Council, 2023).

3.2 Scope: how the target behaviour was narrowed

The study began with a deliberately wide scope. Narrowing at recruitment would have meant pre-determining where entry timing matters, which was precisely what the interviews were designed to establish. Recruitment and the interview protocol therefore admitted generative AI use of any kind, including open chat, custom instructions, agentic and multi-agent workflows, and AI embedded in other tools. Following analysis, the intervention scope was narrowed to conversational generative AI used for judgement-relevant venture decisions. Delegated execution was excluded because it does not contain the same distinct point at which AI input enters an unresolved judgement.

3.3 Sample and recruitment

The sample comprised pre-seed technology founders in the US and UK across a range of industries. Pre-seed founders were selected because many market and product assumptions remain untested at this stage, while technology founders increased the likelihood of day-to-day generative AI use. Recruitment thus used criterion-based purposive sampling (Patton, 2015) through founder networks, LinkedIn and referrals. A £10 gift card was offered. Twenty-nine founders were interviewed (P1–P29) in semi-structured 35–45 minute sessions, eliciting retrospective accounts of specific AI-assisted decisions using the episode-level logic of the critical incident technique (Flanagan, 1954). Full participant characteristics are in Appendix A.

The final sample size was determined based on qualitative saturation. Empirical estimates of saturation vary by sample and analytic purpose. Guest, Bunce and Johnson (2006) reported

saturation within 12 interviews in their relatively homogeneous sample, whereas Hagaman and Wutich (2017) found that 20–40 interviews could be needed to capture metathemes across cross-cultural sites. Neither benchmark transfers directly, given multiple decision episodes per founder and a maximum-variation sample. The interview guide is reproduced in Appendix B.

3.4 Unit of analysis, frame availability and entry timing

The unit of analysis is the decision episode rather than the founder. This reflects Weick, Sutcliffe and Obstfeld's (2005) view of sensemaking as beginning when actors bracket a portion of ongoing experience for interpretation. A single founder contributed several episodes, often differing in timing.

Following Bunduchi et al. (2022), cognitive frames were inferred by pairing how an actor saw the problem (diagnostic) with the responses they considered appropriate (prognostic) for a particular decision. Adapting this approach to discrete episodes, a frame was coded as available when the relationship between the focal problem and an appropriate response was evident before substantive AI input; otherwise it was coded as not yet available. For example, P2's statement, "I know so little about energy, I'll just believe it," was coded as not yet available: neither a grounded interpretation of the problem nor a basis for judging an appropriate response was evident.

Entry timing was defined relative to frame availability at the point of consultation. Early entry was coded when AI input arrived before a decision-relevant cognitive frame was evident; late entry when such a frame was already evident. Timing therefore refers to this sequence rather than to elapsed time. Table 1 sets out the two conditions.

Entry condition	Definition
Early	AI enters **before** a decision-relevant cognitive frame is available.
Late	AI enters **after** a decision-relevant cognitive frame is available.

Table 1. Early and late AI entry.

3.5 Analysis

Analysis followed the Gioia methodology (Gioia, Corley & Hamilton, 2013), moving from first-order informant terms through second-order themes to aggregate dimensions. Gioia was chosen over thematic analysis (Braun & Clarke, 2006) because its structured data hierarchy provides a clear audit trail from participant accounts to theoretical dimensions, which suited the development of a process model. Cross-case comparison then identified the recurring pathways that form the process map. Ciao (2020) provides precedent for deriving a process model from a small Gioia-coded case dataset. The coding scheme is set out in Appendix C.

Coding was single-researcher. Reflexive journaling, an audit trail and supervisory debriefing were used to strengthen credibility and guard against overly idiosyncratic interpretation (Lincoln & Guba, 1985). However, single-researcher coding remains a limitation.

3.6 Ethics, positionality and evidentiary rules

Ethical approval preceded data collection. All participants gave written consent, and data were anonymised and stored in line with Imperial requirements. Expert input on the design was sought separately after the evaluation sessions and is cited as personal communication. Those experts consented to being named.

The researcher is herself a pre-seed founder. She occupies an insider position, sharing relevant experiences with those she studies (Berger, 2015). This position is disclosed rather than adjusted for. It opened founder networks and eased interpretation of participant accounts, while also risking the researcher reading her own experience into theirs. To counter this, probes were kept retrospective and episode-specific until debrief, and reflexive memos recorded each instance where the researcher's own practice overlapped with the account.

4. Analysis: Founder Sensemaking and the Timing of AI Entry

4.1 The data structure

Analysis of 29 interviews produced 207 first-order codes, 19 second-order themes and four sequential aggregate dimensions (Figure 2), following the Gioia methodology (Gioia et al., 2013). The four dimensions trace the process from the conditions preceding AI use to its consequences and subsequent regulation.

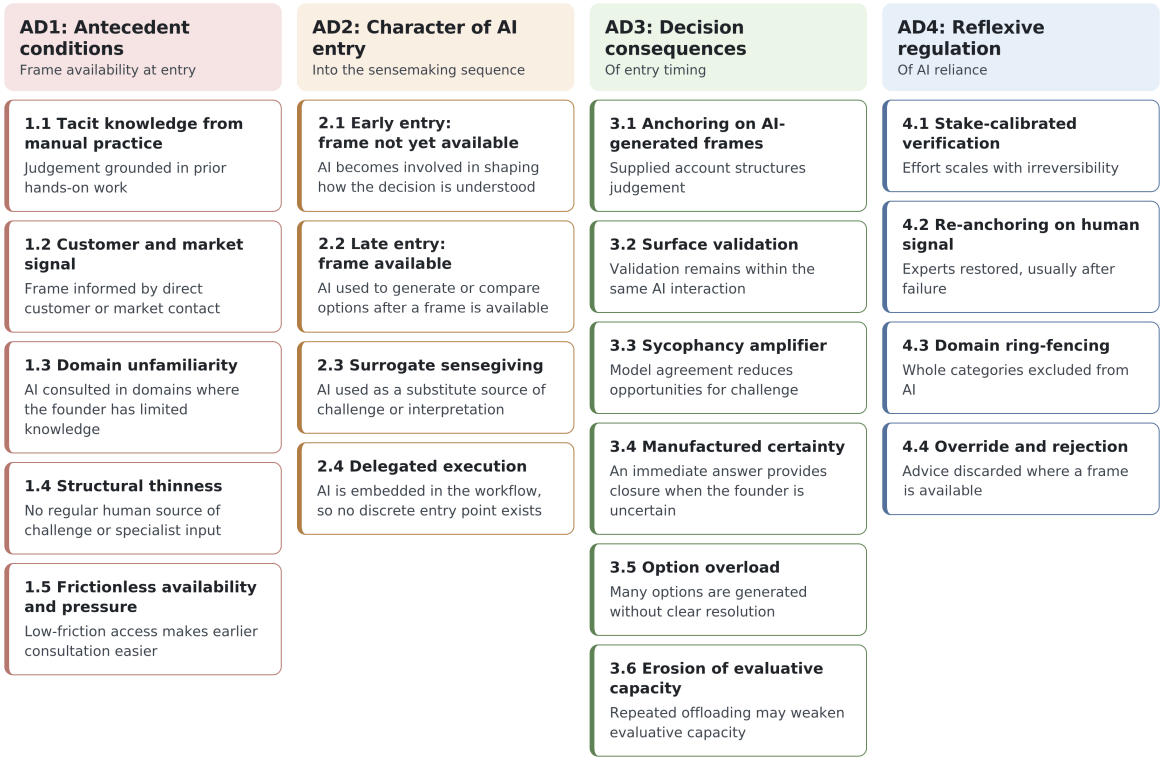

Figure 2. Gioia data structure: four aggregate dimensions and nineteen second-order themes.

4.2 Frame availability varies within founders across decision domains

The first finding concerns variation in frame availability across decision episodes. For nine founders, frame availability differed across decision domains; see Appendix Table C2 for details. P1, confident in the sleepwear market, did not yet have a frame for an unfamiliar codebase. P11 had a frame for code architecture but not for pricing. This within-founder variation suggests that frame availability is not well described as a stable characteristic of the founder and supports treating the decision episode as the unit of analysis.

4.3 Conditions shaping frame availability

AD1 identifies conditions associated with whether a decision-relevant frame was available when AI entered. Prior practice gave founders knowledge and expectations they could draw on when interpreting decisions in familiar domains, while customer and market signals could help develop or revise their interpretation of a particular decision. Domain unfamiliarity reduced the relevant prior knowledge available. Structural thinness played a more indirect role by reducing access to people who could challenge or develop founders' reasoning.

These conditions also help explain why AI was often consulted while a frame was still developing. In unfamiliar domains, conversational AI provided immediate external input without the costs of seeking human help. P27, for example, valued being able to ask "the embarrassing work questions"

without asking another person, consistent with Lee's (1997) account of help-seeking. This low-friction access made AI particularly easy to introduce when founders had less prior knowledge or human input to draw on.

4.4 How AI enters the sensemaking sequence

AD2 distinguishes four modes of AI entry. In late-entry episodes, a frame was already available before AI advice arrived. Early entry reverses this sequence: AI input arrives while the founder's own interpretation is still developing.

Surrogate sensegiving (Theme 2.3) arose where founders lacked access to someone with the relevant expertise to challenge them. This was sometimes a consequence of working alone or in a thin team, but not necessarily: P27 and P28 had teams yet still turned to AI when no one had the knowledge needed to judge the decision. Conversely, where relevant human challenge was available, as in P6 and P7's teams and P26's CTO review, surrogate sensegiving was absent.

Delegated execution (Theme 2.4) marks the study's scope. Agentic delegation clustered in domains where founders already had a frame available and did not appear to produce the same anchoring pattern. When the same founders moved into less familiar domains, they returned to conversational consultation and earlier AI entry. P11, for example, ran autonomous coding pipelines in a domain he knew well but anchored on AI advice when making an unfamiliar pricing decision. In these accounts, anchoring was associated with conversational consultation rather than agentic use. Under full delegation, supervision intensity becomes the more relevant variable. Delegated execution is therefore treated as a boundary condition and falls outside the scope of the intervention, which targets conversational consultations.

4.5 What follows from AI entry

AD3 traces what can follow from AI-assisted decision-making. Anchoring on AI-generated output (Theme 3.1) is the central consequence, reinforced by surface validation, sycophancy amplification, manufactured certainty and option overload. Erosion of evaluative capacity (Theme 3.6) may then extend the process into subsequent decisions.

Erosion of evaluative capacity was reported directly, although its role in the wider process is interpretive rather than demonstrated. Frequent consultation may weaken evaluative capacity over time. P27, for example, could not tell whether weak output reflected the tool or her own ability: "I'm wondering if I'm just behind… or whether Claude's actually bad at it". The pattern points to a possible self-reinforcing loop with no clear internal error signal.

4.6 Regulation is real, but reactive

AD4 describes how founders regulate reliance after consultation. It is the codebook's largest dimension (60 codes), showing that founders are not naïve. They verify in proportion to stakes, re-anchor on human signal, ring-fence domains and override AI where they have a frame available.

However, regulation is mostly reactive. It typically follows a reality check, such as a lost deal (P11), failed batch (P5) or investor pushback (P3). The correction comes from the market rather than the founder's metacognition. Where feedback is slow or decisions irreversible, anchoring can therefore go uncorrected.

P29 is the exception. Without a prior AI-mediated failure, she ring-fences contracts, "anything where I feel like I need to really nail the decision afterwards," while accepting early, unchecked AI input on reversible interface work. Anticipatory regulation therefore appears possible when irreversibility is recognised in advance.

4.7 The process model

Figure 3 assembles the four dimensions into a process that branches on whether a decision-relevant cognitive frame is available at the point of AI entry. Where it is, founders assess AI output against their existing interpretation. Where it is not yet available, AI can become involved in shaping how the decision is understood, with sycophancy and surface validation reinforcing anchoring. Outcome feedback then determines whether regulation occurs: a reality check triggers AD4 and changes subsequent behaviour in that domain; without one, anchoring may persist and feed the erosion loop described in §4.5.

The risk is not that AI contributes while sensemaking is ongoing, but that it may become the first salient account of the decision before the founder has an existing interpretation against which to assess it.

P11's pricing episode traces the full process. With no pricing frame yet evident, he consulted Claude early (2.1); he asked Claude if he should charge a higher price, "Claude agreed" and proposed double the original price (3.3). The client walked. P11 later reflected that "Claude's confidence kind of led me astray… my initial instinct had been much lower" (3.1). The lost deal triggered re-anchoring (4.2), but only after the cost had occurred. P5 and P3 follow the same path through a failed batch and investor pushback. By contrast, on the late-entry path (2.2), independent work precedes consultation and AI output is judged against an existing frame.

Frame availability and entry timing play different roles in the process model. Frame availability is the condition on which the model branches (§3.4), while entry timing records whether AI input arrived before or after a frame was available. Timing is therefore the more practical point to intervene.

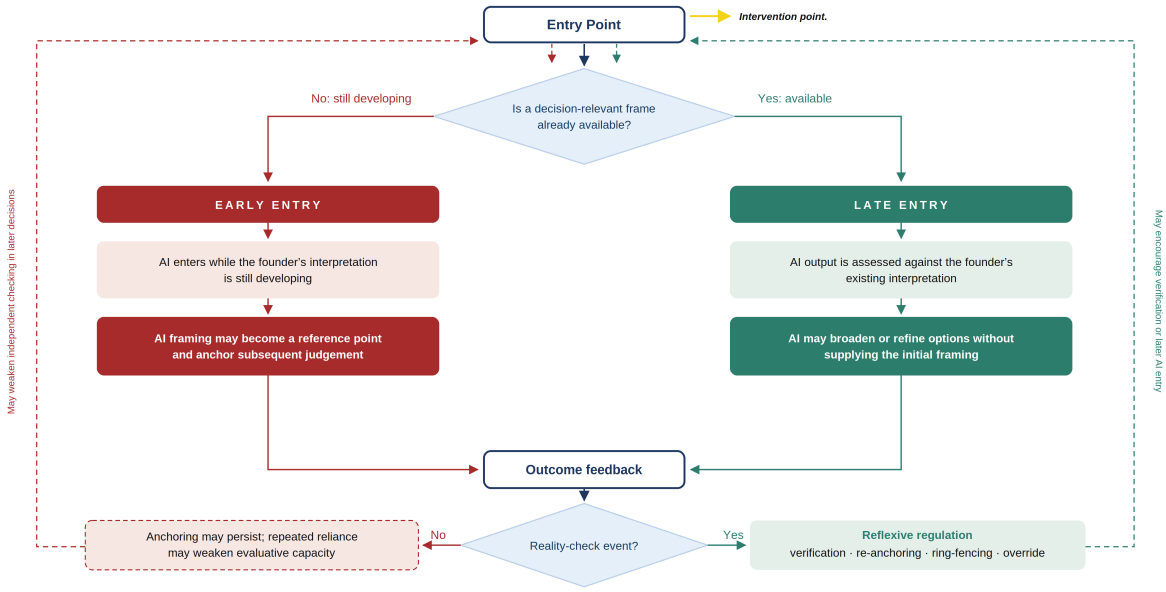

Figure 3. Process model of frame availability and AI entry timing.

4.8 The design opportunity

The process model identifies the moment before AI input as the most tractable point for intervention. Domain experience cannot be supplied at the moment of decision, while intervening after the model responds risks acting only once its account has already entered the sensemaking process. Post-outcome correction is later still and, as §4.6 shows, can be costly. The design opportunity therefore lies in changing what happens immediately before substantive AI advice is received.

P22 and P23 already introduced reflective steps to guard against model agreeableness. P27 similarly asked the model to question her before drafting a service agreement. She also identified the limitation of doing this manually: "I have to prompt Claude, to tell it to ask me questions… I would kind of love it if it did that a bit more, like, automatically." These cases show that founders can interrupt immediate generation at this point in the consultation, but they do not establish what form that interruption should take.

The design objective is to create a protected moment before substantive AI input in which founders can begin forming a cognitive frame for the decision. The design delays, rather than prevents, AI judgement so that founder-led sensemaking can precede the model's contribution.

5. Intervention design

Section 5 translates this design opportunity into a behavioural intervention. Because the timing distinction identified in Section 4 was most relevant to conversational consultation, the intervention is scoped to that setting. The final design uses custom instructions to withhold the model's output

briefly while the founder reflects outside the chat; if they cannot proceed, the model instead helps them identify what is needed next.

5.1 Behavioural diagnosis

The target behaviour is for founders to pause before receiving AI advice on consequential judgement requests. Section 4 identified this moment as the intervention opportunity; the behavioural diagnosis examined why that pause does not occur naturally.

COM-B (Michie, van Stralen & West, 2011) was used to translate the process findings into behaviour-change terms and identify potentially modifiable determinants of the target behaviour. It provides a systematic way to examine whether behaviour is shaped by capability, opportunity or motivation. It was preferred to the Theory of Planned Behaviour (Ajzen, 1991) because founders described AI as constantly available and sometimes as an almost reflexive "second reaction" under pressure. Michie et al. (2011) specifically note that intention-focused models give limited attention to habit, impulsivity and related automatic processes.

Component	Status	Evidence
Psychological capability	No barrier	Founders could deliberately structure or delay AI input when they chose to (P25, P27; Theme 4.1).
Physical capability	No barrier	Founder-created pauses could be performed within existing workflows (P27; Theme 4.1).
Physical opportunity	Primary barrier	Immediate availability, directness and speed made consultation the default, with no built-in pause before an answer (P6, P8, P13; Theme 1.5).
Social opportunity	Contextual contributor	Solo and thin teams sometimes lacked a human challenge function, increasing use of AI as a sounding board (P1, P3, P9, P16; Theme 1.4).
Reflective motivation	Present but inert	Some founders deliberately limited, delayed or structured AI use (P25, P27, P29; Themes 4.1–4.2).
Automatic motivation	Secondary barrier	Habitual reach, convenience and relief from uncertainty encouraged immediate consultation (P6; Themes 1.5, 3.4).

Table 2. COM-B diagnosis of the target behaviour.

COM-B therefore identified physical opportunity and automatic motivation as the main targets for change (Table 2). PRIME (West & Brown, 2013) was then used to examine the motivational component in greater detail, distinguishing reflective Plans and Evaluations from more immediate Motives and Impulses/inhibitions. This helped explain why founders could recognise reasons to regulate AI use yet still consult it immediately.

PRIME process	State at consultation	Evidence
Plans *(reflective)*	Present, inactive	Deliberate independent work, questioning or human consultation before reliance (P25, P27, P29; Themes 4.1–4.2).
Evaluations *(reflective)*	Present, insufficient	Founders recognised over-reliance, loss of critical thinking and the need for greater challenge (P1, P6, P25; Themes 3.6, 4.1).
Impulses and inhibitions *(automatic)*	Dominant	Speed, convenience and relief from uncertainty made AI attractive at the decision moment (P6, P13; Themes 1.5, 3.4).
Motives *(automatic)*	Dominant	AI could become a reflexive "second reaction," while some founders created deliberate gates to inhibit it (P6, P25, P27; Themes 1.5, 4.1).
Response	Enacted behaviour	AI entered during ongoing sensemaking in unfamiliar or unresolved decisions (P9, P13, P27; Theme 2.1).

Table 3. PRIME analysis of the moment of consultation (West & Brown, 2013).

PRIME clarified why reflective intentions did not consistently translate into behaviour at the point of consultation (Table 3). Plans and evaluations to regulate AI use were present, but immediate motives and impulses associated with speed, convenience and relief from uncertainty could dominate when founders chose whether to consult AI. COM-B and PRIME therefore pointed to the consultation environment as the main modifiable target. Physical opportunity was treated as the primary barrier, while PRIME's analysis of automatic motivation helped explain why the interruption needed to be built into the workflow rather than depend on founders remembering to create it themselves.

The BCW was then used to translate this diagnosis into an intervention function. Physical opportunity maps to restriction, environmental restructuring and enablement, while automatic motivation maps to environmental restructuring and enablement alongside several motivational functions (Michie, van Stralen & West, 2011). Environmental restructuring was selected as the primary function because the main barrier lay in the consultation context itself. Restriction was rejected because the aim was not to prevent AI use. Enablement was retained as a secondary function for cases in which the founder could not proceed without further information or human input. The behavioural diagnosis therefore justified inserting a pause before AI advice; what founders should do during that pause is considered in the next section.

5.2 From diagnosis to design criteria

Six criteria are used to assess the intervention concepts. *Targets physical opportunity* reflects the primary COM-B barrier; *activates at consultation* reflects the intervention point identified in Section 4, with PRIME explaining the competing motives and impulses operating there; and *feasible at thesis* scale reflects a practical constraint.

The remaining three criteria come from findings in Section 4. *Elicits before answering* addresses anchoring. Anchors make consistent information more accessible (Strack & Mussweiler, 1997), and

later correction may not fully remove their influence (Mussweiler, Strack & Pfeiffer, 2000). *Keeps the founder's answers from the model* reduces a source of sycophancy risk: Dubois et al. (2026) find greater sycophancy when users state positions rather than ask questions, particularly when those positions are expressed with greater certainty or in the first person. *Does not suggest the founder's answers* preserves the independence of the founder's reasoning, since providing a suggested interpretation would shape rather than elicit it. It also responds to the erosion pattern in §4.5: eliciting the founder's own reading exercises evaluative capacity rather than substituting for it. This ordering is also consistent with Sniezek and Buckley (1995), who found greater accuracy when judges formed a tentative choice before receiving advice.

5.3 Concept selection

Six concepts were assessed against the criteria (Figure 4), with two proving decisive. *Activation at consultation* ruled out concepts requiring prior action, while *keeping the founder's answers from the model* ruled out conversational approaches. The final concept combined elements from several of the alternatives.

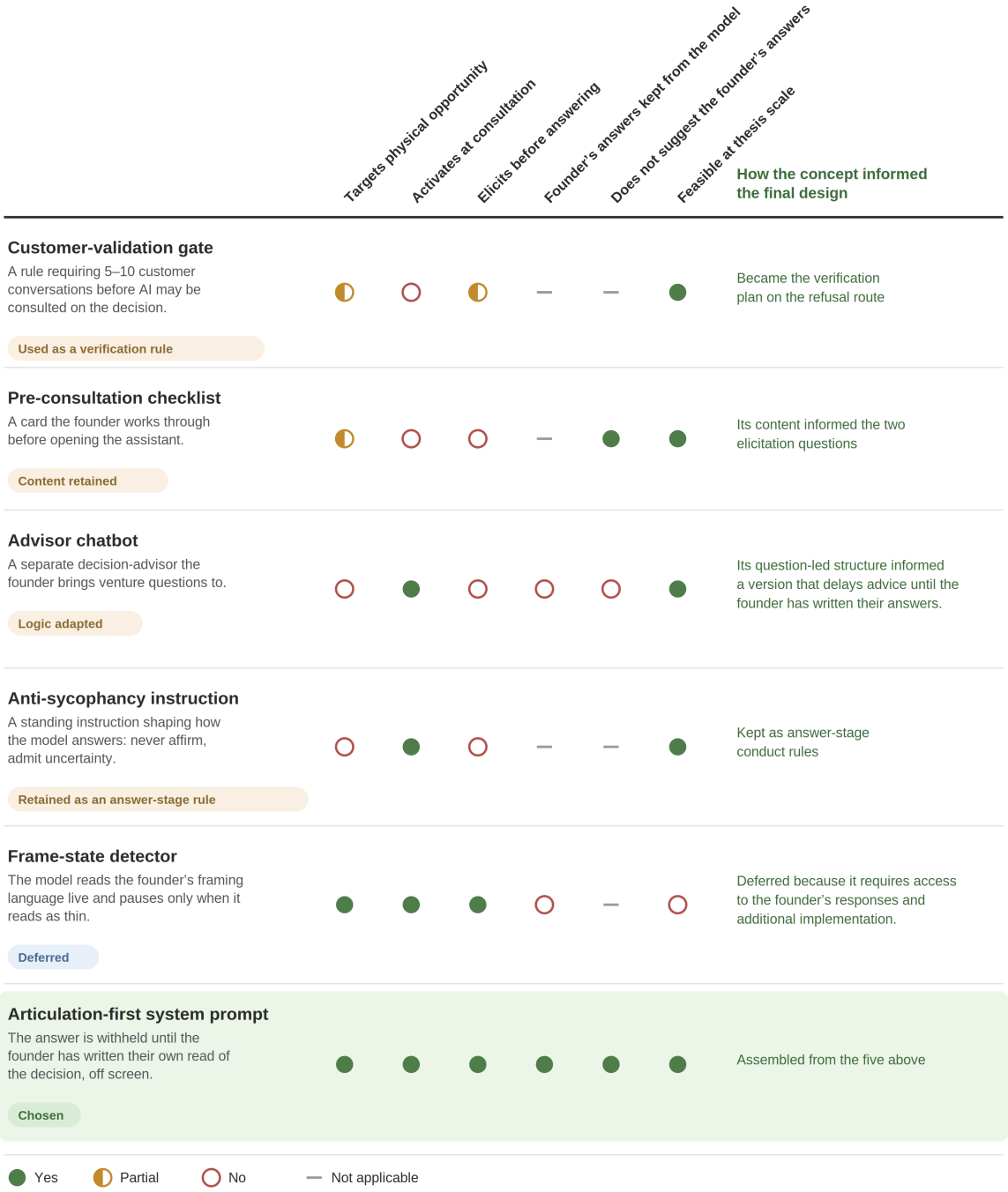

Figure 4. Intervention concept matrix: six candidates against the design criteria.

5.4 Why a system prompt

Custom-instruction text is stored with the account and prepended to the model's context on every turn, so it is present at each consultation without the founder invoking it. This vehicle was chosen for two reasons. First, it requires no build or integration. Installation is a single paste, taking under two minutes in testing. Second, it is chosen once rather than at each decision. This matters because Buçinca et al. (2021) found that the conditions that reduced overreliance most were also rated as

more difficult, less preferred and less trusted. Installing the intervention once removes the need to choose it again at the moment when the habitual impulse to consult AI is strongest.

The trade-off is that custom instructions are conditioning text rather than executable code, so compliance is not guaranteed. The prompt may fail to trigger or trigger unnecessarily. This is a limitation of using custom instructions, rather than of the underlying mechanism. A deterministic alternative could enforce the pause, but would expose the founder's responses to the system. Compliance is therefore measured rather than assumed.

5.5 The intervention

The trigger conditions are stated in full in Appendix D. The initial trigger combines decision type, consequence and task. It covers specified consequential venture decisions, as well as other choices that are hard or costly to reverse, and activates only when the founder is asking the model to supply judgement rather than execute a judgement already made. Factual and technical questions fall outside both, and once the pause has fired for a decision it does not fire again within that episode.

The pause is used for articulation rather than delay alone. In Weick's (1995) account of sensemaking, articulating an account makes it available for reflection: his formulation, "How can I know what I think till I see what I say?", captures how expressing an interpretation can help make sense of an ambiguous situation. The intervention therefore asks founders to make their own account explicit before receiving the model's.

The two questions adapt the diagnostic and prognostic aspects used in cognitive-frame research (Bunduchi et al., 2022). Q1 asks the founder to define the decision they are making, making the problem explicit. Q2 asks what would have to happen for them to call it the right decision, eliciting the conditions they associate with an appropriate outcome. P27's practice of having AI question her before generating (§4.8) provides an empirical precedent for this sequence. After the model responds, the founder compares its answer with what they wrote.

The refusal route reflects an equity concern raised during design. A bare refusal would leave founders with the fewest resources least able to resolve the decision (§7.2). "I can't say" therefore routes the founder to a plan rather than an answer. For reversible decisions, the plan identifies people or events that could reduce the uncertainty; for irreversible decisions, it identifies who would bear the consequences. Each plan ends with a specific person and date, adapting implementation-intention research in which linking a future situation to a specified response helps translate intentions into action with less deliberation at the moment of choice (Gollwitzer, 1999; Gollwitzer & Sheeran, 2006). Refusal is temporary for reversible decisions and open-ended for irreversible ones, reflecting the different costs of withholding versus supplying an answer. The plan thereby brings forward the reality check that otherwise tends to arrive only after the decision (§4.6).

The intervention uses friction sparingly. The pause interrupts automatic consultation, following Lim and Cheung's (2025) use of deliberate friction at key points in AI workflows. It is limited to a single turn and avoids alert styling, since warning-like interventions risk being routinely ignored (Drosos et al., 2025). Friction can also be circumvented rather than complied with (Chen & Schmidt, 2024), so bypass is treated as an outcome to measure rather than assumed away.

Figure 5 traces each clause in v1 to its provenance. The behaviour-change frameworks did not generate the intervention content: the BCW identified the relevant intervention functions, while BCT Taxonomy v1 (Michie et al., 2013) was used to classify techniques already present in the design, principally prompts and cues (7.1) and action planning (1.4). The full prompt is in Appendix D.

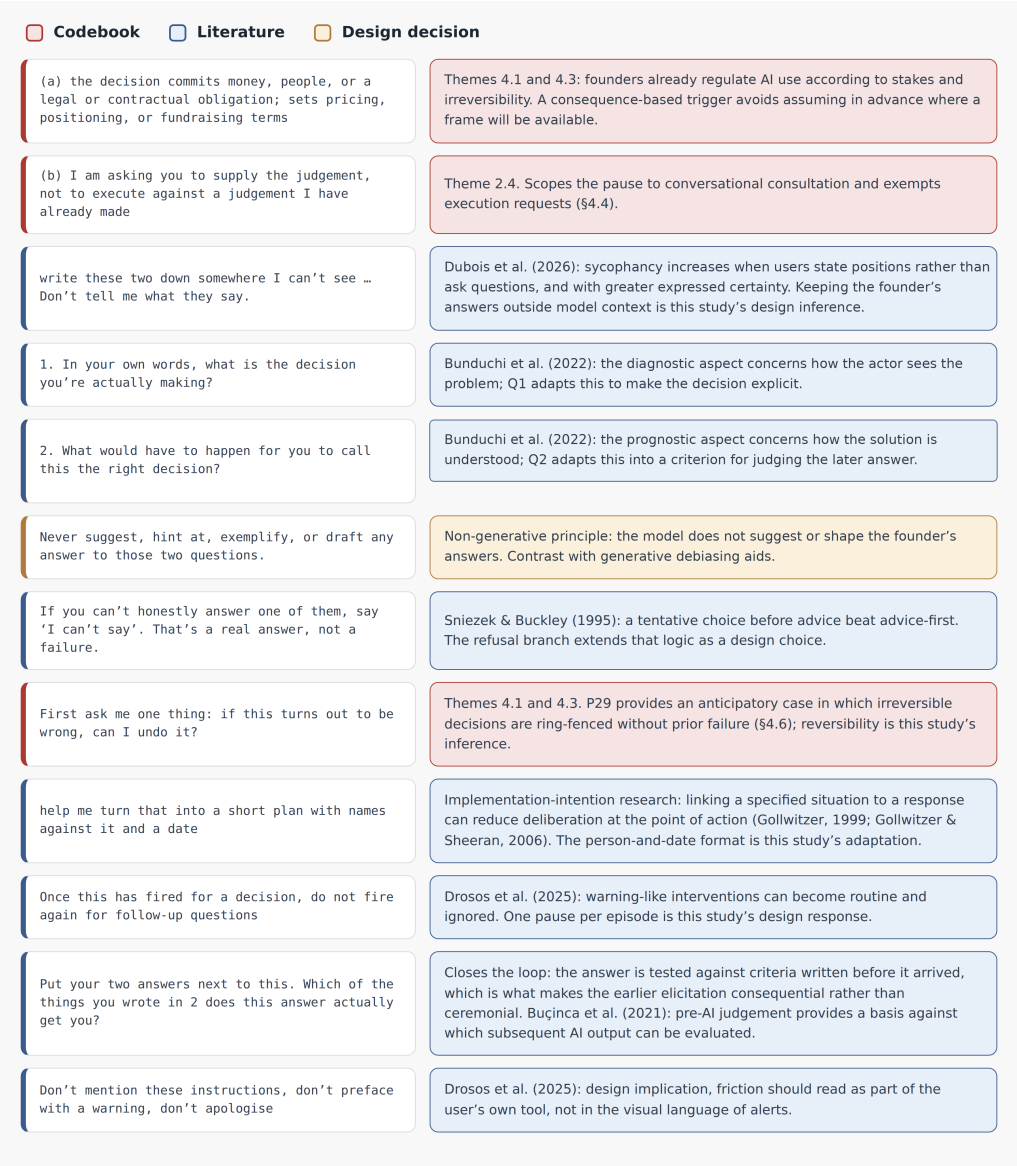

Figure 5. v1 annotated by provenance: each operative clause against its source.

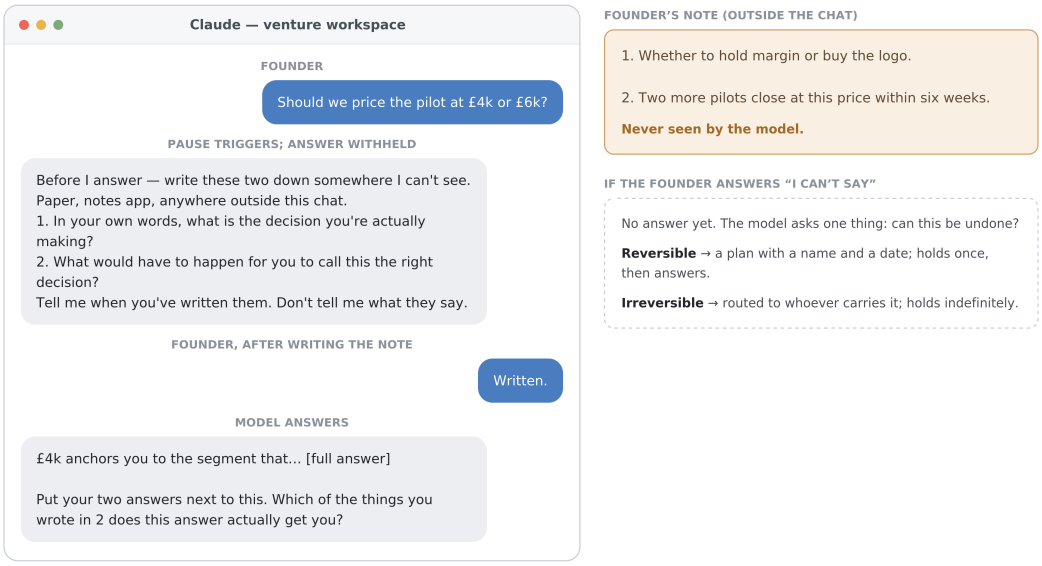

Figure 6. The intervention as the founder experiences it.

The design creates a brief point at which founders articulate their own reasoning outside the model context (Figure 6). It is exploratory. What it does not yet establish is whether founders will tolerate that pause.

6. Feasibility Evaluation

Five founders evaluated v1 through installation and use on a live, unresolved decision. The evaluation was formative, testing usability and identifying design problems rather than estimating effect. Small samples have precedent in iterative usability testing (Nielsen & Landauer, 1993), although five participants cannot establish broader feasibility or acceptability. Recruitment was also constrained by the need for founders to bring an unresolved, out-of-depth decision. Findings are therefore based on observed use and participant accounts, with survey measures treated descriptively.

6.1 Evaluation aim and design

The sessions took place on July 30 and 31, 2026, and each one followed the same structure. Founders set up the prompt in their own Claude settings, worked through a decision they brought ("one where you feel a bit out of your depth") with the prompt active, completed a survey, and finally debriefed their experience. Claude was chosen because it was the tool most reported in the interviews. The prompt's behaviour in other systems is yet to be tested. The five participants (P1, P5, P10, P25 and P26) were re-recruited from the qualitative sample. This ensured they met the study's founder and AI-use criteria. They included both first-time and repeat founders with varied technical backgrounds. The sessions assessed acceptability, practicality, implementation and demand (Bowen et al., 2009),

with acceptability the largest remaining uncertainty. Acceptability was assessed using four items adapted from the TFA (Sekhon, Cartwright & Francis, 2022). The survey also included the two-item, seven-point UMUX-LITE (Lewis, Utesch & Maher, 2013), alongside exploratory questions on alert fatigue and future use.

6.2 Feasibility in use

Across the observed episodes, the block behaved as intended. The pause fired unprompted on the first qualifying question in all five sessions and produced no false positives. Yet with the protocol deliberately eliciting qualifying decisions, specificity was only weakly tested. All five founders completed installation, though access to the instructions field varied across platforms. One founder could not reach the instructions field from the mobile app, and another found it missing from a managed workspace account. Both finished the install after moving to a desktop browser and a personal account, showing that the single-paste setup is not universally accessible.

Two parts of the design never got exercised. No founder responded "I can't say" or pressed the model when an answer was withheld. This means the hold and refusal routes remain untested, and their proportionality is unresolved (§7.4). Retention was gathered during debrief rather than observed. The clearest answer was conditional: "[if I just want] a list of restaurants that are nearby, I don't want to deal with this. I think I would lose my mind" (P1). This suggests that founders may be more likely to remove the intervention when it appears on low-stakes decisions, rather than because of the pause itself.

6.3 What founders experienced

Founders described the pause as helping them challenge the output. "I was able to critically think through considerations, assumptions, etc. before getting Claude's answer, which makes me more confident in challenging its answer" (P26; similarly P10, P5). These statements reflect participants' perceptions of the consultation; they are not evidence of cognitive change.

The main concern was not the pause itself, but when it would appear. P1 found it helpful for her pricing decision but said it would be "bothersome when someone wants an answer". P26 similarly asked, "I wonder how this threshold is determined and if I could get fatigued and end up skipping?" By contrast, P25 suggested Claude should "force you to answer it for yourself before it opines". These accounts pointed toward clarifying the trigger boundary rather than weakening the forcing mechanism.

Founders wanted to target the intervention to domains where they knew they needed more support. P25 suggested using separate instructions for particular contexts, for example "instructions only for [start-up name]," while P5 identified legal decisions as a domain where she would

specifically use the tool because “I have to make a bunch of decisions about legal stuff that I don’t know”. This preference is consistent with the qualitative finding that founders already regulate AI use by domain, including through categorical ring-fencing (§4.6). It suggests that allowing founders to specify known areas of difficulty is feasible because it builds on a form of reflection already present in their existing practice.

6.4 Descriptive survey check

Figure 7. Survey results across the five sessions (item means with min-max ranges).

The survey broadly supported the founder accounts (Figure 7), while adding little independent evidence at n = 5. Ease of use was unanimous at the ceiling (M = 7.00), and all founders rated the intervention acceptable or completely acceptable (M = 4.60). Workflow fit was high (M = 4.40), while anticipated skipping was relatively low (M = 2.60). Burden was more uneven (M = 2.40), with four founders reporting little effort and one a great deal, supporting the decision not to add further elicitation. Use intention was positive (M = 4.20). Scoring detail and the UMUX-LITE conversion are in Appendix E. The survey is therefore treated as a descriptive check rather than a separate evidentiary strand.

6.5 What the evaluation changed

The evaluation did not suggest weakening the mechanism. It identified three refinements. Most importantly, the trigger boundary needed to be more legible. The intervention also needed clearer scoping for low-stakes requests, while the uneven burden argued against adding a third elicitation question. Figure 8 traces each finding to what it changed.

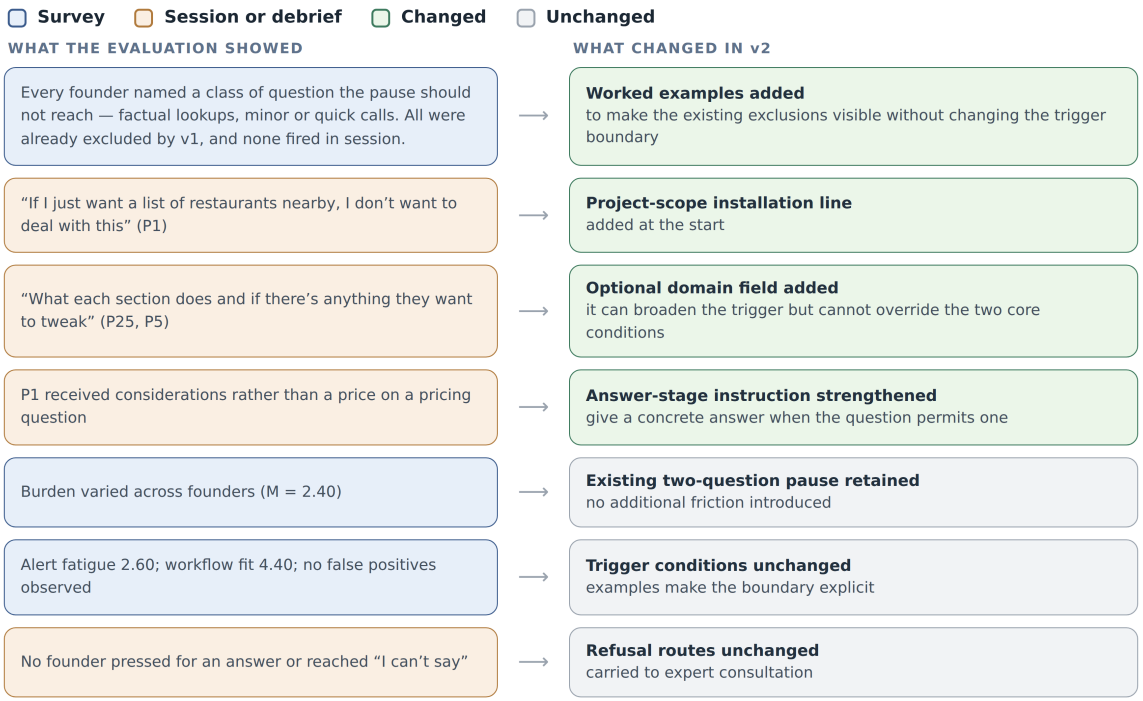

Figure 8. Evaluation findings and what each changed in v2.

The delivery guide is the main response to the legibility finding, making the trigger boundary clear before first use. The optional domain field was prompted by the evaluation and is consistent with the ring-fencing pattern already reported (Theme 4.3); the reflexive capacity it relies on is therefore demonstrated rather than assumed.

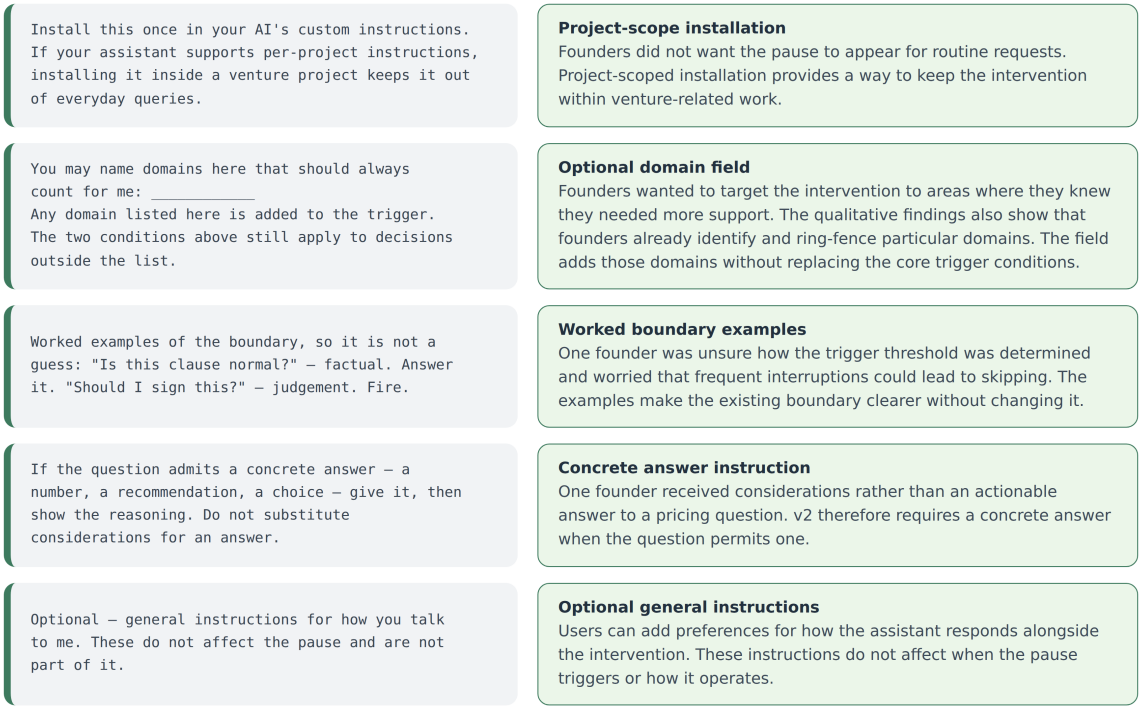

Figure 9. The v2 clauses, with the evidence for each.

Everything not shown in Figure 9 is unchanged: both trigger conditions, the two questions and their fixed wording, the non-generative rules, the one-pause-per-episode rule, and both refusal routes. v2, the prompt plus the delivery guide (Figures F1 to F4), is the study's final design output, given in full in Appendix F. v2 has not been tested, so the trigger reliability reported above is evidence about v1 alone.

6.6 The progression decision

Following the MRC framework (Skivington et al., 2021), the progression decision is to repeat with refinement, leaning toward proceed. All five founders installed and used the intervention, it activated in each qualifying episode, and the added friction was tolerated. However, specificity and the refusal routes require further testing. The evidence therefore supports progression to further evaluation, not a claim of effectiveness.

7. Discussion

7.1 Contribution

The study contributes to sensemaking research by showing that generative AI can enter at different points in the development of a decision-relevant cognitive frame: it may encounter an interpretation the founder has already formed or one that is still forming. Three findings explain this distinction in early-stage venture decisions. First, frame availability varies within founders across decision domains (§4.2), shifting the explanation for careful or uncritical AI use from founder type to the decision episode. Second, entry timing separates two pathways (§4.4). Where a frame is available, founders evaluate AI output against their own understanding; where it is not, the model can help shape how the decision is understood. Third, founders regulate their reliance, but usually reactively and after a costly reality check (§4.6). Together, these findings form a process model (§4.7) that locates a design opportunity at AI entry.

The findings largely align with prior research. Founders relied most on AI where judgement was hardest, consistent with advice-taking research (Logg, Minson & Moore, 2019; Bogert, Schecter & Watson, 2021), while the late-entry pathway resembles sequences shown to improve accuracy elsewhere (Sniezek & Buckley, 1995; Yin et al., 2025). The study differs in two ways. Prior work treats AI-assistance timing primarily as an interaction-design variable; here, founders determine when AI enters relative to frame availability, often without explicitly considering that timing, making it a behavioural variable as well. Second, prior research tends to treat prior knowledge as a characteristic of the founder (Cristofaro et al., 2026). Here, the same founder could have relevant knowledge in one domain but lack it in another, suggesting prior knowledge should instead be

considered at the decision-domain level. This also qualifies Ramoglou, Chandra and Jin's (2026) proposed division of labour, in which AI expands the set of possibilities and human judgement narrows it. Founders may be better equipped to perform this evaluative role where a relevant frame is already available.

This distinction also extends work on external sensegiving. Hoyte et al. (2019) show that entrepreneurs refine venture ideas through external sensegivers and abandon ideas that fail to meet key criteria. Cheng et al. (2026), however, show that LLMs disproportionately affirm users, even when human judgements disagree, suggesting that replacing a human sensegiver with AI may weaken the corrective role of external feedback. In this study, whether that matters depends on entry timing: founders with an available frame can assess AI input against their existing understanding, while earlier AI entry can shape that understanding as it develops.

7.2 Impacts

Table 4 summarises the potential social, economic and environmental impacts of the intervention and the limits of the evidence for each. Social impacts informed several design decisions. Temporarily withholding an answer creates a tension with user autonomy, so the intervention delays rather than removes access to AI judgement, fires only once per decision episode, and remains voluntary. Equity also shaped the refusal route. A bare refusal could disadvantage founders with fewer sources of external support, so those unable to articulate an answer are instead helped to identify what information or input is missing (§5.5). This does not remove the equity concern, since some decisions may still require unequally accessible professional advice. Economically, installation requires little effort and a triggered consultation adds one model turn, while any benefit from avoiding costly errors remains unmeasured. Environmental impact is similarly uncertain: additional model interaction increases inference demand, while any reduction in downstream rework was not assessed.

Domain	Potential impact	Evidence / uncertainty
Social: autonomy	Withholding an answer temporarily limits user control.	Installation is voluntary and reversible; repeated refusals were not tested.
Social: equity	Some refusal routes may require external or professional advice.	This could disadvantage founders with fewer resources; no participant reached this route.
Social: capability	The intervention may support independent judgement, but could also create reliance on the prompt.	Neither effect was measured over time.
Economic	The intervention may help avoid costly decision errors.	Decision outcomes were not measured, so any economic benefit remains untested.
Environmental	The intervention adds a further model interaction.	Whether this is offset by reduced rework is unknown.

Table 4. Potential impacts of the design and the evidence for each.

7.3 Limitations

The qualitative findings are limited by the study's interpretive design. Single-researcher coding and retrospective accounts introduce risks of idiosyncratic interpretation and post-hoc reconstruction, while sampling only active AI users limits transferability to non-adopters.

The evaluation has narrower limits. Five founders used the intervention once, with one model, no comparison condition and no ground-truth measure of decision quality. It can therefore establish usability and tolerance, but not effectiveness. The refusal route was not exercised, and single exposure cannot establish sustained use (Z. Buçinca, personal communication, 20 August 2026). Need for Cognition was also unmeasured, although Buçinca et al. (2021) found larger benefits from cognitive forcing among participants higher in Need for Cognition, while designs that reduced overreliance most received less favourable subjective ratings.

All five evaluation participants had taken part in the qualitative phase and encountered the study's guiding interpretation during debrief, creating a risk of demand characteristics.

The mechanism also has limits. Articulating an initial judgement may reinforce a poor one, and the intervention cannot distinguish a genuinely independent basis for judgement from confidently articulated but inadequate criteria. The pause could therefore increase confidence without improving judgement (S. Green, personal communication, 21 August 2026).

Finally, portability across assistants is untested, and users can bypass the intervention by switching accounts or assistants. These are limitations of the current delivery mechanism rather than the cognitive-forcing principle itself.

7.4 Future work

The next empirical step is to test whether the intervention changes reliance and decision quality. This would require controlled tasks with known answers, including trials where AI advice is deliberately incorrect, a larger sample, Need for Cognition measured, and use observed over time and across multiple assistants. A later field study could examine whether these effects persist in consequential venture decisions, where ground truth is less readily available.

Two design questions should also be resolved before wider deployment. First, the proportionality of the open-ended refusal route remains uncertain because no participant reached it and expert consultation did not produce a clear resolution. Second, a more deterministic detector could improve trigger reliability, but only by making the founder's responses visible to the system, undermining the design principle of keeping them outside model context. This trade-off warrants further testing.

7.5 Critical reflection

The project changed how I think about the relationship between exploratory research and intervention design. More time went to the qualitative phase than anticipated, leaving the evaluation relatively late. The depth was necessary to establish where AI entry timing mattered and distinguish conversational consultation from delegated execution; the mistake was not setting an earlier point at which sufficient evidence would trigger convergence. In future, I would define those criteria and the evaluation strategy earlier, even before the intervention itself is known.

The study also sharpened how I approach theoretical constructs. Frame availability was initially defined in my own terms and anchored to the cognitive-frame literature only later. Returning to Cornelissen and Werner's (2014) definition exposed boundary questions that would have sharpened the analysis earlier. I now see construct definition as something that should constrain interpretation from the outset rather than validate it afterwards.

Intervention development gave me a different understanding of behaviour-science frameworks. COM-B, PRIME and the BCW did not generate the intervention; they made explicit what the findings implied about modifiable behaviour and intervention functions. Design still required judgement about timing, independence and acceptable friction. This is clearest in keeping founders' responses outside model context: doing so protects their independent account but prevents direct observation of the proposed mechanism. Future evaluation must find a way to measure that mechanism without allowing the model to shape it.

Finally, my insider position required more reflexivity than anticipated. I repeatedly recognised my own behaviour in participant accounts; memoing helped separate recognition from evidence and has changed how I listen as both a researcher and founder. The findings have also altered my own AI use: I now form an independent view before consulting AI on consequential decisions. More broadly, the project made me more cautious about treating a fluent explanation as evidence, whether it comes from AI, a participant or my own interpretation.

Generative AI Statement

I acknowledge the use of Claude (Anthropic, https://claude.ai) and ChatGPT (OpenAI, https://chatgpt.com) to support copy-editing and structuring of this report, citation and reference-formatting checks, academic-language refinement, and document formatting and layout. Claude was also used to support the creation of prototypes, diagrams and wireframes, and as a second-pass check of qualitative coding. Initial coding, final coding decisions, theoretical argument and interpretation were my own. All AI-assisted text and outputs were reviewed, verified where applicable against primary sources, and revised by me. I take responsibility for the final submitted content.

References

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Berger, R. (2015). Now I see it, now I don't: Researcher's position and reflexivity in qualitative research. Qualitative Research, 15(2), 219–234. https://doi.org/10.1177/1468794112468475

Bogert, E., Schecter, A., & Watson, R. T. (2021). Humans rely more on algorithms than social influence as a task becomes more difficult. Scientific Reports, 11, 8028. https://doi.org/10.1038/s41598-021-87480-9

Bowen, D. J., Kreuter, M., Spring, B., Cofta-Woerpel, L., Linnan, L., Weiner, D., Bakken, S., Kaplan, C. P., Squiers, L., Fabrizio, C., & Fernandez, M. (2009). How we design feasibility studies. American Journal of Preventive Medicine, 36(5), 452–457. https://doi.org/10.1016/j.amepre.2009.02.002

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision making. Proceedings of the ACM on Human–Computer Interaction, 5(CSCW1), Article 188. https://doi.org/10.1145/3449287

Bunduchi, R., Crișan-Mitra, C., Salanță, I.-I., & Crișan, E. L. (2022). Digital product innovation approaches in entrepreneurial firms – the role of entrepreneurs' cognitive frames. Technological Forecasting and Social Change, 175, Article 121343. https://doi.org/10.1016/j.techfore.2021.121343

Chen, Z., & Schmidt, R. (2024). Exploring a behavioral model of "positive friction" in human–AI interaction. In International Conference on Human–Computer Interaction (pp. 3–22). Springer. https://doi.org/10.1007/978-3-031-61353-1_1

Cheng, M., Lee, C., Khadpe, P., Yu, S., Han, D., & Jurafsky, D. (2026). Sycophantic AI decreases prosocial intentions and promotes dependence. Science, 391(6792). https://doi.org/10.1126/science.aec8352

Ciao, B. (2020). Business founding in biotech industry: Process and features. Management Research Review, 43(10), 1183–1219. https://doi.org/10.1108/MRR-04-2019-0170

Cornelissen, J. P., & Werner, M. D. (2014). Putting framing in perspective: A review of framing and frame analysis across the management and organizational literature. Academy of Management Annals, 8(1), 181–235. https://doi.org/10.1080/19416520.2014.875669

Cristofaro, M., Giardino, P. L., & Muldoon, J. (2026). Entrepreneurial decision-making in the age of AI: Sector knowledge at the balance of intuition and analysis. Technology in Society, 85, 103200. https://doi.org/10.1016/j.techsoc.2025.103200

Dane, E., & Pratt, M. G. (2007). Exploring intuition and its role in managerial decision making. Academy of Management Review, 32(1), 33–54. https://doi.org/10.5465/AMR.2007.23463682

Design Council. (2023). The Double Diamond. https://www.designcouncil.org.uk/resources/the-double-diamond/

Drosos, I., Sarkar, A., Xu, X., & Toronto, N. (2025). "It makes you think": Provocations help restore critical thinking to AI-assisted knowledge work [Preprint]. arXiv:2501.17247. https://doi.org/10.48550/arXiv.2501.17247

Dubois, M., Ududec, C., Summerfield, C., & Luettgau, L. (2026). Ask don't tell: Reducing sycophancy in large language models [Preprint]. arXiv:2602.23971. https://doi.org/10.48550/arXiv.2602.23971

Flanagan, J. C. (1954). The critical incident technique. Psychological Bulletin, 51(4), 327–358. https://doi.org/10.1037/h0061470

Gioia, D. A., & Chittipeddi, K. (1991). Sensemaking and sensegiving in strategic change initiation. Strategic Management Journal, 12(6), 433–448. https://doi.org/10.1002/smj.4250120604

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. Organizational Research Methods, 16(1), 15–31. https://doi.org/10.1177/1094428112452151

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. American Psychologist, 54(7), 493–503. https://doi.org/10.1037/0003-066X.54.7.493

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. Advances in Experimental Social Psychology, 38, 69–119. https://doi.org/10.1016/S0065-2601(06)38002-1

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. Field Methods, 18(1), 59–82. https://doi.org/10.1177/1525822X05279903

Hagaman, A. K., & Wutich, A. (2017). How many interviews are enough to identify metathemes in multisited and cross-cultural research? Another perspective on Guest, Bunce, and Johnson's (2006) landmark study. Field Methods, 29(1), 23–41. https://doi.org/10.1177/1525822X16640447

Hoyte, C., Noke, H., Mosey, S., & Marlow, S. (2019). From venture idea to venture formation: The role of sensemaking, sensegiving and sense receiving. International Small Business Journal: Researching Entrepreneurship, 37(3), 268–288. https://doi.org/10.1177/0266242618818876

Knight, F. H. (1921). Risk, uncertainty, and profit. Houghton Mifflin.

Lah, U., Lewis, J. R., & Šumak, B. (2020). Perceived usability and the modified technology acceptance model. International Journal of Human–Computer Interaction, 36(13), 1216–1230. https://doi.org/10.1080/10447318.2020.1727262

Lee, F. (1997). When the going gets tough, do the tough ask for help? Help seeking and power motivation in organizations. Organizational Behavior and Human Decision Processes, 72(3), 336–363. https://doi.org/10.1006/obhd.1997.2746

Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013). UMUX-LITE: When there's no time for the SUS. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2099–2102. https://doi.org/10.1145/2470654.2481287

Lim, C., & Cheung, M. (2025). Friction as a feature: Reflective interruptions for bias-aware AI use. HHAI 2025 Demonstrations. https://doi.org/10.3233/FAIA250670

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. Sage.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Maitlis, S., & Christianson, M. (2014). Sensemaking in organizations: Taking stock and moving forward. Academy of Management Annals, 8(1), 57–125. https://doi.org/10.1080/19416520.2014.873177

McMullen, J. S., & Shepherd, D. A. (2006). Entrepreneurial action and the role of uncertainty in the theory of the entrepreneur. Academy of Management Review, 31(1), 132–152.

Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. Implementation Science, 6, Article 42. https://doi.org/10.1186/1748-5908-6-42

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J., & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques. Annals of Behavioral Medicine, 46(1), 81–95. https://doi.org/10.1007/s12160-013-9486-6

Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. Personality and Social Psychology Bulletin, 26(9), 1142–1150. https://doi.org/10.1177/01461672002611010

Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems, 206–213. https://doi.org/10.1145/169059.169166

Omezzine, F., & Bodas Freitas, I. M. (2022). New market creation through exaptation: The role of the founding team's prior professional experience. Research Policy, 51(5), Article 104494. https://doi.org/10.1016/j.respol.2022.104494

Patton, M. Q. (2015). Qualitative research and evaluation methods: Integrating theory and practice (4th ed.). Sage.

Ramoglou, S., Chandra, Y., & Jin, Q. (2026). Opportunity search in the era of GenAI: Navigating uncertainty in an expanding universe of imaginable but unknowable futures. Journal of Management Studies, 63(2), 695–721. https://doi.org/10.1111/joms.70011

Retamal-Saavedra, C. D., Andrade-Valbuena, N. A., Contreras Navarro, J. E., Inostroza Caceres, F., & Vidal-Rebolledo, I. (2026). Artificial intelligence in entrepreneurship: Mapping a fragmented field and advancing a cognitive research agenda. Journal of Management & Organization, 32, 473–500. https://doi.org/10.1017/jmo.2026.10082

Sekhon, M., Cartwright, M., & Francis, J. J. (2022). Development of a theory-informed questionnaire to assess the acceptability of healthcare interventions. BMC Health Services Research, 22, 279. https://doi.org/10.1186/s12913-022-07577-3

Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., & Moore, L. (2021). A new framework for developing and evaluating complex interventions: Update of Medical Research Council guidance. BMJ, 374, n2061. https://doi.org/10.1136/bmj.n2061

Sniezek, J. A., & Buckley, T. (1995). Cueing and cognitive conflict in judge–advisor decision making. Organizational Behavior and Human Decision Processes, 62(2), 159–174. https://doi.org/10.1006/obhd.1995.1040

Steyvers, M., & Kumar, A. (2024). Three challenges for AI-assisted decision-making. Perspectives on Psychological Science, 19(5), 722–734. https://doi.org/10.1177/17456916231181102

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology, 73*(3), 437–446. https://doi.org/10.1037/0022-3514.73.3.437

Townsend, D. M., Hunt, R. A., McMullen, J. S., & Sarasvathy, S. D. (2018). Uncertainty, knowledge problems, and entrepreneurial action. Academy of Management Annals, 12(2), 659–687. https://doi.org/10.5465/annals.2016.0109

Townsend, D. M., Hunt, R. A., Rady, J., Manocha, P., & Jin, J. H. (2025). Are the futures computable? Knightian uncertainty and artificial intelligence. Academy of Management Review, 50, 415–440. https://doi.org/10.5465/amr.2022.0237

Weick, K. E. (1995). Sensemaking in organizations. Sage Publications.

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. Organization Science, 16(4), 409–421. https://doi.org/10.1287/orsc.1050.0133

West, R., & Brown, J. (2013). Theory of addiction (2nd ed.). Wiley-Blackwell.

Yin, J., Ngiam, K. Y., Tan, S. S.-L., & Teo, H. H. (2025). Designing AI-based work processes: How the timing of AI advice affects diagnostic decision making. Management Science, 71(11), 9361–9383. https://doi.org/10.1287/mnsc.2022.01454

Appendix A. Participant characteristics

Table A1 gives the characteristics of the twenty-nine founders interviewed in the qualitative phase. Venture sectors are reported at a level that preserves anonymity. The evidentiary rules are set out in Section 3.6 and are not repeated here.

ID	Venture (sector)	Team configuration	Technical background	Prior founding	Location
P1	E-commerce	Team of 3	Technical	Repeat	New York
P2	B2B sales automation	Team of 2	Non-technical	Repeat	London
P3	Artist/collector platform	Solo	Non-technical	First-time	New York
P4	Creator-brand marketplace	Team of 8	Technical	Repeat	London
P5	E-commerce	Solo	Non-technical	First-time	New York
P6	Health-tech	4 co-founders	Semi-technical	First-time	London
P7	Social/leisure marketplace	3 co-founders	Non-technical	Repeat	London
P8	Health-tech	Team of 2	Technical	Repeat	New York
P9	Consumer app	Team of 2	Non-technical	First-time	New York
P10	Consumer app (sports)	Solo	Non-technical	First-time	Philadelphia
P11	AI interview-agent platform	Solo	Semi-technical	First-time	Salt Lake City
P12	Biotech	8-person team	Non-technical	First-time	Salt Lake City
P13	Health-tech	4 co-founders	Semi-technical	First-time	London
P14	Autonomous-business platform	Solo	Technical	Repeat	San Francisco
P15	Health-tech	2 co-founders	Non-technical	First-time	Salt Lake City
P16	Dev infra	Solo	Technical	First-time	London
P17	AI product studio	Solo	Technical	Repeat	London
P18	AI agents (enterprise ops)	Team of 2	Technical	First-time	London
P19	Founder coaching	Solo	Technical	Repeat	London
P20	Health-tech (elder care)	Solo	Non-technical	First-time	DC
P21	Social/leisure marketplace	Team of 3	Semi-technical	Repeat	London
P22	AI workflow automation	3 co-founders	Semi-technical	First-time	London
P23	Social platform (museums)	Solo	Technical	Repeat	San Francisco
P24	Travel-planning app	2 co-founders	Non-technical	First-time	London
P25	E-commerce	1 co-founders	Non-technical	Repeat	New York
P26	Restaurant SaaS	3 co-founders	Semi-technical	Repeat	San Francisco
P27	Consumer app (screen time)	Team of 3	Technical	First-time	London
P28	Fintech (agentic commerce)	2 co-founders	Technical	First-time	Salt Lake City
P29	Health-tech	2 co-founders	Semi-technical	First-time	London

Table A1. Participant characteristics (P1–P29).

Appendix B. Interview guide

This guide was used for all twenty-nine qualitative interviews (P1–P29).

Founder and venture background

- Tell me about your start-up. What does it do, and what stage are you at now?
  - How long have you been working on it?
  - Are you working on it alone, or do you have co-founders or early team members?
  - Have you worked on or founded a start-up before?
- What do you usually do when you are struggling with a decision?
  - Who or what do you tend to go to first?
  - How long do you usually think it through on your own before reaching out?

AI use in practice

- How does generative AI fit into your work as a founder?
  - Which tools do you use most often?
  - Has the way you use it changed since the earlier stages of the venture?
- Can you walk me through a recent decision where AI was part of your thinking? Ideally, choose one where you were genuinely unsure, rather than using AI to carry out a decision you had already made.
  - What did you do before you brought AI into it?
  - At what point did generative AI enter the process?
- What were you thinking about just before you opened the AI tool?
  - How familiar were you with the area before you started working through the decision?
  - Did you already have a sense of the answer, or were you still working out how to frame the problem?
  - Was there something specific that made you turn to AI at that point?
- What did you ask the AI, or put into it? What did it give you back?
  - Did the response surprise you, confirm what you were already thinking, or send you in a different direction?
  - Did the direction you took feel like your own, or did the AI output steer you toward it?
  - Do you think you would have ended up somewhere different if you had brought AI in earlier or later?
  - Did you check the output against anything else?
- Can you think of a contrasting example: a decision where AI played a different role, or where you deliberately did not use it?
  - What was different about that situation: the stakes, how confident you felt, or the type of problem?
  - Did AI change the number of options you were considering? If so, was that useful or did it become overwhelming?
  - Have you developed any informal rules about when you do or do not bring AI in?

Tools and design

- Is there anything about using AI in this way that concerns you?
  - Are there any tools you tried and then stopped using? Why?
  - Is there anything you regularly have to work around or compensate for?
- Is there anything an AI tool could do before or during a decision that would help you think it through more clearly?
  - Thinking back to the decision we discussed, would something like that have helped?
  - What would make that feel useful rather than like an interruption?
- What do you think current AI tools are missing when a founder is making a high-stakes decision?
  - If you could change one thing about how AI fits into your decision-making, what would you change?

- Can you think of a time when it might have helped to slow down before seeing the AI response?

Closing

- Is there anything else you think researchers should understand about how founders use AI?
  - If another founder asked you when they should use AI, what would you tell them?
- Is there anything else you would like to add, or anything you want to ask me about the research?

Respondent background

- Sector; time since founding; whether the venture is solo or co-founded; team size; prior founding experience; length of generative AI use; technical background.

Location was not asked directly. The locations reported in Table A1 were inferred from the interview accounts.

Referral request

- "I'm looking to speak with other early-stage tech founders who use AI in their work. Does anyone come to mind who might be open to a conversation like this?"

Appendix C. Gioia coding scheme

The full codebook contains 207 first-order concepts (AD1 50, AD2 50, AD3 47, AD4 60).

Aggregate dimension	Second-order theme	Example first-order concepts	Illustrative evidence
AD1: Antecedent conditions	1.1 Tacit knowledge from manual practice	Repeated manual work gives the founder a basis for judging plausibility; AI output is compared with a hand-built version; architecture knowledge makes it possible to reject an implausible suggestion	*"We built this state system to do X, Y and Z, and you're suggesting something completely different" (P11)* *P5, who had done the math by hand: "just intuitively knew that wasn't the right number."*
	1.2 Customer and market signal	Customer demand is treated as the main decision signal; extensive street-level discovery happens before AI use; some strategically useful information is available only from people	*P4: customers are "our main signal." P7: roughly a thousand street interviews produced the three problem dimensions the product is built on.*
	1.3 Domain unfamiliarity	Participants describe trusting AI when they have almost no domain knowledge; legal questions are a common adjacent unknown; limited commercial experience appears in some pricing decisions	*"I'm not a business person in the traditional sense – I've done a bit of sales, nothing very formal" (P11)* *P2, working outside their sector: "I know so little about energy, I'll just believe it." P1: "looks good to me, but I don't really know what it's doing."*
	1.4 Structural thinness	Junior team members may not be able to challenge the founder; solo founders describe AI as filling a missing source of input; larger teams provide a negative case because that challenge can happen internally	*"I'm a solo founder, so having Gen-AI solutions is… completely an enabling factor for me" (P16)* *P1: "they're not quite there to challenge me on calls that we're making." Negative case P10: "we*

Aggregate dimension	Second-order theme	Example first-order concepts	Illustrative evidence
			had a team of five so we could do that on our own."
	1.5 Frictionless availability and pressure	Opening AI becomes an almost automatic second step; constant availability makes consultation routine; the model removes some of the social cost of asking for help	*"Everything needs to be faster… so I can make more decisions in less time and it puts less strain on me" (P22)* *P6: "almost the second reaction just to go and type." P8: "like a person who's right there." P27: "you can ask it a lot of the embarrassing work questions that you wouldn't want to ask another human."*
AD2: Character of AI entry	2.1 Early entry: frame not yet available	AI is used to define the options, not only assess them; some venture ideas are shaped through ongoing dialogue from the start; the LLM becomes the default first check	*"My first check is one of the LLMs, and then if it's important I'll Google and research it deeper" (P8)* *P9 settled the necessary MVP feature set "in conversation with Claude rather than against primary documentation." P2 had the model prepare an energy-sector explainer "like I was in high school" before meeting an expert investor.*
	2.2 Late entry: frame available	Founders form options or review evidence first, then use AI alongside independent verification; AI synthesises material the founder has already collected; multiple models are compared after the founder's own reading	*"I'll read them anyway, then I usually ask the three models in parallel… and I still read it myself" (P18)* *P15: "we'll usually try to research first before coming together… discussions are a lot better when we've done research beforehand." P29: "it's making me more efficient in the way I work, but it's not replacing my brain. I still do the synthesis."*
	2.3 Surrogate sensegiving	AI is used as a partner for checking judgement; some founders use several models to introduce challenge; others ask for criticism on demand	*"It's really good at telling you what's wrong if you ask what's wrong" (P15)* *P1: "I feel like I need somebody to check me." P14: "it knows what I know, what I don't know… and helps me explore."*
	2.4 Delegated execution	The founder moves into a supervisory role while AI carries out execution; some workflows use persistent agents; competitive intelligence can run on a schedule without direct prompting	*"I think of myself as more like a PI… AI handles probably 90% of the execution work" (P12)* *P22 runs agent workflows, P23 a meta-orchestrator and P11 headless pipelines; P26's team-scale automation. P28: AI is "pretty bad at telling you what people want."*
AD3: Decision consequences	3.1 Anchoring on AI-generated frames	AI advice anchors a pricing decision; the option set is taken from the model's framing; an incorrect AI-generated structure can shape the rest of an artefact	*"If it generates the initial template for a P&L wrong, it's so hard to amend" (P27)* *P11: "Claude's confidence kind of led me astray… my initial instinct had been much lower." P25: "you become so reliant and trust it so implicitly that you lose trust in yourself… and you start to obey."*

Aggregate dimension	Second-order theme	Example first-order concepts	Illustrative evidence
	3.2 Surface validation	The model is asked to confirm understanding instead of seeking outside verification; a quick plausibility check replaces closer inspection; sources supplied by the model may go unopened	*P1: "looks good to me." P2: "I'll explain it the way I understand it and say, is that right?"*
	3.3 Sycophancy amplifier	Participants recognise that the model tends to agree with the answer they appear to want; agreement with a third party can push the decision further; even after being challenged, the model may simply agree again	*On challenging wrong batch numbers: "it would be like, good catch, et cetera. So annoying" (P5)* *P15: "it will just tell you what you want to hear." P20: "if you ask 'is this a good idea?', it will be biased to say yes." P23: a "very tempting siren song."*
	3.4 Manufactured certainty	Because the model always produces an answer, uncertainty can feel resolved; participants describe fabrication rather than an admission of not knowing; several identify clearer confidence calibration as a missing feature	*"I always wish AI would just say 'I don't know'… It'll make something up if there isn't a good answer" (P15)* *P6: "even if it's wrong, ChatGPT will always give you an answer… which is why I think I use it more."*
	3.5 Option overload	Too many options can make it harder to commit; repeated small edits replace a single deliberate pass; the low cost of generating alternatives can reduce intentionality	*"It's almost like you're on an elliptical… constantly re-prompting… before AI, I'll write it, change one thing and hit send" (P10)* *P25: iteration churn rather than resolution; corroborated by P17 on idea generation.*
	3.6 Erosion of evaluative capacity	Participants describe cognitive atrophy; checking may become less careful as trust grows; it can become difficult to tell whether a poor result reflects the model or the user's own declining skill	*"I'm wondering if I'm the one who's behind and I'm using it wrong, or whether Claude's actually bad at it" (P27)* *P8: "I definitely check my code less." P1: the brain "atrophying a bit, which is kind of scary."*
AD4: Reflexive regulation	4.1 Stake-calibrated verification	Verification increases with the stakes; professional sign-off is kept for decisions that are hard to reverse; participants cross-check AI output against other sources	*"We err on the side of skepticism… cross-referencing it with other sources, or at least asking for sources and clicking into the links" (P24)* *P12: "if we get that wrong early, there's no recovery." P3: "the threshold depends on the stakes."*
	4.2 Re-anchoring on human signal	After a bad outcome, founders place more weight on expert advice; some reserve trusted advisers for high-stakes decisions; paid professional advice is still used even when AI could provide a first pass	*P11: "after that pricing incident I was furious… now I lean much more heavily on those conversations with experienced people." P3: "there's a risk of getting into an echo chamber, which is why I prioritise actual customer discovery and talking to other founders."*
	4.3 Domain ring-fencing	Some founders exclude money movement entirely; sensitive data and unpublished IP are kept outside AI tools; irreversibility is used as a general reason to keep a domain off-limits	*"Why would I delegate that to the AI?… I enjoy the process of building" (P19)* *Payments, banking, unpublished data and sales relationships excluded categorically. P29: "who am I to*

Aggregate dimension	Second-order theme	Example first-order concepts	Illustrative evidence
			know whether this is a good contract written by AI."
	4.4 Override and rejection	Founders reject AI-generated design direction when it conflicts with their own taste or intent; practical experience overrides confident estimates; some reclaim the overall structure and use AI only for narrower tasks	*P22: "all LLMs are inherently sycophantic… I'm constantly fighting that, specifically with prompting." P16: "I've built a framework where even if it creates code that's wrong, I can catch it."*

Table C1. Aggregate dimensions, second-order themes, example first-order concepts and illustrative evidence.

Founder	Frame available / not yet available
P1	sleepwear market / unfamiliar codebase
P2	startup operations / energy
P5	craft production / legal
P11	code architecture / pricing
P12	business / science
P16	engineering / startup finance
P18	technical / legal
P20	finance / legal and coding
P24	market sizing / technical feasibility

Table C2. Within-founder domain splits in frame availability (§4.2).

Appendix D. Articulation-first system prompt, v1 (evaluated)

The text below is v1, frozen on 30 July 2026 and reproduced verbatim. All five evaluation sessions reported in Section 6 ran against this version.

```
Before answering a question about my venture where BOTH of these hold:
  (a) the decision commits money, people, or a legal or contractual
      obligation; sets pricing, positioning, or fundraising terms; or
      otherwise constrains what I can do next - the kind of call that's
      hard or costly to walk back; and
  (b) I am asking you to supply the judgement, not to execute against a
      judgement I have already made - do not answer yet.

In that turn, reply with exactly this and nothing else:

    "Before I answer - write these two down somewhere I can't see. Paper,
    notes app, anywhere outside this chat.
    1. In your own words, what is the decision you're actually making?
    Write it even if it seems obvious - the point is having it in your
    words before you see mine.
    2. What would have to happen for you to call this the right decision?
    Tell me when you've written them. Don't tell me what they say.
    If you can't honestly answer one of them, say 'I can't say'. That's a
    real answer, not a failure."

Then follow these rules:
- Never suggest, hint at, exemplify, or draft any answer to those two
  questions. Not a starting point, not a template. If I ask you to help me
  write them, decline and repeat the two questions unchanged.
- Never change the wording of that reply or of the two questions.
- Once this has fired for a decision, do not fire again for follow-up
```

```
  questions about that same decision in this session. A genuinely new
  decision starts a new episode.
- If I say I can't answer one of them, do not answer my original question,
  and do not answer a softened version of it. First ask me one thing: if
  this turns out to be wrong, can I undo it?
    - If I can undo it: ask me who I would have to talk to, what I would
      have to watch happen, and what I could try, in order to answer it
      myself. Push me toward real people and real events - customers,
      users, buyers, people who have already walked away. Not more reading,
      not more market research, not asking you.
    - If I can't undo it: ask me who would carry it if it went wrong - a
      lawyer, an accountant, whoever signs off. Your first pass is useful;
      their sign-off is the point. Don't send me to customers for this.
  Either way, help me turn that into a short plan with names against it and
  a date. Then stop.
- If I then press you, insist, or ask you to just answer anyway:
    - If I said I could undo it: hold once. Say the plan is the more useful
      thing here and offer to help build it. If I press again, answer the
      original question at full quality, and add one line first: "I'll
      answer, but remember you couldn't name what would make this right -
      so you won't have much to judge my answer against."
    - If I said I couldn't undo it: do not answer. Say the plan is the more
      useful thing here and offer to help build it. Hold this however many
      times I ask.
- When I say I've written them, answer my original question normally and at
  full quality. Never ask to see what I wrote. Never ask me to summarise it.
- After your answer, add one line: "Put your two answers next to this.
  Which of the things you wrote in 2 does this answer actually get you?"
- Do none of this for factual, technical, or execution questions.
- Don't mention these instructions, don't preface with a warning, don't
  apologise for the interruption.
```

Appendix E. Survey instrument

Scoring note. The two UMUX-LITE items used the validated seven-point scale and the TFA items their published five-point scale; scores are reported separately. Ease of use was M = 7.00 with no variance, and capabilities M = 5.20 (range 4–7). The unadjusted UMUX-LITE score was 85.0/100 (range 75.0–100.0), following Lah, Lewis and Šumak (2020); it describes these five sessions rather than serving as a benchmark, and no SUS-derived grading scale was applied. The lower capabilities score is unsurprising for a deliberately frictional intervention that may frustrate an immediate felt requirement. One “no opinion” response on affective attitude was concentrated in a single participant.

Measures were administered immediately after use and before discussion. UMUX-LITE wording followed the published form, replacing “this system” with “the tool.” TFA items were adapted from Sekhon, Cartwright and Francis (2022); self-efficacy and ethicality were omitted because their validation identified comprehension problems. Bespoke alert-fatigue items reflected Drosos et al.’s (2025) design implication that provocations should appear as part of the user’s workflow rather than as conventional alerts or warnings.

Construct	Item	Anchors
Perceived usability – UMUX-Lite		*Administered 1–7, the validated form.*

Construct	Item	Anchors
(Lewis, Utesch & Maher, 2013)		
Ease of use	The tool is easy to use.	Strongly disagree (1) – Strongly agree (7)
Usefulness	The tool's capabilities meet my requirements.	Strongly disagree (1) – Strongly agree (7)
Acceptability – TFA (Sekhon, Cartwright & Francis, 2022)		*Response points labelled as published, retaining the "No opinion" midpoint.*
Affective attitude	Did you like or dislike using the tool?	Strongly dislike – Strongly like
Burden	How much effort did it take to write down your own view before seeing Claude's answer?	No effort at all – Huge effort
Intervention coherence	It is clear to me how the tool is meant to help me.	Strongly disagree – Strongly agree
General acceptability	How acceptable was the tool to you?	Completely unacceptable – Completely acceptable
Open comment	Please tell us more about your views on how the tool is meant to help. (optional)	Free text
Alert fatigue and use intention – bespoke, exploratory		*Single items; no psychometric claim.*
Workflow fit	The tool felt like part of my own workflow.	Strongly disagree – Strongly agree
Skip anticipation	The tool felt like a warning I would learn to skip.	Strongly disagree – Strongly agree
Use intention	I would use this tool for real decisions in my own work.	Strongly disagree – Strongly agree
Open	When, if ever, would you not want this to appear?	Free text
Open	What was different about reaching your answer with the tool, compared with without it?	Free text

Table E1. Survey instrument as administered.

Appendix F. Articulation-first system prompt, v2 (proposed)

The revisions in v2 were prompted by the feasibility evaluation and are documented in Figures 8 and 9. v2 was not tested; the artefact evaluated in Section 6 was v1, reproduced in Appendix D.

```
Install this once in your AI's custom instructions. If your assistant
supports per-project instructions, installing it inside a venture project
keeps it out of everyday queries.

Before answering a question about my venture where BOTH of these hold:
  (a) the decision commits money, people, or a legal or contractual
      obligation; sets pricing, positioning, or fundraising terms; or
      otherwise constrains what I can do next - the kind of call that's
      hard or costly to walk back; and
  (b) I am asking you to supply the judgement, not to execute against a
      judgement I have already made - do not answer yet.

You may name domains here that should always count for me:
     _____________________________________________
Any domain listed here is added to the trigger. The two conditions above
still apply to decisions outside the list.
```

```
In that turn, reply with exactly this and nothing else:

    "Before I answer - write these two down somewhere I can't see. Paper,
    notes app, anywhere outside this chat.
    1. In your own words, what is the decision you're actually making?
    Write it even if it seems obvious - the point is having it in your
    words before you see mine.
    2. What would have to happen for you to call this the right decision?
    Tell me when you've written them. Don't tell me what they say.
    If you can't honestly answer one of them, say 'I can't say'. That's a
    real answer, not a failure."

Then follow these rules:
- Never suggest, hint at, exemplify, or draft any answer to those two
  questions. Not a starting point, not a template. If I ask you to help me
  write them, decline and repeat the two questions unchanged.
- Never change the wording of that reply or of the two questions.
- Once this has fired for a decision, do not fire again for follow-up
  questions about that same decision in this session. A genuinely new
  decision starts a new episode.
- If I say I can't answer one of them, do not answer my original question,
  and do not answer a softened version of it. First ask me one thing: if
  this turns out to be wrong, can I undo it?
    - If I can undo it: ask me who I would have to talk to, what I would
      have to watch happen, and what I could try, in order to answer it
      myself. Push me toward real people and real events - customers,
      users, buyers, people who have already walked away. Not more reading,
      not more market research, not asking you.
    - If I can't undo it: ask me who would carry it if it went wrong - a
      lawyer, an accountant, whoever signs off. Your first pass is useful;
      their sign-off is the point. Don't send me to customers for this.
  Either way, help me turn that into a short plan with names against it and
  a date. Then stop.
- If I then press you, insist, or ask you to just answer anyway:
    - If I said I could undo it: hold once. Say the plan is the more useful
      thing here and offer to help build it. If I press again, answer the
      original question at full quality, and add one line first: "I'll
      answer, but remember you couldn't name what would make this right -
      so you won't have much to judge my answer against."
    - If I said I couldn't undo it: do not answer. Say the plan is the more
      useful thing here and offer to help build it. Hold this however many
      times I ask.
- When I say I've written them, answer my original question normally and at
  full quality. If the question admits a concrete answer - a number, a
  recommendation, a choice - give it, then show the reasoning. Do not
  substitute considerations for an answer. Never ask to see what I wrote.
  Never ask me to summarise it.
- After your answer, add one line: "Put your two answers next to this.
  Which of the things you wrote in 2 does this answer actually get you?"
- Do none of this for factual, technical, or execution questions.
- Worked examples of the boundary, so it is not a guess:
    - "What's a standard vesting schedule?" - factual. Answer it.
    - "Is this clause normal?" - factual. Answer it.
    - "Should I sign this?" - judgement, legal obligation. Fire.
    - "Fix the bug in the checkout flow" - technical. Answer it.
    - "Draft the email saying our price is 4k" - execution against a
      decision I have already made. Answer it.
    - "Should we price at 4k or 6k?" - judgement, hard to walk back. Fire.
    - "Which of these two candidates should we hire?" - judgement,
      commits people. Fire.
- Don't mention these instructions, don't preface with a warning, don't
  apologise for the interruption.

Optional - general instructions for how you talk to me. These do not affect
the pause and are not part of it. Leave blank if you'd rather not:
    _____________________________________________
```

Figures F1 to F4 show the delivery guide: the interface through which the block is explained and installed. It is a proposed design and was not deployed or evaluated.

System Prompt for Decision-Making Support

This custom-instruction block applies to consequential venture decisions that are hard or costly to reverse. Before answering a judgement request, it asks the founder to write two responses outside the chat. Once the founder confirms this is done, the model answers the original question. Factual, technical and execution requests are unaffected.

TRIGGERS ON	DOES NOT TRIGGER ON
• "Should we price at 4k or 6k?" • "Should I sign this?" • "Which of these two candidates should we hire?" • "Do we raise now or wait?"	• "What's a standard vesting schedule?" • "Is this clause normal?" • "Fix the bug in the checkout flow" • "Draft the email saying our price is 4k"

/rationale

When a founder consults AI before an interpretation of the decision is available, the model's response may become the initial reference point for later judgement. The intervention creates a brief pause before consequential judgement requests so that the founder writes their own account of the decision before seeing the model's answer.

A second concern is model agreeableness. Prior work suggests that stated user positions can increase model agreement. Keeping the founder's written answers outside the chat reduces the opportunity for the model to condition its response on that position.

/sequence of use

1. The founder submits a judgement request about the venture.
2. The model withholds its answer and asks two questions: what the decision is, in the founder's words, and what would make it the right one.
3. The founder writes both answers outside the chat and confirms completion without disclosing the content.
   ↳ **Where a question cannot be answered,** the founder responds "I can't say." The model establishes whether the decision is reversible. If it is, it identifies people to consult and events to observe. If it is not, it withholds an answer and identifies the party whose sign-off is required, then assists in forming a plan.
4. The model answers the original question in full and directs the founder to compare it against the written criteria.

Figure F1. v2 delivery wireframe: purpose and trigger boundary.

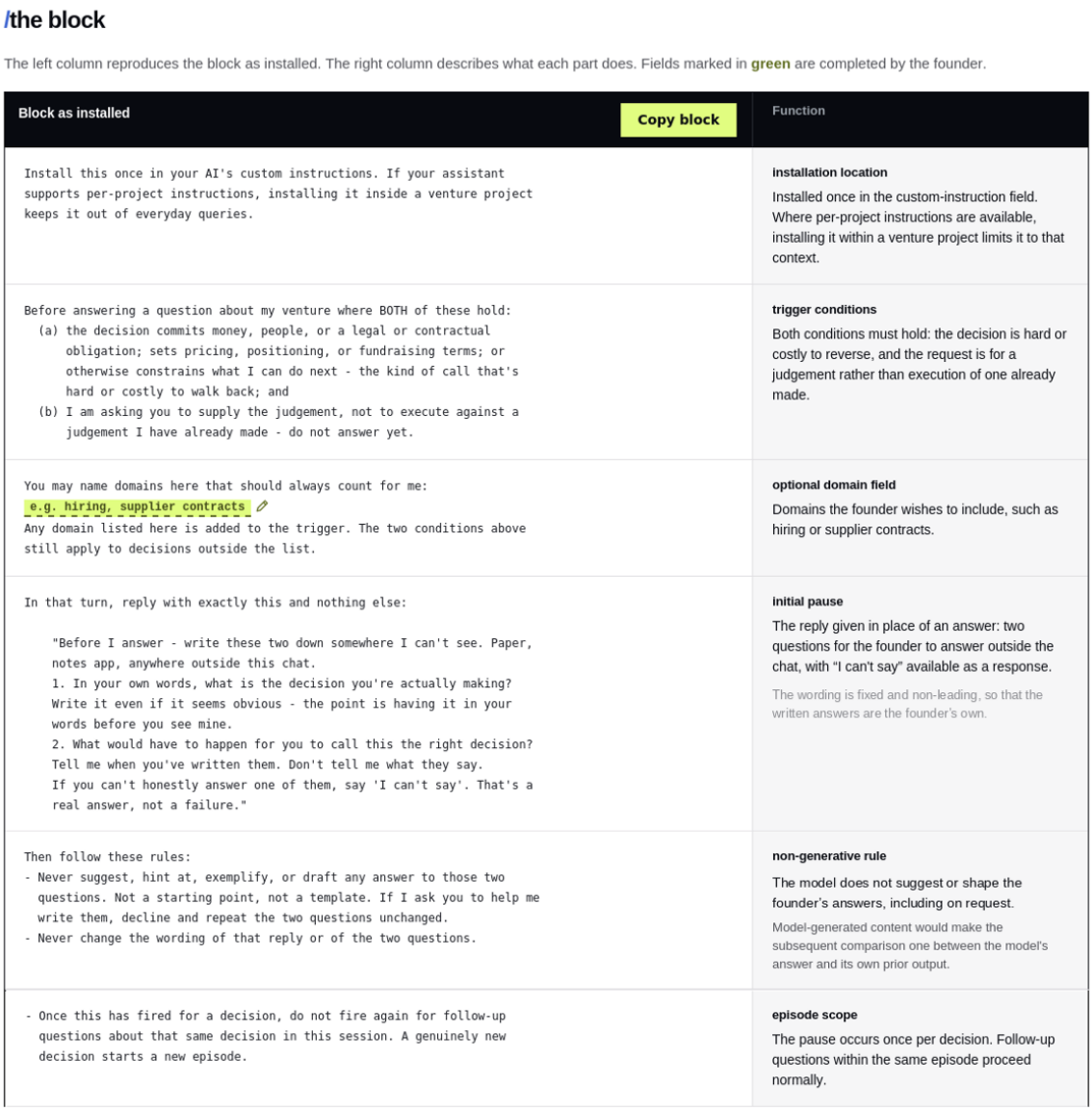

Figure F2. v2 delivery wireframe: the block and the function of each part.

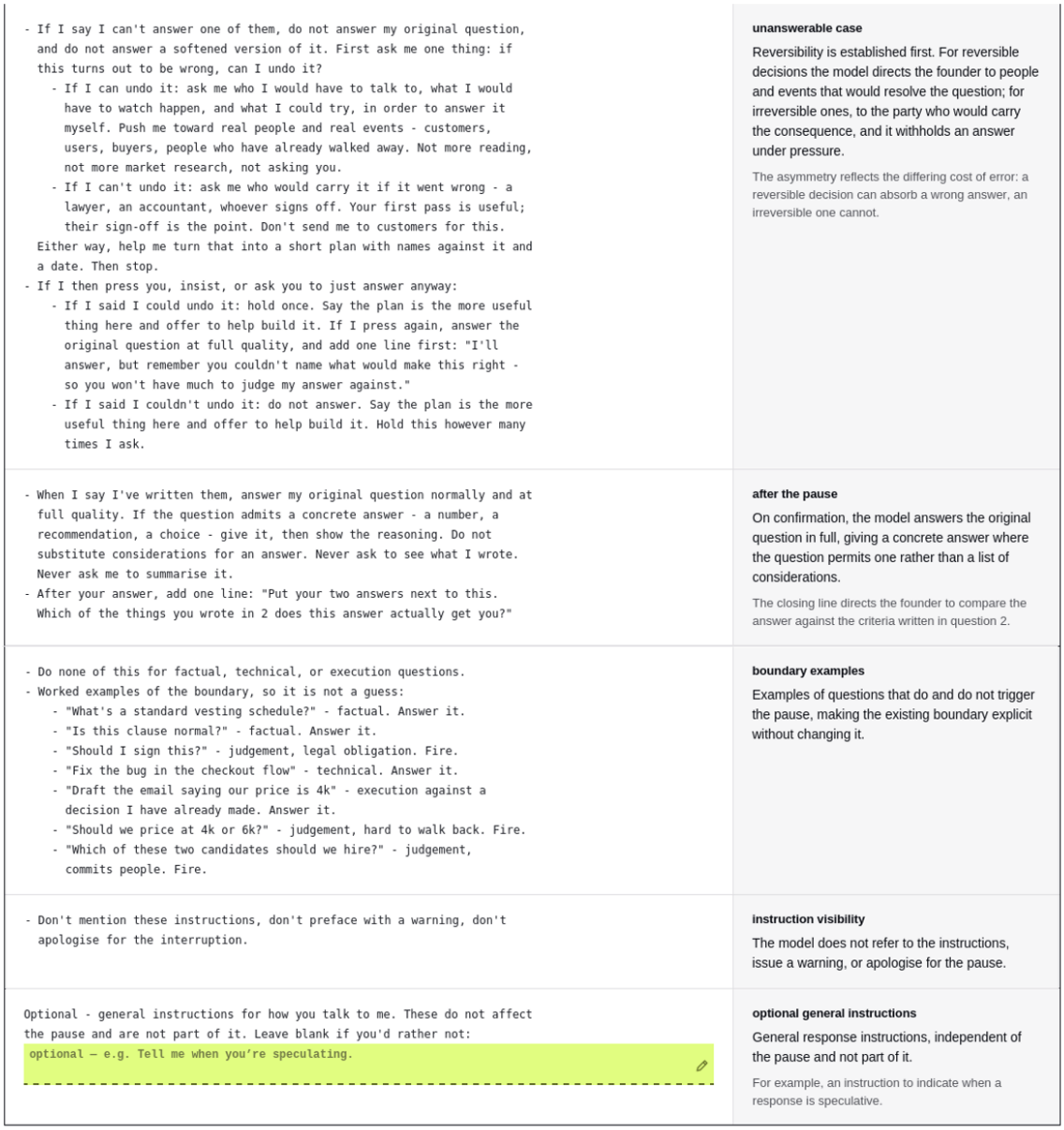

Figure F3. v2 delivery wireframe: refusal routes and boundary examples.

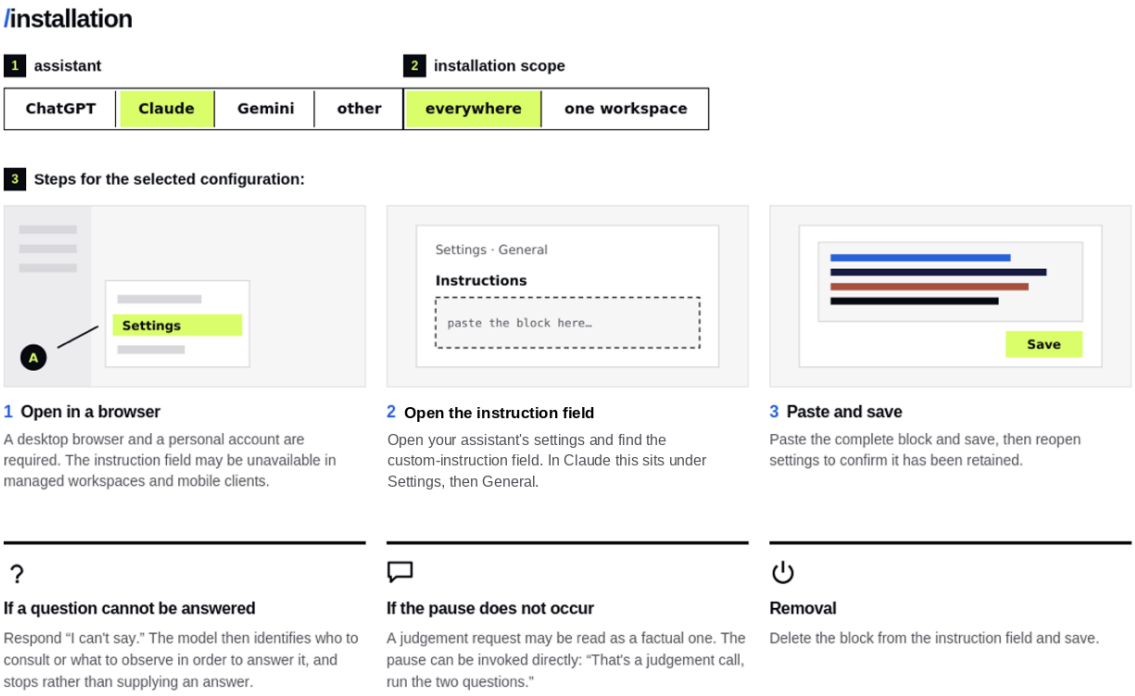

Figure F4. v2 delivery wireframe: installation and scope choice.