Think Through Touch: Designing a Tangible Multimodal System for Embodied Authorship in Human-AI Co-Ideation

Dahae Cheon

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Master of Design

University of Washington
2026

Reading Committee:
Meichun Liu
Jason O. Germany

Program Authorized to Offer Degree:
School of Art + Art History + Design

University of Washington

Abstract

Think Through Touch: Designing a Tangible Multimodal System for Embodied Authorship in Human-AI Co-Ideation

Dahae Cheon

Chair of the Supervisory Committee:
Meichun Liu
School of Art + Art History + Design

This thesis investigates how a tangible, multimodal interface can restructure human-AI co-ideation to support embodied cognitive engagement and authorial agency. As generative AI becomes a partner in thinking, its interface often remains a chat-based exchange: a record organized by time and operated through typed language. This thesis argues that this structure sits in tension with how people think — spatially, gesturally, and through the body — and that the resulting translation cost becomes cognitive. Over a long session, the human can gradually settle into the role of reader and editor of generated text, while authorship of the thinking process drifts toward the system.

In response, the thesis develops the Think-Through Framework, a conceptual model grounded in embodied and extended cognition and extended through Material Engagement Theory. The framework proposes that AI-generated dialogue, when rendered as spatially persistent and handleable digital artifacts, can begin to acquire the properties of cognitive material: something worked, not only read. It describes a loop of Externalize, Materialize, Handle, and Think-Through, in which bodily action and cognitive change continually inform one another, with the recovery of authorship as the condition the loop works toward.

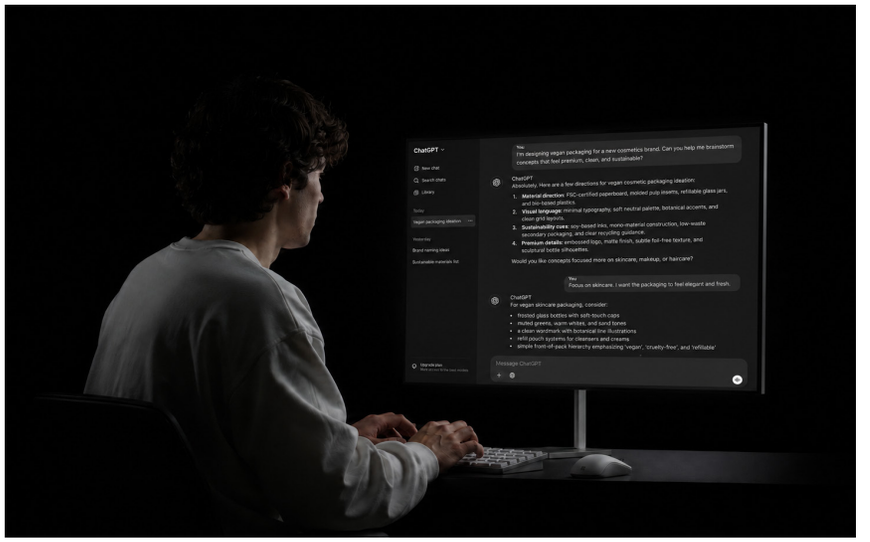
The framework is explored through a gesture-elicitation study with seven design-trained participants, who mapped instinctive gestures to twelve motion stimuli on a back-projected textile surface. The study suggests that motion can function as a grammar of interaction: its temporal quality shaped how participants distinguished momentary triggers from sustained control, while its spatial scope shaped how broadly the body was expected to engage. The study also suggests that the soft, resistant materiality of the surface contributed to the sense of handling digital content as if it had substance.

These insights inform FIELD, the software application developed within the Tangible Multimodal System (TMS): a network graph canvas where ideas hold spatial position, relationships become touchable, and a held fingertip and spoken phrase together direct a thought to its place.

The thesis contributes a conceptual proposition for materializing AI dialogue as handleable artifacts, a working design demonstration for embodied authorship in human-AI interaction, and formative insights for future tangible multimodal AI interfaces. It identifies comparative evaluation against conventional chat-based co-ideation as the central next step. Its proposition is simple to state and demanding to build: that a person might think not only with AI, but through touch.

I am deeply grateful to *Professor Meichun Liu* for guiding this thesis through a research lens and helping me understand Research through Design not only as a method, but as a way of thinking through making.
I would also like to thank *Professor Jason O. Germany* for helping me see the project through storytelling, language, and interpretation. His guidance helped me clarify not only what the project does, but what it means.
My time with the *Microsoft Surface Design Team* profoundly shaped this thesis. Working alongside a team committed to creating seamless experiences between hardware and software helped me ask a more holistic question: how interaction, interface, material, and system can come together as one experience.
I am especially thankful to *Yeha Chung*, who designed the new interface structure with me. Together, we went through an intense iterative process, building and revising hundreds of screens to bring the interface take shape.
I am also grateful to *Gahui Yun* for collaborating on the Motion-Gesture Mapping Study. Her research contribution helped ground the project in observation, analysis, and human response.
This thesis was shaped by the people who helped me think, question, build, and continue.

PART I
THE ARGUMENT



CHAPTER 01-04

CHAPTER 01. Introduction

A person sits down to think with an AI system. The interface before them is a text field, a blinking cursor, and a conversation that accumulates downward as a column of messages. The exchange begins simply: the person asks, the system responds. As the session deepens, however, the interface begins to shape the thinking it contains. Earlier ideas recede upward. Relationships among ideas must be remembered rather than seen. Returning to a prior direction requires scrolling, quoting, or retyping context that the interface has let slip away.

The difficulty is structural rather than a matter of model capability. The dominant chat interface organizes dialogue primarily by time and receives human intent primarily through typed language. For brief, task-oriented exchanges, both properties are efficient defaults. For sustained co-ideation, they become constraints on orientation, return, comparison, and intervention.

These constraints matter because the role of conversational AI is changing. For designers, researchers, writers, and strategists, generative AI increasingly participates in the exploratory stages of thinking: proposing directions, reframing problems, and surfacing unexpected associations. In such settings the interface is no longer a neutral channel of input and output. It is the environment in which thought takes shape, and its properties condition how the human remains oriented and responsible within the process.

Research in cognitive science and human-computer interaction indicates that human thinking is not confined to language, nor to the brain alone, but is shaped through bodily action, external representations, and interaction with the surrounding environment (Clark & Chalmers, 1998; Kirsh & Maglio, 1994; Barsalou, 2008; Dourish, 2001). People think with their hands, tools, and surroundings: they sketch before they can fully articulate, arrange materials to clarify a structure not yet decided, and retrieve ideas partly by where those ideas were encountered. Such actions do not merely express finished thought; they participate in forming it. The dominant chat interface nonetheless compresses the bodily and spatial dimensions of thinking into keyboard input and textual output, so that a spatial or gestural intention must first be translated into a sentence before the system can respond to it.

This thesis examines that translation cost and asks whether it can be addressed through design — not by refining the chat interface, but by reconsidering its premises. The design outcome of the inquiry is a proposed Tangible Multimodal System (TMS) and a software demonstration, FIELD. These names are introduced here only to identify the destination of the argument. Their architecture, interface structure, and feature are deliberately withheld until Chapter 9, after the theoretical framework, the research process, and the empirical findings have established why such an interface might matter and what it must accomplish.

The phrase that organizes the work is Think Through Touch. It names a design hypothesis: that when AI-generated dialogue acquires spatial persistence and becomes available to bodily action, users may better remain oriented within, intervene in, and redirect the development of ideas. The agency at stake is authorial rather than operational. Authorial agency, as used throughout this thesis, denotes the person's continued responsibility for the structure, direction, and meaning of an emerging thought. It is distinct from controlling every system operation, and it is precisely what a conventional interface can quietly erode by positioning the person as a reviewer of outputs produced elsewhere.

1.1 Motivating Question and Research Questions

This thesis is guided by the following motivating design question:

> How might interaction and interface evolve when conversational AI moves beyond linear text scrolls toward tangible, multimodal dialogue?
>
> From this question, the thesis develops two research questions:
>
> RQ1. In what ways does materializing human-AI dialogue as handleable digital artifacts support the formation of human authorial agency in AI-assisted co-ideation?
>
> RQ2. How does gesture-voice multimodal interaction with handleable human-AI dialogue artifacts support thinking-through within the co-ideation process?

The two questions divide the inquiry. RQ1 concerns a change of form: what becomes possible when AI dialogue is rendered as spatially persistent, handleable artifacts rather than as a transcript. RQ2 concerns a change of channel: how gesture and voice, acting on those artifacts, support thinking as an ongoing, embodied process.

The structure of the thesis follows this progression. Chapter 2 identifies the interface conditions that motivate the work: the linear scroll and keyboard-centered input. Chapter 3 develops the theoretical framework that explains why spatial, bodily, and material engagement matter for thinking, and extracts the insights on which the design argument rests. Chapter 4 translates those insights into the Think-Through Framework. Chapters 5 through 8 report the design-research process and synthesize the findings of the Motion-Gesture Mapping Study. Chapters 9 and 10 translate framework and findings into the designed system. Chapter 11 states contributions, limitations, and future directions; Chapter 12 concludes.

1.2 Scope and Posture

This thesis is situated within human-computer interaction and interaction design, drawing on cognitive science, material engagement theory, tangible computing, and research on human-AI collaboration. It does not claim that tangible AI interfaces are universally superior to text-based ones, nor that all AI dialogue should become spatial or gestural. It investigates a narrower proposition: that sustained AI-assisted co-ideation may benefit from interfaces in which ideas and relationships persist in space and can be acted upon through the body.

The empirical work reported here is exploratory and qualitative. Its findings inform and constrain the design logic of the proposed system; they are not presented as a validated gesture vocabulary or as generalizable evidence of effectiveness. They are treated as formative grounding: an account of how participants interpreted motion, materiality, depth, and resistance on a soft projected surface, and of how those interpretations can guide the design of tangible multimodal AI interfaces.

The system described in this thesis is a research prototype and a design argument. Its purpose is to make a possible interaction structure legible and examinable, not to propose a finished product. The working prototype demonstrates software logic and gesture-voice interaction under constrained technical conditions; the fuller material experience of pressure, depth, elasticity, and physical resistance remains a target for future development.

CHAPTER 2: Background

The interface that most commonly mediates human-AI ideation is the conversational chat interface: a text field for input, a scroll of exchanges for output, organized primarily by time. This chapter examines two structural properties of that interface that sit in tension with what research describes about human thinking. The first is the linear scroll as an organizational paradigm; the second is keyboard-centered input as the primary channel of human intent. The argument is that the two properties stem from a shared structural condition — inherited rather than designed — and that naming this condition is the precondition for designing differently.

2.1 The Limits of Linear Scroll

When a person converses with an AI system, the exchange typically accumulates as a vertical column of text. Each turn appends to the bottom; earlier content recedes upward and out of view. This is the linear scroll: an organizational model carried over from document processing, messaging, and web reading, and inherited by AI dialogue interfaces largely as a default.

The scroll provides movement through space but not meaningful spatial organization. Position carries little semantic meaning beyond sequence, and information is retrievable primarily in the order it was produced. For reading a transcript, this is adequate. For exploratory thinking, it becomes a constraint.

Research in spatial cognition indicates that human memory is not organized

only as a temporal sequence. People encode and retrieve information in relation to place — physical places in the world, but also constructed spatial schemas associated with layouts, diagrams, and environments. When an idea cannot be located in a meaningful space, when it can only be described as having occurred earlier in the conversation, the cognitive work of retrieval increases. The interface offers few spatial footholds for return. The practical consequence is a progressive loss of access. Ideas generated early in a session become harder to revisit, compare, or recombine as the session continues; a person who wishes to return to a prior thread must scroll through time or reconstruct the idea from memory, and neither strategy scales with the complexity or duration of sustained ideation. The scroll gradually becomes part of the work the user must manage rather than the container that supports it.

This property reflects an assumption embedded in chat-based interfaces: that dialogue is a sequence of messages whose primary unit of value is the most recent response. The assumption is reasonable for short, task-completion contexts. It fits poorly with exploratory ideation, where value accumulates across the session and where the relationships between ideas — not only the ideas themselves — are the substance of the work.

Figure 2.1. Problems emerging from the structural gap. When AI dialogue accumulates as a linear log, users may experience cognitive overload and a gradual loss of creative control.

2.2 Away from Keyboard

The keyboard is a technology for encoding language. It operates through discrete, symbolic input: a key is pressed, a character is produced. This precision serves text composition well. As the primary channel for communicating intent during open-ended ideation, however, keyboard-centered input imposes a translation cost that is easy to overlook precisely because it is habitual.

Human thought does not always begin in fully verbal form. Before a thought can be typed, it must be formulated as a sentence, segmented into words, and entered in sequence. Ideas that are spatial, relational, or gestural — ideas that first appear as orientations, proximities, or directions — must be converted into a representational format that cannot preserve all of their structure. The system receives a linguistic representation of the thought, not the bodily or spatial structure through which the thought emerged.

The cost concerns fidelity as much as effort. Cognitive science has documented that people routinely perform epistemic actions: manipulations of the environment carried out not to accomplish a task directly but to support the cognitive work of understanding it, as when a person points at an object while describing it or physically separates items to clarify a categorization. When keyboard input dominates the interaction, this dimension of cognition has limited means of expression.

The phrase away from keyboard — familiar in digital communication as a marker of absence — is useful here if inverted. It describes a design ambition: AI interaction in which the keyboard is no longer the sole conduit between human intent and machine response. The move away from keyboard-exclusive input is treated in this thesis not primarily as a matter of convenience or accessibility, though it may be both, but as a matter of cognitive architecture. A system that accepts only symbolic, sequential, language-like input receives only a portion of what human thinking produces.

2.3 A Structural Gap

The two limitations are distinct but not independent. The linear scroll constrains the spatial organization of AI dialogue; keyboard-centered input constrains the bodily expression of human intent. Together they produce an interface in which several of the ways people think — spatially, gesturally, relationally, continuously — have few direct points of entry.

This observation is not a critique of any particular product. The scroll and the keyboard are not aberrations; they are inherited infrastructure, shared across most current conversational AI interfaces. Stated carefully, the claim of this chapter is that such interfaces provide limited structural support for two things that matter in sustained ideation: meaningful spatial organization and bodily expression.

Whether this limitation constitutes a cognitive problem rather than a stylistic difference is a theoretical question, and it is the subject of the following chapter.

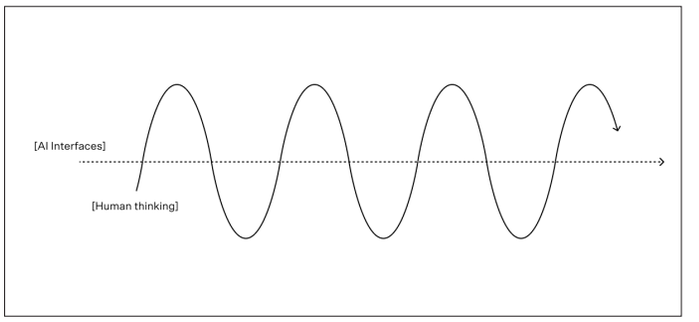


Figure 2.2. The structural gap. Current AI interfaces move linearly through text, while human thinking moves spatially, gesturally, and continuously. This mismatch frames the design problem of the thesis.

CHAPTER 3: Theoretical Framework

Chapter 2 closed with an observation: dominant chat-based AI interfaces are organized with limited spatial structure and operated through limited bodily engagement. This chapter asks why that observation matters — why the absence of spatial organization and bodily expression might constitute a cognitive problem rather than a design preference.

The chapter develops four theoretical layers and extracts from each a working insight. Embodied and extended cognition explain why thinking can involve the body, external representations, and the surrounding environment. Material Engagement Theory explains why the properties of a material environment shape the actions through which thought develops. Co-location, spatial memory, and cognitive load explain why the relationship between where information is seen and where action occurs matters for orientation and return. Mixed-initiative interaction and generative latent space supply the context for treating human-AI collaboration as a negotiated exchange rather than a command sequence. The purpose throughout is not to anticipate the designed system, but to assemble the premises from which the conceptual framework of Chapter 4 is built.

3.1 Embodied and Extended Cognition: The Foundation

The premise underlying this thesis is that human cognition is not confined to the brain. Two related traditions establish it.

Embodied cognition holds that cognitive processes are shaped by the body's structure, sensorimotor capacities, and physical interaction with the environment. On this view, concepts are not only abstract symbols; they are frequently grounded in bodily experience — in moving through space, exerting force, orienting attention, and manipulating objects. To think about expansion, connection, or compression may draw on patterns originally formed through physical experience, and conceptual content can retain traces of the bodily experience through which it was learned.

Extended cognition develops this premise further. As Clark and Chalmers (1998) argue, the boundaries of cognition need not end at the skin. When an external resource — such as a notebook, a map, or a spatial arrangement of objects — performs a function that would otherwise be carried out internally, it can become part of the cognitive system itself. In this view, the notebook is not simply a record of finished thought, but, under certain conditions, part of the thinking process (Clark & Chalmers, 1998).

The implication for interface design is considerable. An interface that structures and stores information is not only a display; it is a candidate component of the cognitive system of the person who thinks through it, and its organizational properties may shape the cognitive properties of the work it supports. The insight this section contributes is therefore foundational: the spatial and bodily properties of a human-AI interface are cognitive

Figure 3.1. Gesture as thought in motion. Gesture is treated not as expression after thought, but as part of thought in formation (McNeill, 1992; Goldin-Meadow et al., 2001).

concerns, not aesthetic or ergonomic ones. How information is organized in space, and how the body can act in relation to it, conditions what can be thought through it.

3.2 Material Engagement Theory: The Primary Lens

Material Engagement Theory (MET) provides the primary lens for understanding why material conditions matter to cognition. Developed by Lambros Malafouris, MET argues that human cognition and material culture are not separate domains, but are co-constituted through ongoing material engagement (Malafouris, 2013). The properties of the materials people engage with — resistance, texture, weight, responsiveness, and constraint — shape the actions through which cognitive processes unfold. Thinking does not happen entirely inside the head and then find expression through materials; it happens, at least in part, through the dynamic relation between brain, body, and material world (Malafouris, 2013, 2019).

The operative point is not that material objects think on behalf of people. It is that material properties shape the actions available to a person, and those actions can participate in the development of thought. A surface that can be pressed, stretched, arranged, marked, or erased invites different forms of action than a surface that can only be read, and the difference matters because action changes what the person notices, remembers,

tests, or commits to. Materiality contributes to cognition by structuring the bodily actions through which thinking proceeds.

This formulation is what allows MET to be applied to digital and AI-mediated contexts without overstatement. The thesis does not claim that AI-generated content becomes literal physical material. It asks what follows when digital content is given some of the interactional conditions associated with material engagement: location, persistence, manipulability, tactile feedback, constraint, and consequence. Under such conditions, digital content may become workable as material for thought. The claim is relational rather than literal — material conditions shape action, and action shapes thinking — and the insight this section contributes follows directly: whether AI dialogue can function as cognitive material depends not on what it is made of, but on what it allows the thinker to do.

3.3 Co-location, Spatial Memory, and Cognitive Load

Three converging lines of research explain why the spatial organization of information bears on the quality of thinking: co-location in tangible interface research, spatial memory in cognitive psychology, and cognitive load theory.

In tangible user interface research, co-location names the condition in which the representation of information and the means of controlling it occupy the same physical space. In many screen-based interfaces these are separated: output is read in one location, input is entered in another, and the user bridges the gap between perception and action. Co-location collapses this separation; the user acts on what they see, in the place where they see it.

Spatial memory research offers a complementary rationale. O'Keefe and Nadel's account of the hippocampus as a cognitive map established that spatial location plays a structural role in memory and navigation, and subsequent spatial cognition research indicates that people use location, layout, and mental mapping to encode and retrieve information. For extended ideation, the implication is that persistent spatial arrangement may allow users to return to ideas by place rather than reconstructing them from a temporal transcript.

Cognitive load theory adds the third layer. When an interface forces repeated translation between what is seen, what is meant, and where action must occur, it imposes extraneous load: effort generated by the structure of the task environment rather than by the substance of the thinking. Co-locating representation and action may reduce this burden, freeing cognitive resources for interpreting, comparing, and developing ideas rather than managing the interface that contains them.

Taken together, these theories yield a design principle without specifying a design — and this is the insight the section contributes: if AI dialogue can be organized so that ideas persist in meaningful locations, and if users can act in relation to those locations, the interface may support orientation,

Figure 3.2. Thinking through spatial structure. Ideas are not only remembered by sequence, but by position, relation, distance, and return. Spatial arrangement gives thought a map to move through (Tversky, 2019; O'Keefe & Moser, 2014).

return, and intervention more effectively than a linear transcript.

3.4 Mixed-Initiative Interaction and Latent Space

Two further contexts inform the thesis without constituting its primary justification. Mixed-initiative intweraction addresses the dynamics of initiative in human-computer collaboration: who leads, who follows, and how transitions between leadership states are negotiated. In conventional AI dialogue this negotiation is conducted almost entirely through language — a person requests expansion, convergence, or critique, and the system responds. The value of mixed-initiative theory here is its framing: human-AI co-ideation is an ongoing negotiation of direction, not a one-directional command sequence, and the channels through which that negotiation is conducted are themselves open to design.

Latent space, in the context of generative AI, refers to the high-dimensional representational space in which a model organizes learned patterns; input activates regions of this space, and output emerges from those activations. This thesis makes no attempt to visualize latent space literally. The concept serves instead as a reminder that generative systems operate through relational possibilities not directly visible to users — and that interface design can make some of those emerging relations more inspectable, negotiable, and open to human judgment.

The four layers converge on a single argument, and each contributes a premise that the conceptual framework of Chapter 4 must satisfy.
First, from embodied and extended cognition: space, body, and external representation can be constitutive features of a thinking environment, so an interface is a cognitive condition rather than a neutral container.
Second, from Material Engagement Theory: materiality matters because it shapes action, and action participates in thought, so the cognitive value of materializing AI dialogue lies in what it allows the thinker to do. Third, from co-location, spatial memory, and cognitive load: when ideas persist in meaningful locations and action coincides with perception, orientation and return are supported and extraneous effort is reduced. Fourth, from mixed-initiative interaction and latent space: the human-AI exchange is a negotiation of direction whose channels are designable, operating over relational possibilities that an interface can help make visible.
These premises do not prescribe an interface. They specify what a conceptual model of embodied human-AI co-ideation must describe: how intent is externalized beyond language, how AI output becomes available for action, how the person works with it, and how that work contributes to thought. The model that answers this specification — the Think-Through Framework — is the subject of the following chapter.

CHAPTER 4: Think-Through Framework

Chapter 3 closed with four premises: the interface is a cognitive condition; the value of materialization lies in what it allows the thinker to do; spatial persistence and co-located action support orientation and return; and the human-AI exchange is a designable negotiation of direction. This chapter turns those premises into a framework. The goal is not yet to describe an interface. It is to define the interaction logic that the research of Part II will examine and the design of Part III will demonstrate.
Think-Through Framework describes how a person might think with AI-generated content when that content is treated not only as text to be read, but as materialized dialogue that persists, can be referenced, and can be acted upon. The framework comprises four stages — Externalize, Materialize, Handle, and Think-Through — which identify recurring dimensions of a loop rather than a rigid sequence: a person frames an intention, encounters AI-generated material, works with it through bodily action, and uses that work to continue thinking.
The framework fixes how this thesis understands authorial agency. Authorial agency is not operational control; it does not mean determining every system action. It means remaining responsible for the structure, direction, and meaning of an evolving thought. The framework asks how an interface can keep the person situated within the formation of ideas, rather than positioned outside the process as a reviewer of outputs.

The foundational premise of the framework is that AI-generated content must change form before it can be worked with through the body. In a conventional conversational interface, output arrives as flowing text: linearly structured, temporally ordered, readable, appended to the end of a transcript. Text can be scanned, copied, quoted, and edited; it is difficult to locate spatially, compare relationally, or handle as part of an ongoing field of thought.
Materialized dialogue names a different condition, defined by three properties: location, persistence, and manipulability. Location means an idea can be found somewhere rather than only earlier. Persistence means it remains available as the session develops. Manipulability means it can be acted upon — moved, connected, transformed, removed — rather than only read and rephrased. The claim is not that digital content becomes physical material in a literal sense; it is that content holding these three properties becomes available for material-like engagement, as material for thought.
This transformation is the precondition for the rest of the framework. Once AI dialogue holds the three properties, the person no longer depends on linguistic recontextualization alone to return to an idea. They can return by place, relation, and action.

4.2 Externalize: Making Intent Visible

The first stage is Externalize. Before a system can respond meaningfully, the person must make intent available. In conventional chat this happens almost entirely through language: the typed prompt is the system's main evidence of what the person wants.
The framework widens this notion of intent. Intent is also expressed through where attention is directed, what is pointed toward, how content is grouped, how long contact is held, and how the body is oriented in relation to an idea. These actions do not replace language; they supplement it, allowing situated aspects of thought — reference, emphasis, uncertainty, proximity, direction — to enter the interaction without being fully converted into sentences. Externalize names the moment when intent becomes visible as a combination of semantic expression and bodily orientation: the point at which an emerging thought acquires a form the interface can respond to.

4.3 Materialize: Giving AI Output Location and Persistence

The third stage is Handle. Handle names the range of actions through which a person works with materialized dialogue, and it is where the difference between receiving an AI output and shaping an AI-assisted

thought becomes most concrete.
To handle is not simply to operate a control. It is to treat AI-generated content as something that can be opened, moved, connected, set aside, revised, or removed — actions that allow the person to test relations, commit to directions, withdraw emphasis, and create new structure without translating every intention into a new prompt.
Handle is where materiality becomes cognitively consequential. Touching or moving content is not automatically meaningful; the point is that actions performed on persistent, situated content can become epistemic actions — actions that help the person understand what they think, what belongs together, what should be discarded, and where the inquiry should go next.

4.3 Materialize: Giving AI Output Location and Persistence

The second stage is Materialize. If Externalize describes how human intent becomes available to the system, Materialize describes how the system's response becomes available to the person. In conventional chat the response is appended as another turn in a transcript; in the framework proposed here, the response enters a spatial field as content holding the three properties defined in 4.1, where it can be revisited and worked with.
Materialize is where the premises of Chapter 3 take conceptual form. If external representations can support thought, and if material conditions shape the actions through which thought develops, then AI output must become more than a message. It must become an object of engagement within the thinking process.

4.4 Handle: Working with Materialized Dialogue

The third stage is Handle. Handle names the range of actions through which a person works with materialized dialogue, and it is where the difference between receiving an AI output and shaping an AI-assisted thought becomes most concrete.
To handle is not simply to operate a control. It is to treat AI-generated content as something that can be opened, moved, connected, set aside, revised, or removed — actions that allow the person to test relations, commit to directions, withdraw emphasis, and create new structure without translating every intention into a new prompt.
Handle is where materiality becomes cognitively consequential. Touching or moving content is not automatically meaningful; the point is that actions performed on persistent, situated content can become epistemic actions — actions that help the person understand what they think, what belongs together, what should be discarded, and where the inquiry should go next.

Figure 4.1. Thinking through touch. Touch becomes a way of holding thought in formation. Image by author.

4.5 Think-Through: Action as a Method of Thought

The fourth stage is Think-Through. It names the cognitive dimension of what Handle describes physically: where Handle names the action, Think-Through names what happens through the action in the development of thought.
A pragmatic reading of interaction describes task execution: the person acts, the system responds. A Think-Through reading recognizes that the action may also change the person's understanding. Moving two ideas closer together tests whether a relation holds; removing an idea clarifies the boundary of a direction; holding attention on one artifact makes its weight apparent. The thought is not always completed before the action — it may be clarified through it.
This is the central claim of Think Through Touch as a framework: the manipulation of materialized AI dialogue can participate in cognition, not merely represent cognition after the fact. Surface, artifact, gesture, and response form a loop in which the person continues to think by acting.

4.6 The Framework as a Whole: Diagram and Purpose

Taken as a whole, the Think-Through Framework differs from conventional chat-based dialogue in three respects. It is spatial rather than only

temporal: content persists in place rather than receding into a scroll. It is bodily rather than exclusively symbolic: gesture, touch, orientation, and voice carry human intent alongside language. And it is iterative rather than sequential: its core is a loop between action and cognitive consequence, not a pipeline from input to output.

Diagrammatically, the framework begins with human intent. Intent is externalized through language, gesture, and orientation; AI output is materialized as spatially persistent content; the person handles that content through situated action; and through this handling, thinks through the material — questioning, connecting, revising, releasing, reframing. The result is not simply an output. It is a restructured field of thought the person has actively shaped.

The framework is the bridge between theory and method. Chapter 5 turns it into a research plan; Chapters 6 through 8 examine how people interpret motion, gesture, materiality, depth, and resistance on a soft projected surface; and Part III translates those findings into design.

Figure 4.2. Think-Through Framework. Intent becomes visible, dialogue becomes material, and action becomes part of thought.

PART II
THE RESEARCH

CHAPTER 05-08

CHAPTER 5: Research Framing and Methodology

5.1 From Research Questions to Research Design

Chapter 1 stated the motivating question and the two research questions that guide this thesis. This chapter returns to them from a methodological standpoint: how a conceptual proposition about embodied human-AI co-ideation was translated into a sequence of design-research activities.
In compressed form, RQ1 asks what the materialization of AI dialogue contributes to authorial agency, and RQ2 asks how gesture-voice interaction with materialized dialogue supports thinking-through. Neither question can be answered by evaluating an existing system, because the interface condition they presuppose — spatially persistent, handleable AI dialogue on a responsive material surface — does not yet exist in mature form. The methodological task was therefore to construct that condition in examinable layers: a material substrate, a bodily action space, and a relationship between visual motion and user expectation.

5.2 Research through Design Approach

The methodological posture of this thesis is Research through Design (RtD): an approach in which the iterative making of artifacts functions as a method of inquiry, and in which knowledge is produced through the act of designing rather than only applied to it.
RtD suits the questions this thesis poses. The research does not ask whether an existing system performs against a benchmark; it asks what becomes possible under a design condition that does not yet fully exist. Answering required constructing the condition — the material substrate, the gesture space, and the study situation through which tangible multimodal dialogue could be experienced and examined. Within this posture, empirical study and design practice informed each other in sequence: material exploration narrowed the design space, gesture exploration mapped the bodily possibilities of the chosen material, and a participant study examined how people responded to motion-based artifacts on that material.

5.3 Three-Phase Study Plan

The research consists of three connected phases — a material-interaction probe, a gesture configuration matrix, and a motion-gesture mapping study — each addressing a different layer of the proposed interface condition: the material substrate, the bodily action space, and the relationship between visual motion and user expectation.

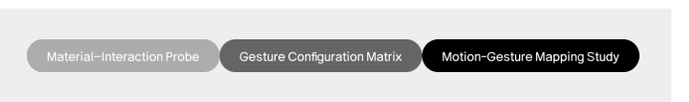


Figure 5.1. Three-Phase Study Plan

Phase 1 — Material-Interaction Probe

The first phase asked which material could function as a surface for handling AI dialogue by hand. Because the proposed interaction logic depends on pressure, stretch, resistance, and depth rather than on discrete touch alone, the research first needed a material substrate capable of supporting it.

The probe compared two candidate surfaces. A non-stretch membrane afforded mostly discrete input — contact or no contact. A stretch membrane suggested a wider analog range: light, medium, and deep press; slow and fast stretch; sustained tension; gradual release. Although it introduced calibration complexity and required structural support for consistent depth sensing, the stretch membrane was selected as the interaction substrate most conceptually aligned with exploring pressure, depth, elasticity, and material resistance.

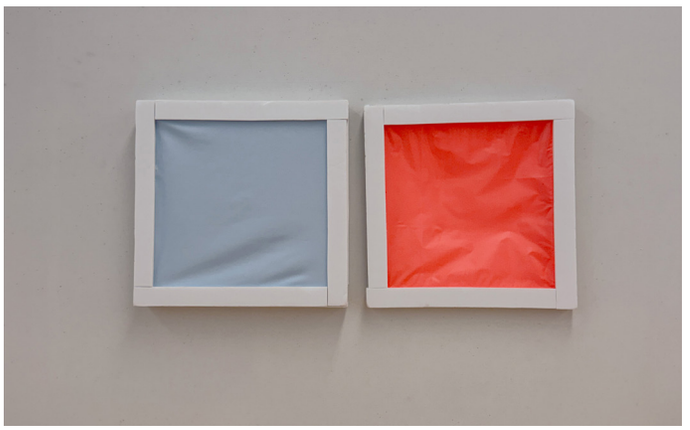


Figure 5.2. Comparing Stretch and Non-Stretch Surfaces. Two membrane conditions were tested at the same 30 cm × 30 cm scale to compare their potential as interaction surfaces. Image by author.

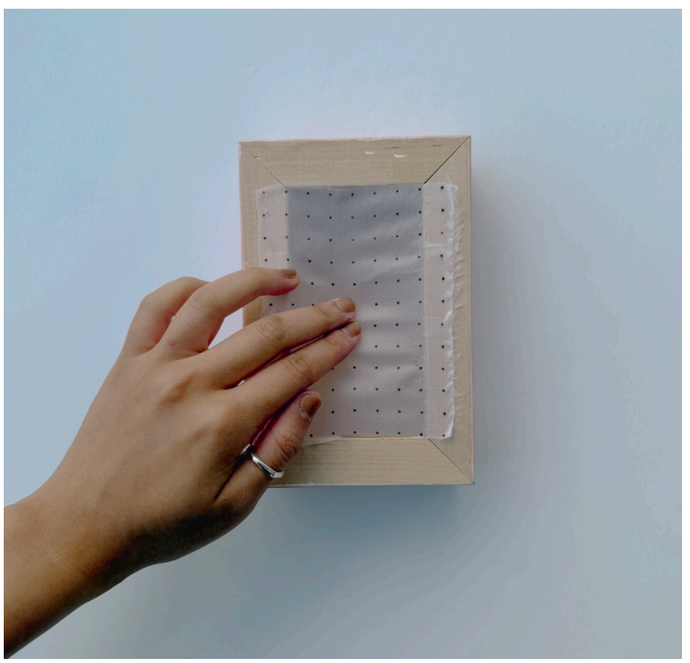


Figure 5.3. The Limit of Discrete Touch. The non-stretchable surface offered contact, but little material dynamics beyond touchpad-like input.

Figure 5.4. Material Variations. Layered and partially overlapped fabrics were explored as possible ways to introduce richer tactile behavior.

Phase 2 — Gesture Configuration Matrix

The second phase catalogued the gesture families and variants the selected membrane appeared to afford. Conducted as a first-person exploratory probe, its purpose was to map the basic bodily action space of the material before introducing participants, AI content, or motion-based tasks.

Direct exploration identified five high-level gesture families — tap, press, drag, sweep, and lift — each with variants based on duration, pressure, direction, finger count, contact area, and trajectory.

The resulting gesture configuration matrix served as a descriptive scaffold for the participant study, defining how observed gestures could later be named and coded while showing that the membrane supports a richer range of action than discrete touch.

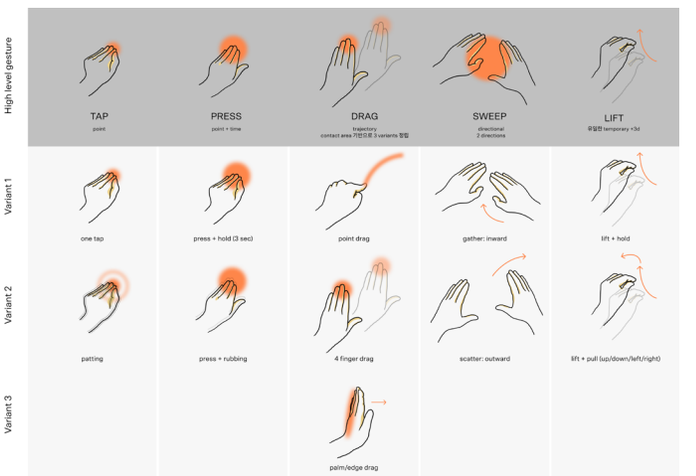


Figure 5.5. Gesture Configuration Matrix Five high-level gesture families — tap, press, drag, sweep, and lift — and their variants identified through first-person exploratory probing of the stretch membrane.

Phase 3 — Motion-Gesture Mapping Study

The third phase examined how visual motion behavior shaped participants' instinctive gestures, their interpretations of system function, and their responses to the textile surface. The phase was necessary because the proposed interface represents AI-generated

content as moving, transforming, spatially located artifacts; before specifying final interaction behavior, the research needed to understand what gestures and meanings such motion invites.
Phase 3 produced two connected outputs: a structured UX research dashboard, which organized the qualitative data into a traceable visual analysis structure, and a set of synthesized findings distilling the recurring patterns. Together, they provide the empirical basis for understanding how motion, gesture, and textile materiality shaped expectations of a tangible multimodal AI interface.

CHAPTER 6: Motion-Gesture Mapping Study

This chapter documents the design of the third research phase: a gesture-elicitation study examining how motion-based artifacts on a textile surface shape the gestures people propose and the meanings they assign. The analysis of the resulting data is the subject of Chapter 7.

6.1 Study Design

The study asked how temporal quality and spatial scope shape users' proposed gestures and interpretations when interacting with motion-based AI dialogue artifacts.
The design crossed three dimensions. The first was ideation theme: Expansion, Bridging, and Narrowing, corresponding to three kinds of thinking action within co-ideation — opening or exploring content, connecting or combining ideas, and condensing or reducing information.

Design Axis	Levels	Definitions	Design Rationale
Temporal Quality	Sudden Sustained	Describes the speed and onset characteristics of a motion. Sudden motions appear or complete within a short time frame, while sustained motions unfold gradually over an extended duration.	This axis was selected to examine whether motion timing influences the speed, intensity, and continuity of users' proposed gestures.
Spatial Scale	Global Local	Describes the extent of the textile surface engaged by a motion. Global motions occupy the majority of the surface, while local motions remain confined to a specific region.	This axis was selected to examine whether the scale of projected motion shapes the range of bodily engagement, from single-finger gestures to whole-hand or multi-hand movement.
Ideation Theme	Expansion Bridging Narrowing	Describes the cognitive action represented by the motion: opening or exploring content, connecting or combining ideas, or condensing and reducing information.	This axis was selected to connect motion behavior to different forms of thinking within AI-assisted co-ideation.

Table 6.1. Motion Design Matrix
Temporal quality and spatial scale were crossed to define the motion conditions used in the study.

The second was spatial scope: global motions engaged the broader surface and suggested surface-level transformation, while local motions remained within a smaller region and suggested precise interaction with a specific artifact or cluster. The third was temporal quality: sudden motions appeared or completed quickly, while sustained motions unfolded gradually.
Crossing three themes with two spatial scopes and two temporal qualities produced a twelve-motion stimulus system, presented under a single material condition: a back-projected, stretchable textile surface.
The study was conducted under a University of Washington IRB-approved protocol, STUDY00025547, and followed a gesture-elicitation paradigm: participants viewed a motion animation and proposed the gesture they would expect to trigger or guide it. The method was chosen because the goal was not to validate a predefined gesture set, but to surface intuitive mappings between motion, meaning, and bodily action before the system's final interaction language was specified.

	Global	Local
Sudden	Motion appears or completes abruptly across the full textile surface.	Motion appears or completes abruptly within a confined region of the surface.
Sustained	Motion unfolds gradually across the full textile surface.	Motion unfolds gradually within a confined region of the surface.

Table 6.2. Motion Condition Matrix
Sudden/Sustained temporal quality and Global/Local spatial scale were crossed to define the 4 motion conditions used in the study.

6.2 Participants and Procedure

Seven participants took part, referred to anonymously as P1–P7. All had design backgrounds, ranging from graduate students to professional designers with approximately seven to fifteen years of design-related experience including undergraduate training. All also reported regular use of AI tools, including large language models and image generation, in creative work.
The group was selected because the study concerns AI-assisted co-ideation rather than general touch usability. Participants therefore needed sufficient creative and professional experience with AI tools to interpret abstract motion not only as animation, but as potential information behavior within an ideation system.
Each session lasted approximately thirty to forty minutes. After introduction and consent, a short background interview established the participant's design practice and AI use. A textile sensitizing activity followed, in which participants explored an eight-by-eight-inch swatch of the stretch membrane while thinking aloud, becoming familiar with its softness, elasticity, and depth response before the elicitation task. In the

core elicitation, participants viewed twelve motion animations on the back-projected textile surface; for each, they proposed a gesture that could trigger or guide it, described what feature or information behavior the motion might represent, and reflected on whether the material influenced their response. The session closed with a short reflection on gesture comfort, motion-gesture fit, and the perceived difference between the textile surface and conventional glass touchscreens.

Five forms of data were collected: hand-movement video, audio recording, observation notes, gesture proposals, and participants' interpretations of motion and material experience. The study was reviewed under University of Washington IRB protocol STUDY00025547; participation was voluntary, faces were not recorded, and recordings were retained for research use only.

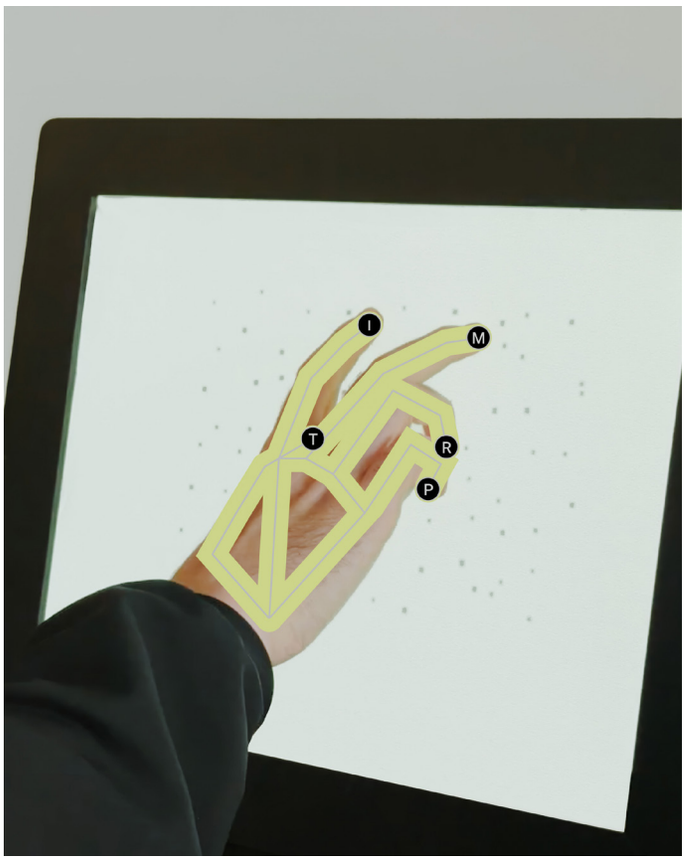


Figure 6.1. Material Variations. Layered and partially overlapped fabrics were explored as possible ways to introduce richer tactile behavior.

Ideation Theme	Global + Sudden	Global + Sustained	Local + Sudden	Local + Sustained
Expansion	M1	M2	M3	M4
Bridging	M5	M6	M7	M8
Narrowing	M9	M10	M11	M12

Table 6.3 Stimulus Condition Matrix
M1–M12 identify the twelve video stimuli constructed from three ideation themes and four motion conditions.

Table 6.3 summarizes the twelve motion stimuli using condition-based labels that identify each stimulus by ideation theme, spatial scale, temporal quality, and visual behavior.

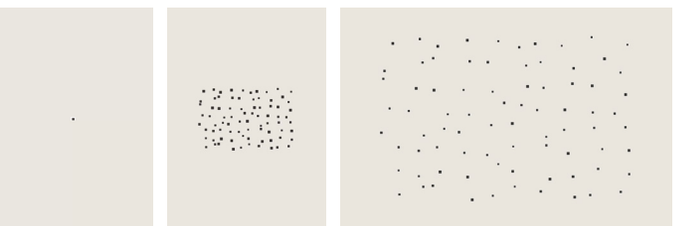
Each stimulus was designed as a transition between an initial state and a resulting motion state. The small dots represent handleable digital artifacts, each implying a unit of information within the projected field. Participants were asked to interpret how the motion occurred and to propose what gesture input would trigger or guide the transition. In this sense, the stimuli functioned not as finished interface animations, but as motion prompts for eliciting gesture expectations.

M1 — Expansion / Global / Sudden

A compact cluster of items rapidly disperses across the full surface.

M2 — Expansion / Global / Sustained

A compact cluster of items slowly disperses across the full surface.



M3 — Expansion / Local / Sudden

A compact group of items rapidly emerges outward, completing the wider field.

M4 — Expansion / Local / Sustained

A compact group of items slowly emerges outward, gradually completing the wider field.

M5 — Bridging / Global / Sudden

One item from each distant cluster rapidly moves inward until the two meet at the center.

M6 — Bridging / Global / Sustained

One item from each distant cluster slowly moves inward until the two meet at the center.

M7 — Bridging / Local / Sudden

One item from each nearby cluster rapidly moves inward until the two meet at the center.

M8 — Bridging / Local / Sustained

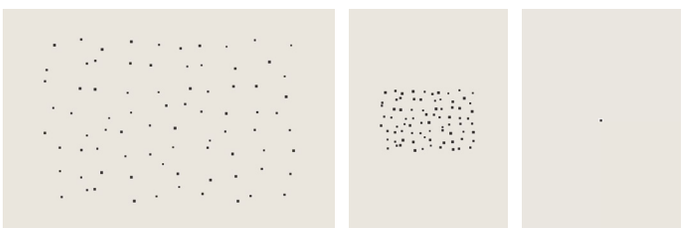
One item from each nearby cluster slowly moves inward until the two meet at the center.

M9 — Narrowing / Global / Sudden

Items scattered across the full surface rapidly collapse into a compact cluster.

M10 — Narrowing / Global / Sustained

A scattered field of items slowly converges into a compact cluster.

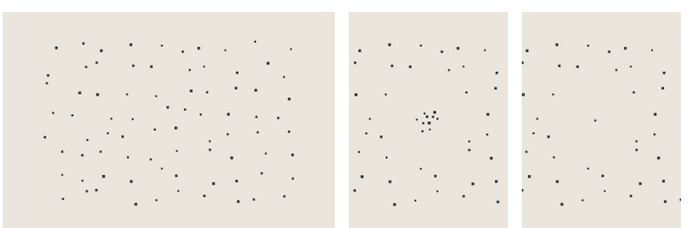


M11 — Narrowing / Local / Sudden

Items within a confined region rapidly converge into a compact group.

M12 — Narrowing / Local / Sustained

Items within a confined region slowly converge into a compact group.



CHAPTER 7: Analysis

This chapter describes how the study data were analyzed: how raw observations were organized, how a coding framework was developed, and how cross-participant patterns were synthesized into findings. The analysis moved from raw data, to coded interpretation, to a structured UX research dashboard that served as the main visual analysis environment.

7.1 Analysis Pipeline and Coding Framework

The analysis proceeded in five stages. First, raw study data — hand video, audio transcripts, gesture proposals, motion interpretations, and material comments — were organized on an analysis board. Second, the data were reviewed participant by participant and motion by motion. Third, recurring patterns were coded across three dimensions: motion–gesture correspondence, motion–meaning interpretation, and material-mediated interaction. Fourth, the coded material was organized into a structured UX research dashboard. Finally, the dashboard was used to compare patterns across participants, motions, gesture families, and interpretations, forming the basis for the synthesized findings in Chapter 8.

The first coding dimension, motion–gesture correspondence, examined how visual motion qualities shaped proposed gestures, coding for gesture type, hand and finger use, direction, duration, pressure, and scale. The second, motion–meaning interpretation, examined how participants read each motion as an information behavior — expanding, revealing, connecting, merging, summarizing, cleaning up, navigating, or generating. The third, material-mediated interaction, examined how the textile altered bodily engagement and sense of control, including softness, resistance, elasticity, tactile feedback, Z-axis pressure, and commitment.

Coding Dimension	What It Examined	Coded For
Motion–Gesture Correspondence	How motion qualities shaped proposed gestures	gesture type, hand/finger use, direction, duration, pressure, scale
Motion–Meaning Interpretation	How participants interpreted motion as information behavior	expand, reveal, connect, merge, summarize, clean up, navigate, generate
Material-Mediated Interaction	How the textile altered bodily engagement and control	softness, resistance, elasticity, tactile feedback, Z-axis pressure, commitment

Table 7.1 Analytic Coding Framework
Three coding dimensions used to translate raw observations into structured findings.

The coded data were built into a structured UX research dashboard that organized gesture selections, interpretations, quotes, and material-experience patterns. This allowed comparison across participants and motions without reducing situated responses to isolated anecdotes.

7.2 Patterns Across the Three Coding Dimensions

Motion–Gesture Correspondence

Across the coded proposals, pinch was the most frequently proposed gesture family, followed by tap, hold, swipe, sweep, and poke. The strongest pattern was the relationship between temporal quality and gesture character: sudden motions tended to invite single, immediate gestures such as tap or quick swipe, while sustained motions more often invited maintained contact, such as hold or slower pressure-based engagement.

This pattern suggests that participants read motion speed as a cue for the kind of input expected. Sudden motion was often interpreted as a trigger-based system response; sustained motion suggested that the user might remain involved in controlling, pacing, or sustaining the action.

MOTION	P1	P2	P3	P4	P5	P6	P7	CONSENSUS
M1 Single→Multi Expand	Pinch	Pinch	Pinch	Tap	Pinch	Pinch	Pinch	86%
M2 Cluster Expand (Fast)	Hold	Tap	Pinch	Tap	Tap	Tap	Tap	71%
M3 Cluster Expand (Slow) Low	Hold	Pinch	Pinch	Hold	Pinch	Hold	Pinch	67%
M4 Single Pop-up	Tap	Tap	Poke	Tap	Tap	Tap	Tap	86%
M5 Dual Cluster Merge (Fast)	Swipe	Swipe	Swipe	Hold	Swipe	Swipe	Swipe	86%
M6 Dual Cluster Merge (Slow)	Hold	Swipe	Swipe	Hold	Swipe	Swipe	Swipe	71%
M7 Proximity Merge	Pinch	Swipe	Pinch	Hold	Pinch	Pinch	Pinch	71%
M8 Proximity Merge (Slow) Low	Hold	Hold	Pinch	Hold	Pinch	Pinch	Pinch	57%
M9 Full Converge (Fast) Low	Sweep	Sweep	Pinch	Pinch	Pinch	Sweep	Pinch	57%
M10 Full Converge (Slow) Low	Sweep	Sweep	Pinch	Hold	Pinch	Sweep	Pinch	43%
M11 Selective Converge Low	Tap	Sweep	Pinch	Hold	Pinch	Tap	Pinch	43%
M12 Float / Random	Tap	—	Hold	Tap	—	Tap	Tap	86%

Pinch Tap Hold Swipe Sweep Poke

Figure 7.1. Gesture–Motion Mapping Matrix
Gesture selections across seven participants and twelve motion stimuli, with consensus rates shown for each motion.

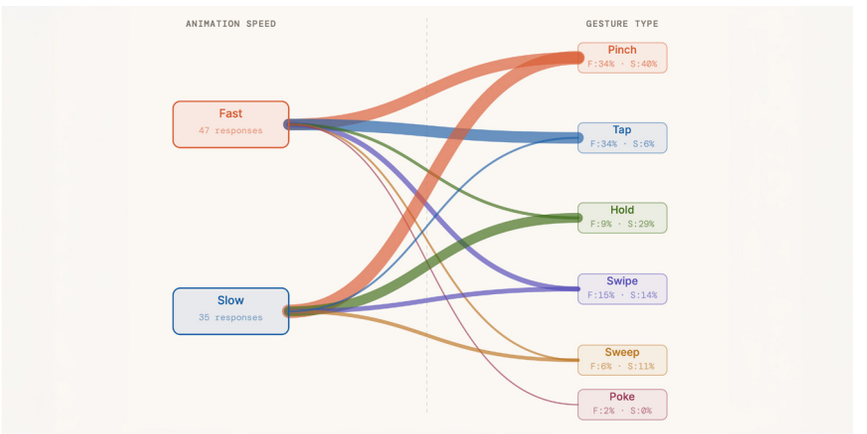


Figure 7.2 Temporal Quality and Gesture Expectation
Fast and slow motion conditions were compared to show how motion timing shifted expected gesture types.

Motion–Meaning Interpretation

Participants did not read the motions only as visual effects; they interpreted them as information behaviors. Expansion was generally read as opening, exploring, zooming into detail, or revealing layers beneath a summary. Bridging was read as connecting, merging, recombining, or discovering relationships. Narrowing was read as summarizing, condensing, cleaning up, closing, or completing a step. Where motion behavior was clear, gesture and meaning tended to align. Where it sent multiple signals, participants diverged. Divergence was most visible in sustained motions, where some participants interpreted gradual movement as continuous control and others as auto-playback requiring a hold.

Material-Mediated Interaction

The textile surface shaped engagement in ways that exceeded conventional touchscreen interaction. Participants described its tactile feedback as helping them feel whether contact had registered. The elasticity and spring of the membrane invited pressing, leaning, resting, and playing with the surface; several participants contrasted the warmth and softness of fabric with the hardness of glass.

The material also introduced limitations. Fingernails and dry skin sometimes created friction, and the softness of the projected image made precise targets harder to identify. Participants perceived the surface's depth dimension — pressing into rather than only onto the surface — although depth-based interaction did not consistently emerge without prompting.

7.3 Cross-Participant Patterns and Low-Consensus Motions

High-consensus motions revealed which mappings appeared relatively intuitive: sudden local appearance often invited precise touch, and sudden global bridging often invited broader movement. The low-consensus motions, however, were more analytically valuable, because disagreement revealed where motion behavior sent conflicting signals. Three recurring sources of disagreement emerged. The first was temporal ambiguity: in sustained motions, some participants read gradual unfolding as continuous interaction, others as an animation to be triggered and held until complete. The second was competing cues between spatial scope and temporal quality: a local sustained bridging motion could suggest a small, precise gesture because the elements were close together, but also a sustained hold because the motion unfolded slowly. The third was spatial scale: for global narrowing motions, participants differed in whether gathering everything called for a finger-scale action such as pinch or an arm-scale action such as sweep — differences suggesting that spatial scope shaped not only gesture size but participants' understanding of what kind of control the system was asking for.

The most fundamental unresolved issue concerned selection. In localized narrowing motions, participants sometimes recognized a two-step structure — select a region, then condense it — but the motion did not always make the selectable target clear. When elements were already moving, participants struggled to determine what could be selected, where the gesture should begin, and which elements their action would affect. This issue is carried forward as a design constraint in Part III.

Category	What It Examined	Coded For
Temporal ambiguity	Sustained motion was read either as continuous control or as an animation triggered by holding.	Slow expansion Slow merge
Competing cues	Spatial closeness suggested a small, precise gesture, while slow timing suggested sustained contact.	Local sustained bridging
Spatial scale	Global narrowing invited either finger-scale pinch or arm-scale sweep.	Global convergence
Design constraint: selection ambiguity	Participants were unsure what to select, where to begin, and which elements would be affected.	Local convergence

Table 7.2. Analytic Coding Framework
Three coding dimensions used to translate raw observations into structured findings.

CHAPTER 8: Research Synthesis

This chapter consolidates the analysis of Chapter 7 into synthesized findings. Its role is transitional: it turns the Motion-Gesture Mapping Study from a set of coded observations into design knowledge for Part III. The findings are not validated gesture vocabulary. They are formative evidence about how motion, materiality, depth, and resistance shaped participants' interpretations on a soft projected surface.

The findings fall into two groups: motion — how the visual behavior of artifacts communicated function, timing, and scale — and material engagement — how the textile surface changed the way participants approached, felt, and imagined interaction with projected content.

8.1 Motion as Interaction Grammar

F1 — Motion operated as interaction grammar.

Participants did not treat visual motion as styling. They treated it as information: about what the system was doing, what kind of content was being represented, and what form of bodily response felt appropriate.

Motion shaped perceived function — Expansion read as exploration or revealing detail, Bridging as connecting or combining, Narrowing as condensing, cleaning, or completing — so motion operated as a functional cue rather than only as visual feedback. Temporal quality suggested different distributions of control: sudden motion was often read as system-triggered or already completed, while sustained motion suggested the participant might still modulate, continue, pause, or reverse the action; speed signaled not only timing but the user's role. Spatial scope shaped gesture scale: global motions tended to invite larger gestures, sometimes with multiple fingers or both hands, and local motions invited smaller, more precise ones — though hand count itself carried no stable symbolic meaning, varying instead with reach, comfort, and scale.

Design requirement carried forward: the designed system should make motion legible as interaction grammar, and should avoid relying on motion alone where timing or system agency could be ambiguous.

8.2 Material as Thinking Surface

Three material-engagement findings summarize how the surface shaped behavior and interpretation.

F2 — Soft materiality shifted bodily engagement.

Participants did not describe the textile interface only as a screen. Through tactile feedback, elasticity, and pressure, the projected graphics began to feel like spatial things: content that could be moved, grouped, returned to, and physically worked. Instead of only tapping or pointing, participants pressed, leaned, rested their hands, and played with the surface. The material invited a more active kinetic relationship to the interface and expanded what participants imagined the system could sense and support.

F3 — Depth suggested a dimension of intent.

The textile introduced a Z-axis absent from conventional flat touch. Moving across the surface suggested exploration, navigation, and arrangement; pressing into it suggested selection, commitment, or stronger intent. Depth therefore emerged as a promising channel for distinguishing casual movement from deliberate action. Because depth-based gestures did not consistently emerge without prompting, however, the channel is treated as a design opportunity rather than a validated interaction convention.

F4 — Material resistance provided a cognitive foothold.

Participants used the fabric's elasticity, pushback, and tactile feedback to anchor their actions. The surface made interaction feel more consequential: touch could be confirmed, pressure could be felt, and digital information could be imagined as having weight or solidity. Material resistance did not simply make the interface feel different; it helped projected content feel like something the participant could hold onto, work through, and return to during thinking.

8.3 Carrying the Findings Forward

The findings are carried into Part III as interpreted empirical insight, not as a finished specification. Where a designed interaction matches a condition the study examined directly, the observed mapping informs the design; where the design moves beyond the studied conditions, the findings are used more cautiously, as reasoning tools rather than prescriptions.

The low-consensus motions serve an equally important role. They identify interaction problems the design must resolve rather than inherit: temporal ambiguity, competing spatial and temporal cues, and the difficulty of selecting or shaping moving artifacts. These become design constraints in the following chapters.

Read against the conceptual framework of Chapter 4, the findings support the premise that physical engagement with materialized content can participate in sensemaking. Participants' gestures were not merely commands issued after thought; in many cases, pressing, sustaining contact, scaling the hand to the surface, and feeling material feedback were part of how participants interpreted what the system was doing.

At this scale the findings remain supportive rather than conclusive. The following chapter translates this formative grounding into the design principles that organize the proposed system.

PART III
THE DESIGN

CHAPTER 09-10

CHAPTER 9: Design Development

This chapter begins Part III by translating the thesis argument into design. The previous chapters established why conversational AI may need a different interface for sustained co-ideation: one that supports spatial organization, bodily engagement, material interaction, and authorial agency. Chapter 9 turns those commitments into design principles, clarifies the relationship between TMS and FIELD, and introduces the Network Graph structure through which FIELD materializes human-AI dialogue.

The chapter serves as a hinge between research and demonstration. Chapter 8 provided formative findings about motion, materiality, depth, and resistance. Chapter 9 asks how those findings should shape the system. Chapter 10 then shows how the principles become specific FIELD features.

9.1 Design Principles

The design of TMS is organized around three principles. These principles are not feature descriptions. They are commitments that guide how the system should behave, what kinds of interaction it should support, and how the user's role in AI-assisted co-ideation should be protected.

Principle	Commitment	System Requirement	User Value
P1. Materialize the Dialogue	AI dialogue should become spatially persistent and handleable.	Represent AI-generated ideas as nodes, cards, and visible relationships.	Supports return, comparison, and active manipulation.
P2. Co-locate Sensemaking and Steering	Reading, interpreting, and directing should happen in the same spatial field.	Combine voice-based semantic input with gesture-based spatial reference.	Reduces the split between thinking about content and acting on it.
P3. Enable Authorial Agency	The interface should keep the user responsible for direction, relation, and meaning.	Support framing, expansion, synthesis, pruning, and reframing as user-initiated actions.	Keeps the user actively shaping the process rather than only selecting outputs.

Table 9.1
Design principles translated from the thesis framework and research findings.

The principles intentionally move from representation to action to agency. First, AI dialogue must be materialized so that ideas can persist in space. Second, sensemaking and steering must be co-located so that the user can act where they are thinking. Third, the interface must preserve authorial agency by making framing, connection, subtraction, and reframing available as user-initiated actions.

Each principle redeems a specific line of evidence from the preceding chapters. The first, materialization, rests on the premises of Chapters 3 and 4 and on F1 and F4: participants treated projected content as functionally

meaningful when its motion was legible, and as physically available when the surface offered resistance. The second, co-location of sensemaking and steering, follows from the co-location and cognitive-load rationale of Chapter 3 and from F2: the soft surface drew participants into acting where they were looking, rather than dividing attention between display and control. The third, authorial agency, responds to the temporal-quality pattern within F1 — sudden motion read as system-triggered, sustained motion as user-controllable — which showed that interaction design itself distributes control between user and system, and can therefore be designed to keep that distribution in the user's favor.

9.2 System Overview: TMS and FIELD

This thesis uses two names deliberately. TMS, or Tangible Multimodal System, refers to the whole designed system: the hardware concept, the software application, the interaction model, the gesture logic, and the design argument that connects them. FIELD is the software application developed within TMS. It functions as the design demonstration of the software and interaction logic of TMS: the named interface through which users engage in AI-assisted co-ideation.

TMS is conceived as a unified hardware-software proposal for transforming human-AI dialogue from a linear exchange into an embodied, multimodal thinking medium. Its target architecture consists of three connected parts: a responsive display, a Network Graph canvas, and a multimodal interaction model.

At the hardware level, TMS is designed for a soft responsive display. The intended display is not a conventional touchscreen but a surface capable of responding to touch, pressure, depth, and deformation. This material level matters because the thesis argument depends on more than visual organization. It depends on the possibility that digital artifacts can become available for bodily action.

At the software level, FIELD organizes the dialogue on a Network Graph canvas. The canvas represents AI-assisted ideation as a field of ideas and relationships rather than as a vertical transcript. Ideas become nodes. Relationships become connection lines. The user can move across the canvas, open ideas, inspect relationships, generate new directions, and evaluate what has not yet been developed.

At the interaction level, voice and gesture work together. Voice carries the semantic content of the interaction: the user's goals, questions, constraints, doubts, and reframings. Gesture carries the situated direction of the interaction: where the user is acting, what they are selecting, how they are connecting, and what they are releasing. The user's point of contact on the canvas becomes part of the context for the AI response.

Figure 9.1
TMS as a Soft Responsive Display System
The Tangible Multimodal System is shown as a unified hardware–software proposal, with the FIELD interface presented on a soft responsive display.

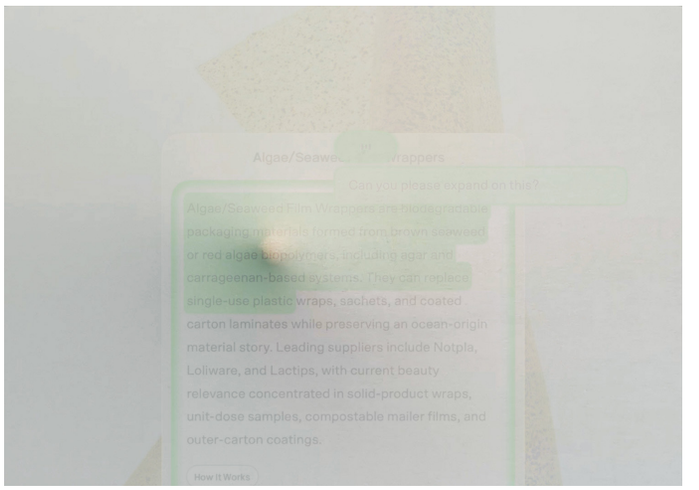
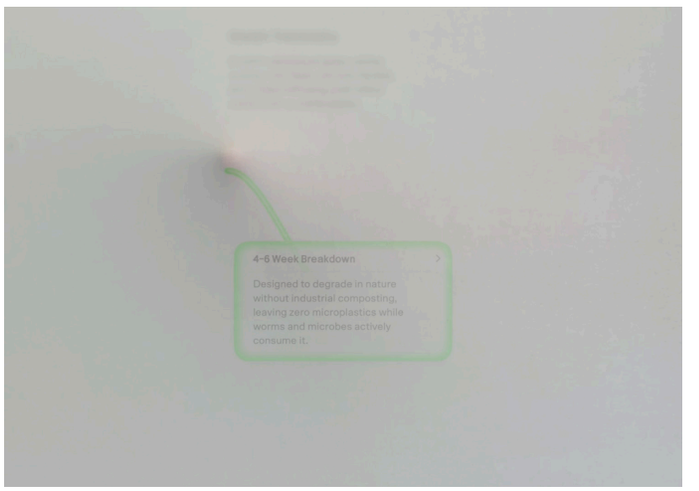


Figure 9.2
FIELD Interface on the Responsive Surface
A close-up view of the FIELD interface, where AI-generated content is presented as spatially situated material for touch, inspection, and further development.



9.3 Network Graph Canvas

The FIELD canvas is organized as a network graph: ideas are represented as nodes, and the relationships between them as visible connection lines. The graph is not a literal model of the brain, nor a literal visualization of a generative model's latent space. It is an interaction structure — a way of making the evolving relationships among ideas visible, addressable, and open to action.

The canvas is composed of four core elements: node, connection line, network graph, and NORTH STAR. Together, they define how AI dialogue becomes spatially organized and available for user intervention.

Node

A node is a compressed idea artifact: a single unit of AI dialogue that the user can open, move, branch, connect, or release. It allows generated content to persist as something the user can return to and manipulate, rather than as text that disappears into the scroll.

Connection line

A connection line is a visible relationship between two nodes. Because the relation is drawn rather than implied, it can be touched, inspected, and questioned in its own right. It turns conceptual relationships into addressable parts of the interface.

Network Graph

The network graph is the full node-and-edge canvas. It represents the evolving field of co-ideation: the space where the user compares ideas, returns to earlier directions, connects fragments, and reframes the conversation. It makes visible which ideas are connected, which areas have grown dense, and which directions remain underdeveloped.

NORTH STAR

NORTH STAR is the central framing node of the canvas. It holds the session's goal, question, or intention, and can be revised at any point to redirect the field. Rather than treating framing as something fixed at the beginning of a chat, FIELD keeps the frame present and editable throughout the process.

This organization responds to a problem described earlier in the thesis. In a linear chat interface, ideas appear in the order they were generated, and their relationships must be remembered, inferred, or re-explained through text. In FIELD, relationships are placed on the canvas. The scroll keeps these relations implicit; the graph makes them explicit.

The canvas also responds directly to the selection problem identified in Section 7.3. In the study, participants struggled to act on targets that were already in motion: when elements moved, it was unclear what could be selected, where a gesture should begin, and which elements an action would affect. FIELD therefore keeps nodes spatially persistent and stationary by default. Motion is reserved for user-initiated transitions, so that the target remains stable before any gesture begins.

Selection precedes motion rather than competing with it — a direct design response to a central constraint identified through the low-consensus findings.

The network graph is therefore not only a visual style. It is the structural mechanism through which FIELD makes AI dialogue returnable, comparable, and open to intervention. It is the ground on which the features described in Chapter 10 operate.

9.4 Canvas Modes and Views

FIELD uses two canvas modes and one focused view to manage how much information is shown and what kind of thinking the user is doing at a given moment. SURFACE and BLINDSPOT are modes of the whole network graph; DIVE is a focused view opened from a single node.

SURFACE

SURFACE is the default mode of the canvas. It supports the exploratory development of the visible field — generating, connecting, and arranging ideas — and shows the user what they are currently building. In this mode, the graph functions as an active workspace for expanding and organizing AI-assisted co-ideation.

BLINDSPOT

BLINDSPOT is the evaluative counterpart to SURFACE. Rather than displaying a different field, it reads the same network through the lens of NORTH STAR, surfacing what may be missing, underdeveloped, or disconnected from the session's intention. Its purpose is not simply to extend the field, but to help the user critically reframe it.

SURFACE and BLINDSPOT share an essential property: they preserve the same node layout. What changes between them is not where things are, but the lens through which the network graph is read. This allows the spatial memory built during exploration to remain intact during evaluation.

DIVE

DIVE is different in kind. It is not a mode of the whole graph, but a focused view opened from a single node. It unfolds that node into detail while keeping the main canvas visually minimal in the background. DIVE gives depth without losing the field: the user can examine one idea closely without surrendering their place in the larger structure.

Together, these three views allow FIELD to support co-ideation at three scales: the whole field in SURFACE, the overlooked field in BLINDSPOT, and the focused idea in DIVE. The system can move between these scales without requiring the user to lose track of where their attention rests within the evolving structure of the session.

CHAPTER 10: FIELD Feature System

Having established FIELD as the software demonstration of TMS and the Network Graph as its organizing structure, this chapter turns from structure to action. Each feature is framed as an authorial action: something the user does to shape the direction, relation, depth, or boundary of an AI-assisted thinking process.

The point of the feature system is not to maximize the number of controls available to the user. It is to distribute agency across the co-ideation process. FIELD gives the user recurring opportunities to frame, deepen, expand, read relationships, connect, refine, and reframe. These actions correspond to the Think-Through Framework introduced in Chapter 4 and to the motion and material-engagement findings synthesized in Chapter 8. The FIELD feature system is organized around authorial actions rather than commands. Each feature gives the user a way to frame, deepen, expand, question, synthesize, remove, or reframe the evolving ideation field.

Feature	Authorial Action	Primary Interaction	System Response	Purpose in Co-ideation
NORTH STAR	Frame	Tap the central NORTH STAR node	Opens the session frame for editing	Keeps the user's goal, constraints, and criteria available as the center of the ideation field
DIVE	Deepen	Tapa a node	Opens a focused detail view	Allows the user to examine the fuller content inside a compressed idea artifact
BRANCH	Expand	Press, hold, and pull outward from a node	Generates related directions from the selected idea	Supports divergent exploration b letting the user expand from a chosen point
EDGE	Interpret	Touch a connection line	Surfaces an insight about the relationship between two nodes	Helps the user understand how two ideas relate
BRIDGE	Synthesize	Press and drag one node toward another	Connects two previously unconnected nodes to see what emerges	Expands the idea space by exploring unexpected connections
RELEASE	Remove	Rub a node with two fingers	Deletes the selected node	Lets the user actively shape the field by removing irrelevant ideas
BLINDSPOT	Reframe	Switch from Surface to BLINDSPOT mode	Reveals underdeveloped nodes, weak connections, and North Star gaps	Gives the user a critical overview of the field to strengthen their line of reasoning

Table 10.1. FIELD Feature System as Authorial Actions

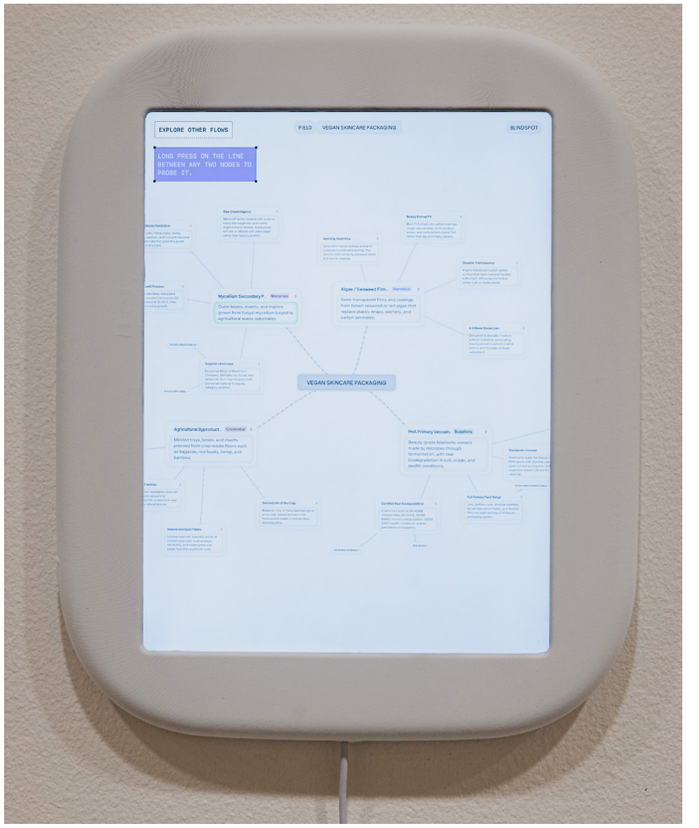


Figure 10.1. FIELD network graph canvas on the iPad-optimized prototype. Ideas are displayed as nodes and connection lines across the canvas, showing how AI dialogue is materialized as a spatial ideation field.

In the Henry Art Gallery installation, visitors were able to explore the FIELD interface through direct touch gestures. Because ambient gallery noise made live voice input unreliable, the voice layer was simulated through pre-scripted internal input. The installation therefore foregrounded the gesture-based interaction model while maintaining the intended relationship between voice, touch, and canvas response.

Figure 10.2. FIELD prototype installed for public interaction at the Henry Art Gallery.

10.2 NORTH STAR — Framing

NORTH STAR anchors the user's initial intention as the organizing center of the session. In a linear chat, the original prompt often disappears into the history of the conversation, even though it continues to shape what counts as relevant. FIELD makes that frame spatially persistent by placing it at the center of the network graph, where it remains visible as the field grows around it.

The interaction begins with voice. The user speaks the topic, goal, or concern they want to explore. FIELD then translates this input into a set of framing statements, allowing the user to review how the system has understood their intent before the ideation space expands. The user can swipe through these statements with a finger, reading them as editable pieces of the session frame rather than as a fixed prompt.

Because spoken intent is often incomplete, NORTH STAR is designed to remain revisable. If the user notices something missing, unclear, or misread, they can press and hold the relevant statement or position. A local voice input box appears directly above the finger, allowing the user to add, correct, or clarify the frame in place. Once revised, the framing statements are consolidated into the central NORTH STAR node, which becomes the organizing anchor for the network graph.

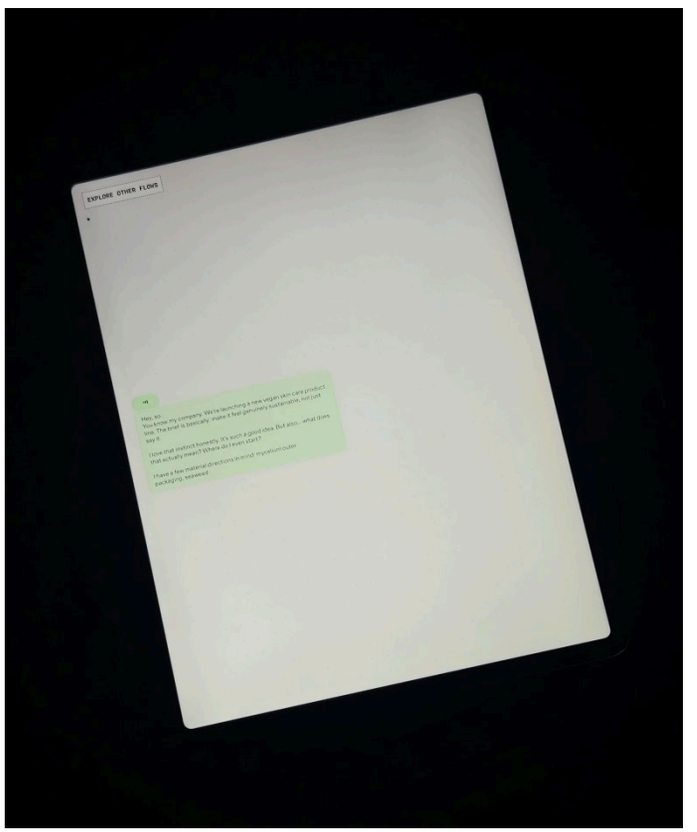
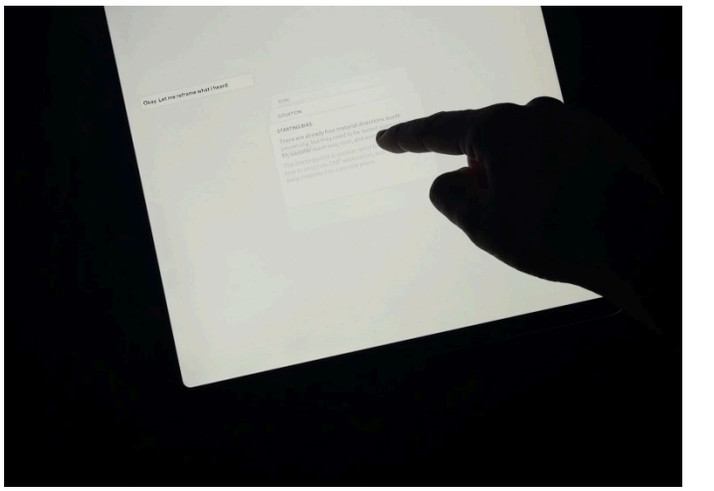
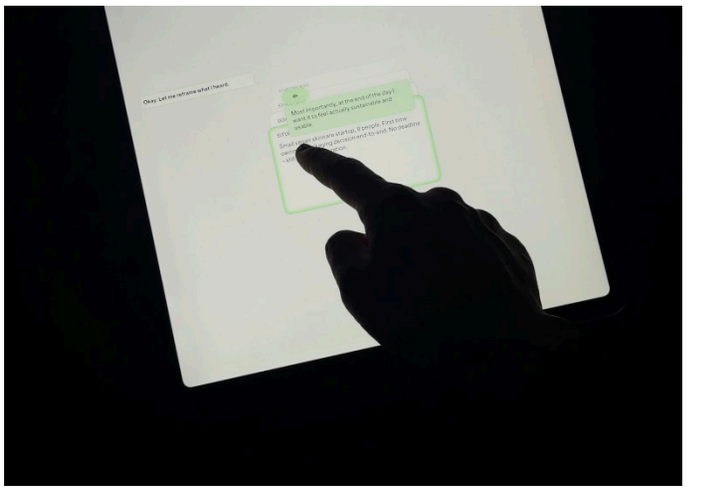


Figure 10.3. NORTH STAR framing workflow.
1. Speak ideation topic

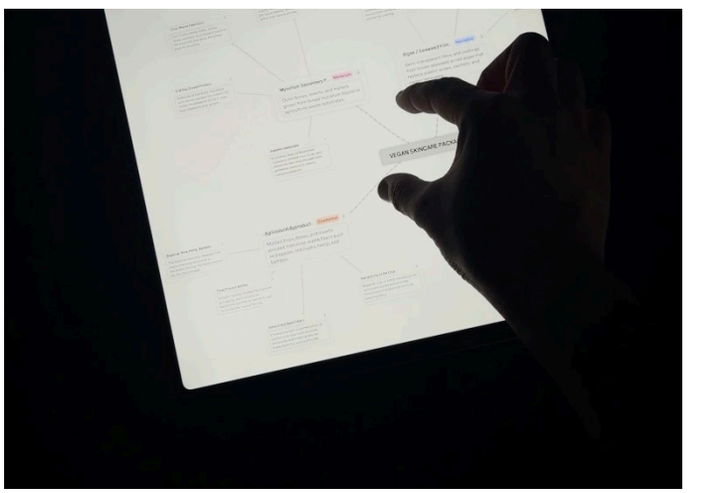
2. Review framing intent



3. Revised in place



4. Generate network graph



The sequence shows how the user begins a FIELD session through voice input, reviews and revises the generated framing intent, and then generates the NORTH STAR node as the center of the network graph. Once the canvas is created, the user begins co-ideation with AI.

As an authorial action, NORTH STAR is not only setup. It is a way of keeping the session accountable to the user's purpose. The user can return to the frame, revise it, or use it as the lens through which BLINDSPOT later identifies what may be missing. The design decision here is deliberately conservative: rather than letting the AI infer intent from the accumulating conversation, FIELD holds the user's stated intention in a fixed, returnable place.

10.3 DIVE — Deepening

DIVE allows the user to open a compressed node and examine its content in greater detail. On the main canvas, nodes remain visually minimal so that the user can read the overall structure of the ideation field without being overwhelmed by text. Tapping a node opens DIVE, where the fuller content of that idea becomes visible for closer reading and further interaction.
DIVE is designed not only for inspection, but also for revision. If the user notices a phrase, sentence, or section that needs clarification, expansion, or correction, they can select the relevant area by rubbing across it with a finger. Once the area is selected, the user can press and hold to invoke a local voice input box, using the same interaction pattern introduced in NORTH STAR. The user then speaks the revision, addition, or question, and the selected content is updated in place.

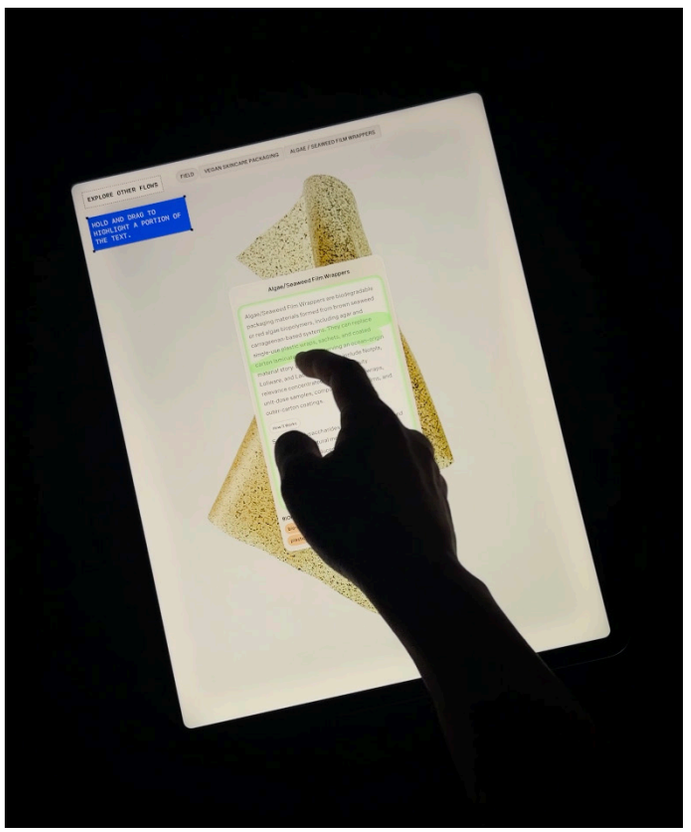
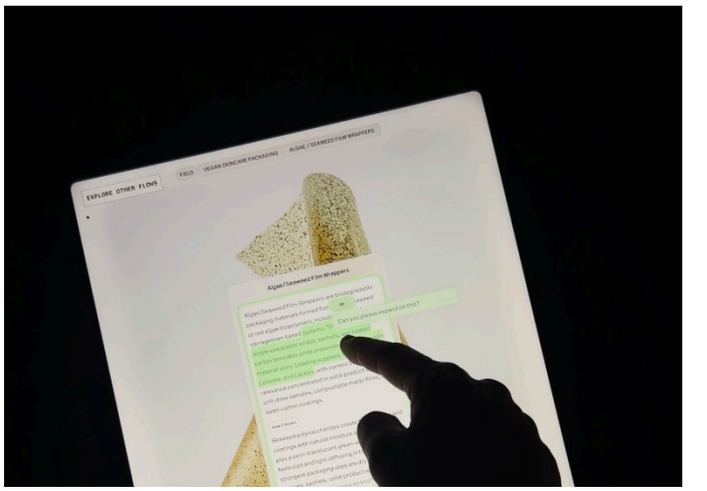


Figure 10.4. DIVE: tapping a node opens focused detail.
1. Open nod in Dive
2. Select text by rubbing

3.Press and hold to revise
4. Update content in place

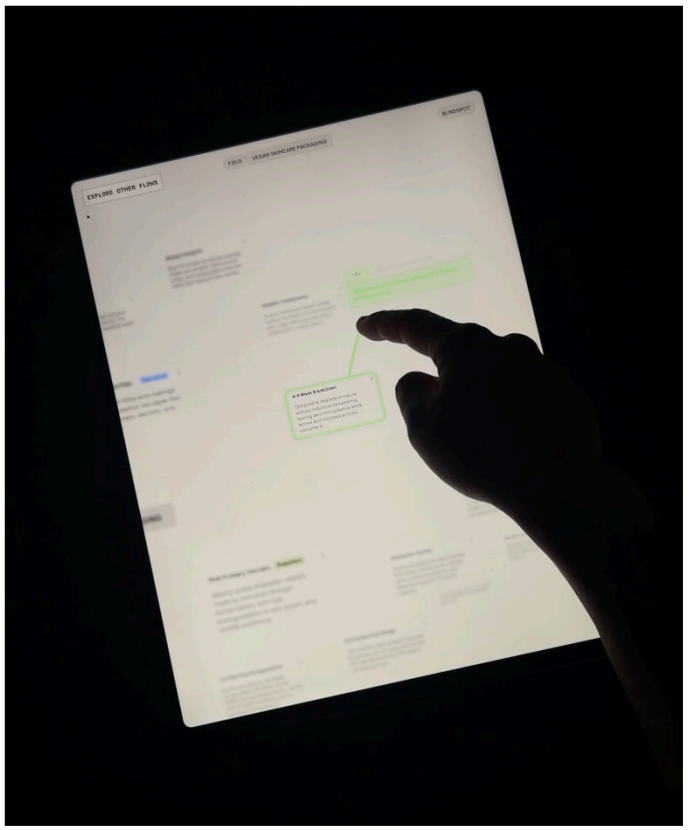


The sequence shows how a node is opened into DIVE for detailed reading, how a sentence or area is selected by rubbing across it, and how the user invokes a local voice input box to revise the selected content in place.
As an authorial action, DIVE supports depth without abandoning the larger field. It allows the user to move from overview to detail, intervene at the level of a single idea, and then return to the canvas with the revised node still connected to the wider ideation structure.

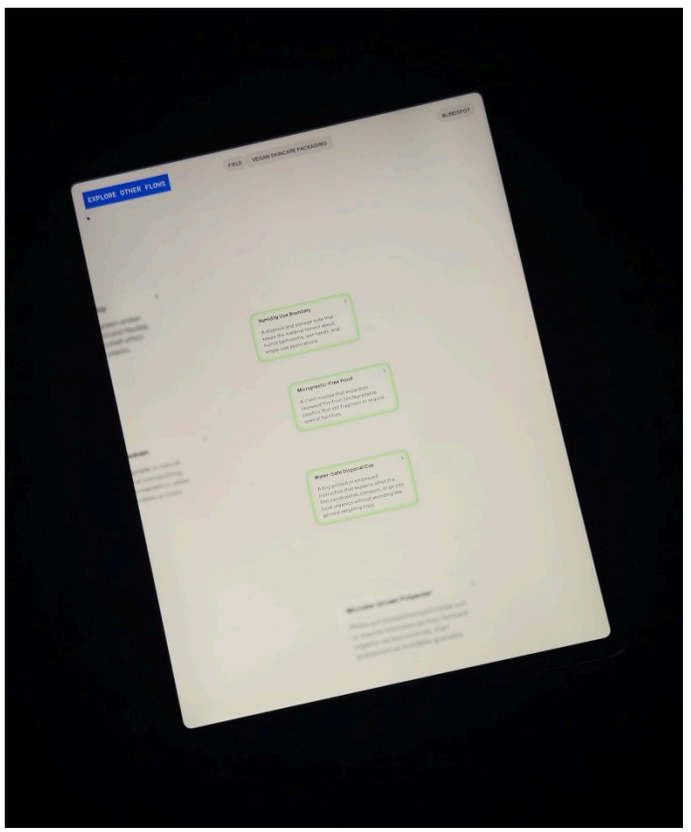
10.4 BRANCH — Expanding

BRANCH expands one idea into related possibilities without losing its origin. The user presses, holds, and pulls outward from a node, and new candidate directions grow from that point while the source node stays in place. This makes divergence spatially situated: expansion begins from a chosen idea rather than from the bottom of a transcript.
The design carries forward the study's Expansion theme — which participants consistently read as opening and exploring — at the level of meaning, while the press-and-pull gesture itself draws on the outward-scattering variants catalogued in the Phase 2 gesture matrix. The two were kept distinct on purpose: the theme settles what the action means, while the matrix settles how the hand performs it.
During this interaction, the user can also speak how they want the idea to expand. Voice provides the semantic direction of the expansion, while the gesture specifies where that expansion should begin. In this way, BRANCH combines embodied selection with verbal steering.
The authorial function of BRANCH is to make expansion directional. The AI proposes related possibilities, but the user decides where divergence should begin and how far it should extend. Generation is initiated by a held, deliberate gesture rather than delivered as an unprompted feed — a small but consequential difference in who is leading.

Figure 10.5. BRANCH: pressing, holding, and pulling from a node grows related directions.
1. Hold node
2. Pull outward

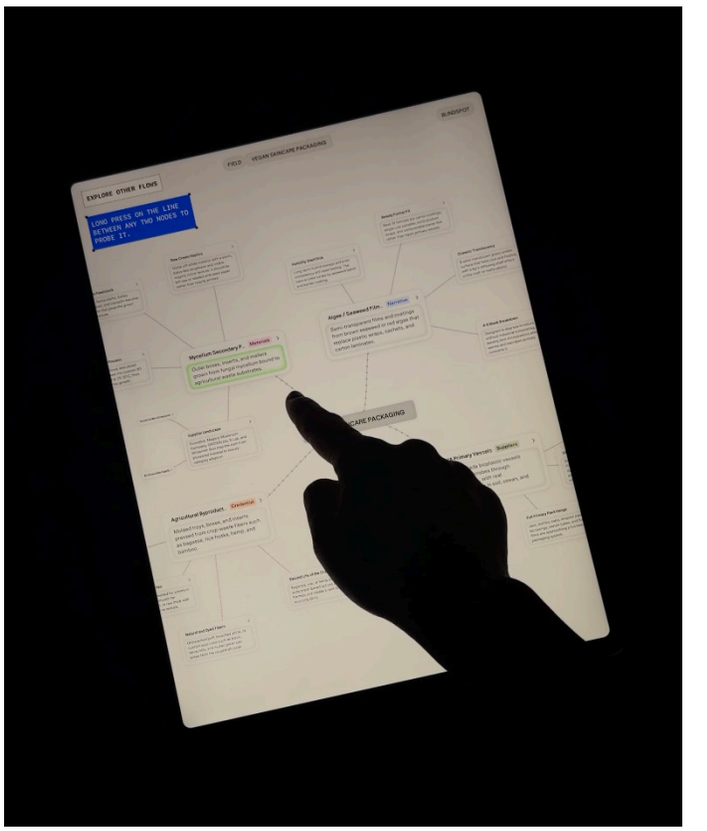


3. Speak expansion direction
4. Generate branches



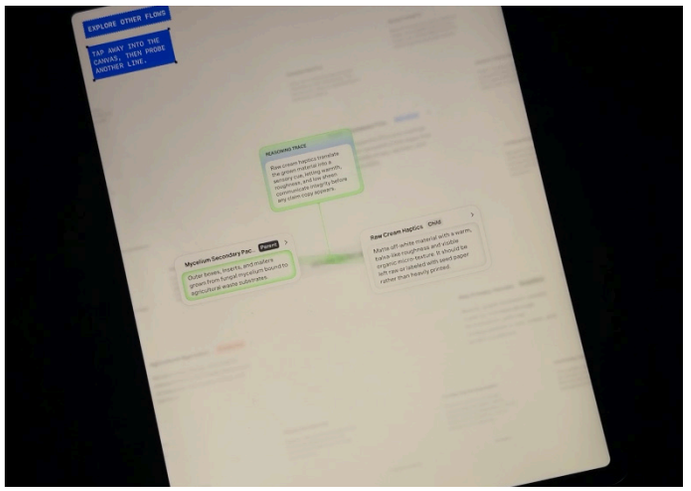
EDGE lets the user touch a connection line and read what lies between two nodes. In a conventional chat interface, the relationships among ideas remain implicit: they must be remembered, inferred, or asked about through text. FIELD makes relationships visible and addressable as first-class elements of the canvas. Making connection lines directly touchable applies the co-location principle introduced in Chapter 3 — the relationship is interpreted at the place where it is shown, without translation into a separate textual query.

Figure 10.6. EDGE: touching a connection line interprets the reasoning between two nodes and what would strengthen it.
1. Touch a connection line



As an authorial action, EDGE is interpretive rather than interrogative. Touching an edge does not simply return a definition of the link; it surfaces the reasoning that joined the two ideas — why the system, or the user's own earlier move, placed them in relation — and what might strengthen that relationship: a missing premise, an untested assumption, a connection that is suggestive but not yet substantiated. EDGE is a reasoning-surfacing tool: it makes the basis of a connection inspectable and shows the user how to make it more deliberate. Where BLINDSPOT reframes the whole field by revealing what is missing, EDGE works locally and relationally, deepening one connection at a time. It supports relational sensemaking — letting the user ask not only whether two ideas belong together, but how their pairing might be made better grounded and more clearly their own.

2. Reveal the relationship



As an authorial action, EDGE is interpretive rather than interrogative. Touching an edge does not simply return a definition of the link; it surfaces the reasoning that joined the two ideas — why the system, or the user's own earlier move, placed them in relation — and what might strengthen that relationship: a missing premise, an untested assumption, a connection that is suggestive but not yet substantiated. EDGE is a reasoning-surfacing tool: it makes the basis of a connection inspectable and shows the user how to make it more deliberate. Where BLINDSPOT reframes the whole field by revealing what is missing, EDGE works locally and relationally, deepening one connection at a time. It supports relational sensemaking — letting the user ask not only whether two ideas belong together, but how their pairing might be made better grounded and more clearly their own.

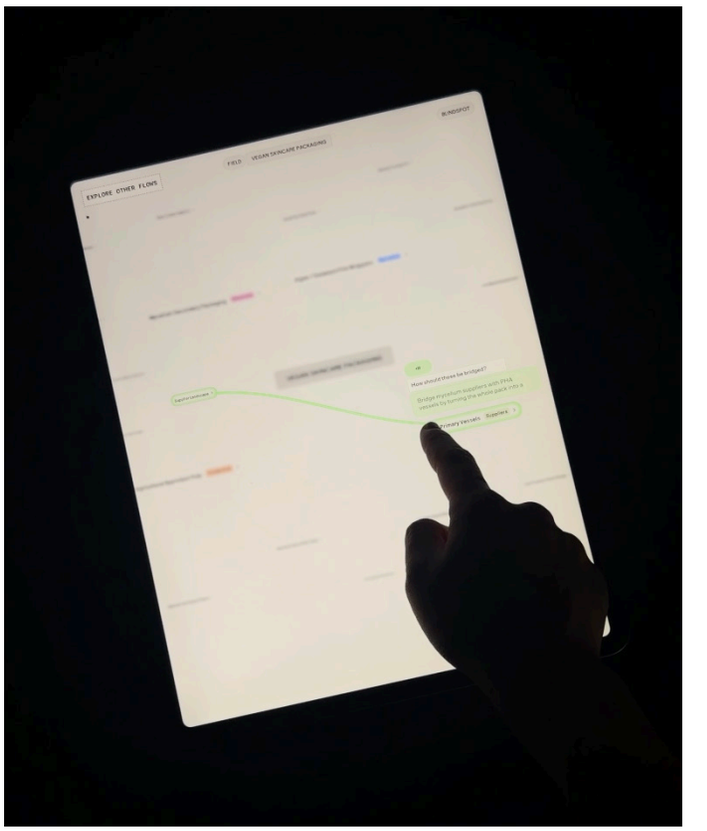
10.6 BRIDGE — Connecting

BRIDGE tests whether two ideas can produce something together. The user presses one node and drags it toward another, turning synthesis into a deliberate, visible act. The system may attempt an experimental synthesis, but the relation begins with the user's gesture. The drag-together mapping follows the study's Bridging theme, the most consistently interpreted of the three: participants read convergent motion as connecting or merging, and proposed convergent gestures to drive it — so the gesture FIELD assigns to synthesis is the one users already associated with bringing things together.

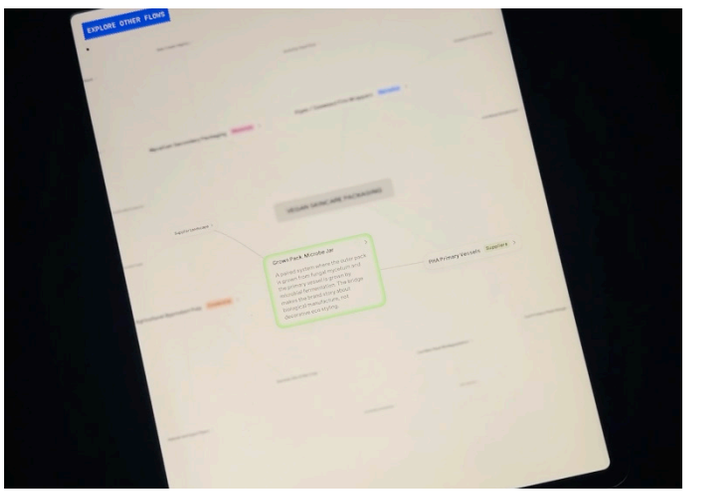
This feature matters because synthesis is often where authorial agency becomes vulnerable in AI-assisted work. A generative system can combine ideas fluently, but fluency does not guarantee relevance, and a system that synthesizes on its own tends to narrow toward the combinations it already favors. BRIDGE makes synthesis something the user initiates, observes, judges, and either keeps or releases. By requiring the user to choose which two nodes collide — including distant or unlikely pairings the system would not have proposed — it also works against the convergence that can settle over a long session. The user does not merely receive a combined output; they author the conditions under which the collision occurs.

The sequence shows how the user drags one unconnected node toward another, speaks the intended direction of synthesis, and receives a new idea node generated from the connection.

Figure 10.7. BRIDGE: dragging one node toward another attempts a synthesis.
1. Drag one node toward another
2. Speak the sunthesis direction



3. Generate a new idea node



RELEASE removes what no longer belongs. The user rubs a node with two fingers, and the selected node fades from the active canvas. In a system designed for abundant generation, subtraction is not a secondary housekeeping action; it is part of how the field is shaped. The two-finger rubbing gesture derives from the press-and-rub family identified in the Phase 2 gesture matrix and from the study's reading of the Narrowing theme as cleaning up. Performing removal as a felt, effortful physical act — friction sustained until the node releases — draws on F4, in which material resistance gave an action a sense of consequence: deletion should feel like a decision, not a reflex.

The authorial function of RELEASE is boundary-making. Co-ideation requires deciding what to carry forward and what to leave behind, and in an interface that can generate endlessly, that deciding is easily lost. By making removal a deliberate physical action, FIELD gives pruning a visible and embodied place in the thinking process — the negative space of authorship, as much a part of shaping the field as generation is.

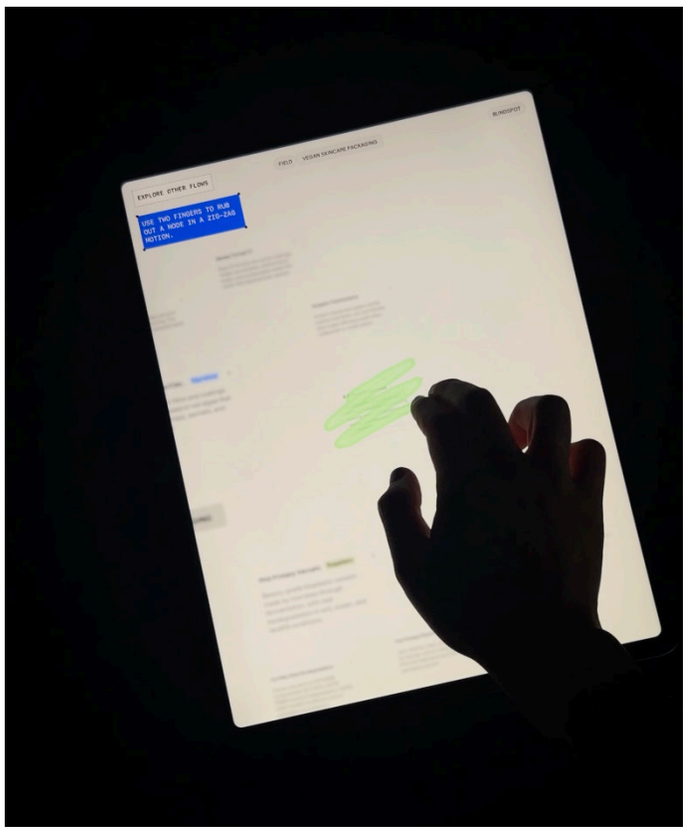


Figure 10.8. BRIDGE: dragging one node toward another attempts a synthesis.
1. Rub slected node with two fingers
2.Node fades
3.Node is removed

BLINDSPOT reframes the whole field by revealing what has been underdeveloped, over-assumed, or left outside the current direction. Unlike DIVE, which focuses on a single node, BLINDSPOT acts on the same canvas as a whole: the layout remains familiar, but the interpretive lens changes. Preserving the node layout across the mode shift applies the spatial-memory rationale of Chapter 3 — reframing changes the lens, not the map, so the user's accumulated sense of place survives the change of view, and evaluation does not mean starting over in an unfamiliar arrangement.

This feature supports metacognitive monitoring, and it does so in a specific way. Rather than asking the AI to evaluate the conversation from nowhere — a stance that tends to produce agreeable, generic assessment — the user activates a reframing mode in relation to the NORTH STAR and the existing field. Evaluation becomes a user-initiated act anchored to the user's own stated intent: the person asks the system to reveal what the current structure may be hiding relative to where they said they were going, then decides whether and how to redirect the session. It is the design's answer to a tendency identified earlier in the thesis — that human and machine, left to reinforce each other, can narrow together without noticing what has fallen outside the frame.

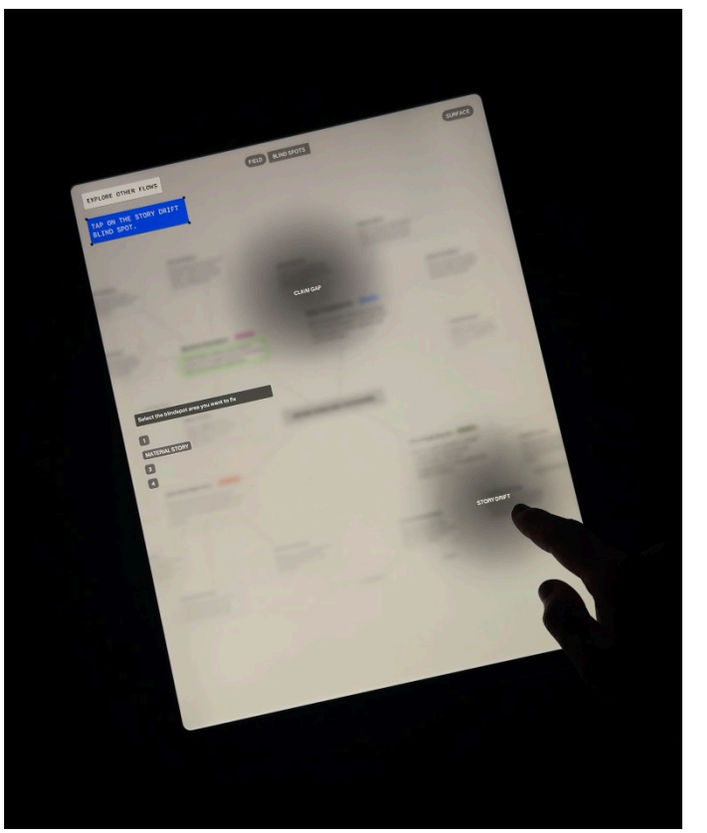
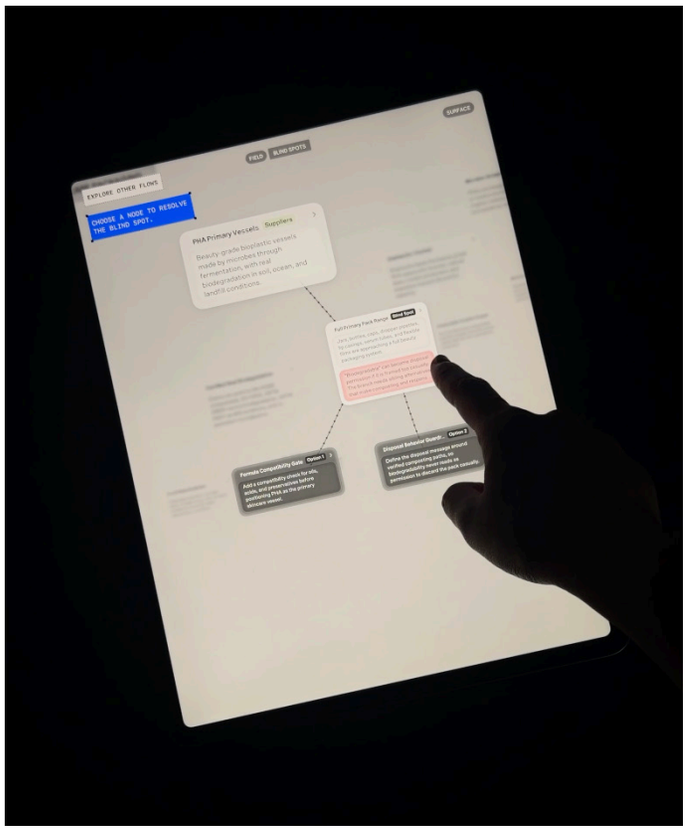


Figure 10.9. BLINDSPOT: reframing the same network graph through what has been overlooked.
1. Activate BLINDSPOT
2. Reveal weak areas

3. Review AI suggestions
4. Strengthen the field



10.9 The Feature System as an Authorial Loop

Taken together, FIELD's features form an authorial loop rather than a menu of isolated tools. NORTH STAR frames the field; BRANCH expands it; DIVE deepens it; EDGE interprets its relationships; BRIDGE tests synthesis; RELEASE prunes what no longer belongs; and BLINDSPOT reframes the whole. The order is not fixed — a session moves among these actions as the thinking requires — but the set is designed to keep the user present at every scale of the work: framing the field, growing it, reading within it, combining across it, cutting from it, and stepping back to see it whole. This is the concrete form of the authorship the thesis set out to restore. In a chat interface, the human is most active at the two ends of generation — writing the prompt and judging the result — and comparatively passive in between. FIELD's loop is designed so that the user remains an author throughout the movement of the session: not only at the beginning and end of AI generation, but in the continuous handling of the field as it forms.

Part IV.
Discussion and Conclusion

CHAPTER 11-12

CHAPTER 11: Contributions, Limitations, and Future Work

This thesis set out to ask how interaction and interface might evolve when conversational AI is used not only to produce answers, but to participate in the formation of ideas. The answer developed across the thesis is neither a claim that text chat should disappear nor a claim that touch alone solves co-ideation. The contribution is more specific: this thesis proposes and demonstrates an interaction structure in which AI dialogue becomes spatially persistent, handleable, and open to authorial intervention.

11.1 Contributions

The thesis contributes to human-AI interaction design in three related ways: a conceptual framework, a design demonstration, and formative research insights. Each corresponds to a different part of the thesis arc: theory, design, and empirical grounding.

Conceptual Framework — Materializing AI Dialogue

The first contribution is the proposition that AI-generated dialogue can be treated not only as text to be read, but as materialized dialogue that can be spatially organized, returned to, and handled. The Think-Through Framework — Externalize, Materialize, Handle, Think-Through — operationalizes this shift by describing how human intent and AI output may form a continuous loop of framing, manipulation, interpretation, and revision. The core argument is not that touch simply adds another input method, but that spatial and bodily action can participate in reasoning during human-AI co-ideation.

Design — FIELD as a Demonstration of Spatial, Gesture-Voice Co-Ideation

The second contribution is the realization of this theory through FIELD. FIELD translates the conceptual framework into a Network Graph canvas where ideas and relationships become nodes and connection lines that can be addressed through voice and gesture. Rather than proving a complete product model, FIELD demonstrates how co-ideation — framing, expansion, deepening, synthesis, refinement, and reframing — can be structured spatially rather than only temporally.

The feature system is central to this contribution: NORTH STAR, DIVE, BRANCH, EDGE, BRIDGE, RELEASE, and BLINDSPOT distribute authorial agency across the process of co-ideation, allowing the user to frame, deepen, expand, inspect, connect, subtract, and reframe ideas before the final moment of editing.

Research: Formative Insights on Motion, Gesture, and Materiality

The third contribution is a set of formative research insights from the Motion-Gesture Mapping Study. The findings suggest that visual motion can operate as an interaction grammar, that soft textile materiality can shift bodily engagement, and that resistance and depth may provide footholds for working with projected content.

These insights are exploratory rather than generalizable. Their value lies in informing the design logic of TMS and identifying problems future tangible multimodal AI interfaces will need to address: temporal ambiguity, competing spatial and temporal cues, and the challenge of selecting or shaping moving artifacts.

Returned to the research questions, the contributions read as follows. With respect to RQ1, the thesis offers a conceptual and demonstrated answer: materializing dialogue as handleable artifacts supports authorial agency by making framing, return, comparison, subtraction, and reframing available as user-initiated spatial actions — an answer whose empirical confirmation is assigned to the comparative evaluation described in 11.3. With respect to RQ2, the formative findings indicate that gesture and voice operate as complementary rather than interchangeable channels — gesture anchoring intent in place while voice carries semantics — and that motion, depth, and resistance shaped how participants understood their own role in the interaction. Both answers are bounded by the formative scope stated in 11.2.

Across these contributions, the iterative design process clarified an important thesis-level lesson: authorial agency is not preserved by giving users more options alone, but by designing ways for them to stay oriented, intervene, and redirect the process as ideas evolve.

11.2 Limitations

Exploratory Sample

The empirical study involved seven design-trained participants. While this sample was appropriate for exploratory gesture elicitation and formative design research, the findings are indicative rather than prescriptive. They should be read as a grounded baseline for design development, not as broadly generalizable evidence.

Hardware Proxy

The Think Through Touch conceptual model relies on material resistance, depth, and elasticity. Because FIELD was prototyped on an iPad Pro as a technical proxy, the system demonstrated spatial and gestural logic under approximated material conditions, but could not yet evaluate the cognitive impact of a fully responsive tactile surface.

Task Realism

The prototype demonstrated interaction flows using bounded stimuli rather than a live LLM backend. Consequently, the cumulative cognitive dynamics of a sustained, open-ended AI co-ideation session have not yet been fully tested.

11.3 Future Work

Addressing these limitations defines the immediate trajectory for this research, moving it from formative design toward more robust empirical evaluation.

Real-Time AI Integration

The immediate technical next step is connecting FIELD to a live generative API. This requires developing a robust state model capable of dynamically tracking the interplay among the user's NORTH STAR framing, spatial canvas history, selected nodes, connection lines, and real-time AI outputs.

Responsive Hardware Development

A second direction is the development of the soft responsive display assumed by the TMS design argument. Future prototypes should test how pressure, depth, elasticity, and resistance change the user's ability to anchor, question, and redirect AI-generated artifacts.

Comparative Empirical Evaluation

Future studies should evaluate FIELD directly against dominant linear-scroll AI workflows. Moving beyond reductive metrics such as speed or efficiency, this evaluation should examine the qualities prioritized by the framework: cognitive continuity, relational sensemaking, spatial return, and the user's felt sense of authorial agency.

In this form, the future work is not separate from the thesis contribution. It extends the same question at larger scale: how might interfaces for generative AI make space for human orientation, judgment, and authorship throughout the process of thinking?

CHAPTER 12: Conclusion

This thesis began with a person at a blinking cursor: an hour into a session, dozens of ideas deep, scrolling to find what had emerged, retyping context the interface had let slip away. The scene was ordinary, and that was the point. The difficulty was not the failure of a single product. It was a property of a default interface: one that organizes thought by time and receives intent primarily through typed language. Against that default, this thesis has argued for a spatial, embodied, and authorable way of thinking with AI.

Human thought is not only verbal. It is spatial, gestural, and continuous. To bridge the gap between the linear text scroll and the embodied nature of human cognition, this research proposed materializing AI-generated dialogue as handleable digital artifacts. TMS is the system architecture of that proposition. FIELD is the software application that gives it form: a Network Graph canvas where ideas can be held in place, relationships can be questioned, and a held fingertip and a spoken sentence together can direct a thought.

The thesis is careful about what it has and has not shown. It has shown that the proposition is constructible: that materialized, multimodal AI dialogue can be organized as a coherent interface system, grounded in formative research and transparent about the platform on which it runs. It has not yet shown that the system delivers its intended effects in comparison with existing linear-scroll AI workflows. That evaluation remains future work. The claims here end where the evidence ends.

Yet the larger implication of this thesis is not only the prototype. It is a reorientation of the design question. In many dominant computing interfaces, human intent is translated into forms the machine can process: typed commands, structured inputs, and linear sequences of interaction. This thesis asks what happens when the direction is reversed. Instead of asking human thought to fit the interface, how might the interface move closer to the ways humans have long reasoned with one another — through speech, gesture, shared attention, spatial reference, hesitation, return, and touch?

Seen this way, the work of a designer is not simply to add another object, screen, or feature to the world. It is to design the relationships among a person, a generative system, an interface, a material surface, and the space in which thinking unfolds. FIELD is one attempt at that composition. It is bounded by the hardware available to it and the time a thesis allows, but the direction it points toward is larger than the prototype: interaction that begins from the conditions of human reasoning, and interfaces that follow from that interaction rather than dictating it.

The relationship between a person and a generative system does not have to settle into reading and editing. AI can generate possibilities fluently and abundantly. The human can author direction: framing what matters, shaping what forms, connecting what belongs, pruning what does not, and seeing the field whole. The interface between them shapes whether that relationship is possible.

The goal was never efficiency alone. It was authorship — restored to the hand, anchored in space, and exercised one touch at a time.

References

Allen, J., Ferguson, G., & Stent, A. (2001). An architecture for more realistic conversational systems. In ***Proceedings of the 6th International Conference on Intelligent User Interfaces*** (pp. 1–8). Association for Computing Machinery.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In ***Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*** (Paper 3, pp. 1–13). Association for Computing Machinery.

Barsalou, L. W. (2008). Grounded cognition. ***Annual Review of Psychology, 59***, 617–645.

Clark, A., & Chalmers, D. (1998). The extended mind. ***Analysis, 58***(1), 7–19.

Dourish, P. (2001). ***Where the action is: The foundations of embodied interaction***. MIT Press.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. ***American Psychologist, 34*** (10), 906–911.

Frayling, C. (1993). Research in art and design. ***Royal College of Art Research Papers, 1*** (1), 1–5.

Goldin-Meadow, S. (2003). ***Hearing gesture: How our hands help us think.*** Harvard University Press.

Hornecker, E., & Buur, J. (2006). Getting a grip on tangible interaction: A framework on physical space and social interaction. In ***Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*** (pp. 437–446). Association for Computing Machinery.

Horvitz, E. (1999). Principles of mixed-initiative user interfaces. In ***Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*** (pp. 159–166). Association for Computing Machinery.

Ishii, H., & Ullmer, B. (1997). Tangible bits: Towards seamless interfaces between people, bits and atoms. In ***Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*** (pp. 234–241). Association for Computing Machinery.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. ***Cognitive Science, 18*** (4), 513–549.

Lakoff, G., & Johnson, M. (1999). ***Philosophy in the flesh: The embodied mind and its challenge to western thought.*** Basic Books.

Malafouris, L. (2013). ***How things shape the mind: A theory of material engagement.*** MIT Press.

Malafouris, L., & Renfrew, C. (Eds.). (2010). ***The cognitive life of things: Recasting the boundaries of the mind.*** McDonald Institute for Archaeological Research.

McNeill, D. (1992). ***Hand and mind: What gestures reveal about thought.*** University of Chicago Press.

O'Keefe, J., & Nadel, L. (1978). ***The hippocampus as a cognitive map.*** Oxford University Press.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? ***International Journal of Human-Computer Studies, 45*** (2), 185–213.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. ***Cognitive Science, 12*** (2), 257–285.

Tversky, B. (2019). ***Mind in motion: How action shapes thought.*** Basic Books.

Ullmer, B., & Ishii, H. (2000). Emerging frameworks for tangible user interfaces. ***IBM Systems Journal, 39*** (3–4), 915–931.

Varela, F. J., Thompson, E., & Rosch, E. (1991). ***The embodied mind: Cognitive science and human experience.*** MIT Press.

Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in HCI. In ***Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*** (pp. 493–502). Association for Computing Machinery.

Written and designed by Dahae Cheon