MAIK

Modular Artificial Intelligence Kit

Emma Furia

Colofon

Master Thesis

Author

E.E.F (Emma) Furia

Student number 5091527

MSc Design for Interaction

Faculty of Industrial Design Engineering

Delft University of Technology

Chair TU Delft

Prof. Gerd Kortuem

Mentor TU Delft

JD (Derek) Lomas

Collins Aerospace

Arnau Castillo

Jeff McKee

Summary

Designers' individual already use AI. Inside every design team, each designer has their own approach for working with ai. What one person learns rarely reaches the rest of the team. This thesis asks how AI can become part of a design team's early-stage workflow as a shared practice, not as twelve personal habits running in parallel.

This thesis followed a Research through Design approach with the design team of a aviation company. Three iteratives' cycles go rough round of: interviews, workflow analysis and low-fidelity AI probes reframed the original brief twice. First, the workflow problem was not communication between disciplines. It was sensemaking under pressure. Designers have to build a shared understanding from fragmented inputs before they can communicate anything at all. Second, the real gap was not a missing AI tool. It was the missing structure for using the tools already in the room. Without that structure, AI use stays personal and what one designer learns stays with that one designer.

These findings led to MAIK, a Modular AI Kit. MAIK treats AI as a team practice, not an individual skill. It has three layers. Layer 1 is a deck of physical cards that give designers a tangible way to browse, sort and argue about AI capabilities together. Layer 2 is the MAIK Canvas, a worksheet a team fills in together to build reusable Claude Skills. Layer 3 chains those Skills into longer workflows with a human review step between each.

The three layers are built on scaffolding theory (Wood, Bruner & Ross, 1976). In learning theory, scaffolding describes temporary support that lets a learner work beyond their independent ability and withdraws as competence grows. MAIK applies this to team AI literacy. Layer 1 offers the most support. The cards lower the barrier of the blank page. Layer 2 offers less. The Canvas structures the thinking behind a good Skill, but it does not write the Skill for the team. Layer 3 sits outside the current scope and offers almost no support at all. The team orchestrates its own workflows.

Each layer builds on what the one before it taught. The framework is designed to become less necessary over time. MAIK is learning infrastructure, not production infrastructure. Its job is to build a team's AI literacy, not to automate its output.

MAIK was evaluated in two studies. The first ran with seven IDE Master's students at TU Delft across three in-person sessions. The second ran with a CXD designer at Collins Aerospace on two real project briefs. Both studies point to the same answer. Where designers already use AI, the work is not to give them new tools. It is to give them a shared structure, so individual AI use becomes a team capability.

AI Statement
This thesis is about AI, and I used AI while working on this thesis. As a designer who works with these tools every day, it felt natural to bring them into the process of building and documenting this project.

The framework and the prototype application were built with Claude Code. I used Claude (Anthropic) during the process for brainstorming, for structuring arguments, critique and refining my writing.

This thesis is mine. Every chapter, every argument, and every decision went through my own judgment and voice.

Disclaimer

Due to confidentiality, are some parts not visible

Glossary

Term	Definition
Agent	An AI system that can take a sequence of actions toward a defined goal, using tools and memory to do so. Unlike a simple prompt-response exchange, an agent can call external services, read and write files, browse the web, or trigger other processes. The designer sets the goal and the constraints; the agent determines the steps. What matters for practice is the boundary between what the agent decides autonomously and what requires human input, a boundary known as the human-in-the-loop checkpoint.
AI (Artificial Intelligence)	A broad field of computer science focused on creating systems capable of performing tasks that typically require human intelligence, such as reasoning, learning, and decision-making.
AI Capability	A potential functional behaviour an AI system can express through interaction (e.g. summarisation or clustering), emerging from the combination of model behaviour and prompting rather than from fixed features.
AI Capability Framework	An analytical framework developed in this thesis that describes AI systems in terms of interaction, system processing, and expressed capabilities rather than interface elements.
Automation Paradox	The observation that as automation increases, human situational awareness decreases, meaning the moments when human oversight matters most are precisely the moments when the human is least prepared to provide it (Parasuraman & Riley, 1997). At Collins Aerospace, this grounds the requirement for layered human review checkpoints in MAIK.
Boundary Object	An artefact that satisfies the informational requirements of multiple communities of practice simultaneously (Star & Griesemer, 1989). In this thesis, the PowerPoint deck at Collins functions as a boundary object: simultaneously a thinking tool, coordination artefact, and evaluation document.
Context Engineering	The practice of designing the full informational environment a model operates in, including roles, memory, documents, and tool access, rather than optimising individual prompts. Represents the shift from prompt engineering to environment design (Rajasekaran et al., 2025).
Context Window	Everything a model can see at once: the instruction, the conversation history, any documents provided, and the system-level configuration. By 2024, context windows had expanded from a few thousand tokens to over 200,000, making it possible to feed entire reports and knowledge bases into a single session.
CXD (Cabin Experience Division)	The division within Collins Aerospace responsible for passenger-facing cabin interior design, including galleys, seating, lighting, and experience concepts.
Divergent Exploration (AI Capability)	An AI capability involving generating alternatives, surfacing unexpected connections, and broadening the solution space. Most critical in the early Discover phase of the First Diamond and the least consistently supported by existing AI tools.
Double Diamond	A design process model describing four phases: Discover, Define, Develop, Deliver, characterised by cycles of divergence and convergence (Design Council, 2024).
Externalisation	The act of making tacit or implicit knowledge explicit through artefacts such as sketches, text, diagrams, or visualisations.
First Diamond	The early phase of the Double Diamond (Discover and Define), characterised by ambiguity, sensemaking, and problem framing. This is the primary focus of the research.
Foundational Models	Large-scale pre-trained architectures, such as GPT-4, Claude, Gemini, and Llama, that serve as the building blocks from which specific AI tools are constructed. They determine what kinds of inputs the system can process and what kinds of outputs it can produce.
Fuzzy Front End (FFE)	The early stage of design where problem definitions are unclear, information is fragmented, and multiple interpretations coexist (Sanders & Stappers, 2008). Overlaps with the First Diamond.
Generative AI (GenAI)	A subset of AI that generates new content such as text, images, audio, or video.
Gulf of Envisioning	The cognitive distance between what a user can imagine and what they can actually produce through unstructured interaction with an AI system. It manifests as three gaps: Capability (not knowing what the system can do), Instruction (not being able to express intent as a prompt), and Intentionality (not being able to evaluate whether the output meets the underlying need) (Subramonyam et al., 2024).
Hallucination	The tendency of large language models to generate confident-sounding statements that are factually incorrect or not grounded in verifiable sources. Human verification is structurally necessary, not optional.
Human-in-the-loop	A boundary in AI workflows where the system pauses for human review, decision, or approval before proceeding. In MAIK, human review checkpoints are a structural requirement at every layer.
Interaction Layer	The layer of human-AI interaction where designers articulate intent through prompts, constraints, and feedback.

Glossary

Term	Definition
Knowledge Conversion	The process by which knowledge transforms between tacit and explicit forms within organisations. Nonaka and Takeuchi (1995) describe four modes: socialisation (tacit to tacit), externalisation (tacit to explicit), combination (explicit to explicit), and internalisation (explicit to tacit). MAIK cards function as a combination mechanism; the canvas Team Alignment Zone is where genuine externalisation happens.
Large Language Models (LLMs)	AI models trained on large corpora of text data that can generate, summarise, translate, and reason over language. In this research, LLMs are treated as design materials whose behaviour emerges through interaction.
MAIK (Modular AI Kit)	The three-layer AI interaction toolkit designed for the CXD team, consisting of MAIK Cards (Layer 1), MAIK Canvas with Claude Skills (Layer 2), and an orchestrated multi-agent workflow (Layer 3). The name stands for Modular AI Kit.
MAIK Canvas	An A3 group artefact adapted from the Prompt Canvas (Hewing & Leinhos, 2024) that structures AI interaction configuration across eight components: Persona/Role, Task and Intent, Context, Output, Audience, Step-by-Step, References, and Tone. Used as a shared team activity before AI-assisted research or synthesis tasks.
MAIK Cards	A physical card set of 56 cards across 7 categories (Research, Analysis, Ideation, Critique, Synthesis, Writing, Collins-specific) that present structured AI prompt templates for design tasks. Each card provides a trigger question, a fill-in prompt format, and usage guidance.
MCP (Model Context Protocol)	An open standard developed by Anthropic that enables AI models to connect securely to external data sources, tools, and services, allowing them to access real-time information and execute actions beyond their training data.
Multimodal AI	AI systems capable of processing and generating multiple types of input and output (e.g. text, images, audio), relevant for externalising tacit design knowledge through mixed representations.
Prompt Canvas	A practitioner-facing framework developed by Hewing and Leinhos (2024) that externalises the tacit knowledge of effective prompting into eight defined dimensions. It is the strongest prior art for Layer 1 of MAIK, but explicitly scoped to single-prompt, text-to-text use.
Prompt Engineering	The practice of crafting a single instruction to guide AI model behaviour. Distinguished from context engineering, which designs the full informational environment the model operates in.
Prompting	The practice of guiding AI behaviour through linguistic instructions, constraints, examples, and iterative refinement. Treated as a structured interaction practice in this thesis.
Research through Design (RtD)	A research methodology in which knowledge is generated through iterative cycles of designing, prototyping, reflecting, and refining. Used in this thesis to explore AI in a specific organisational context (Stappers & Giaccardi, 2017).
Scaffolding	An educational concept describing temporary support structures that enable a learner to achieve tasks beyond their current independent capability (Vygotsky, 1978; Wood, Bruner & Ross, 1976). In MAIK, the three layers function as progressive scaffolding: each layer is achievable given the competence built at the previous one, and the scaffolding withdraws as AI literacy grows.
Semantic Transformation	An AI capability referring to summarising, rewriting, translating, or restructuring existing information into more usable forms.
Sensemaking	The process by which designers interpret, structure, and give meaning to fragmented, ambiguous, or evolving information during early-stage design work (Weick, 1995).
Sensemaking and Structuring (AI Capability)	An AI capability involving clustering, organising, and relating unstructured inputs to support understanding and pattern recognition.
Situational Awareness	A model from aviation human factors (Endsley, 1995) describing three levels of awareness: perception (noticing relevant information), comprehension (understanding what it means), and projection (anticipating what happens next). Used in Chapter 5 to ground the automation paradox and justify layered human review in MAIK.
Skills	Persistent, purpose-specific AI instruction sets in Claude's architecture that configure the model's behaviour for a defined task. Skills are lazy-loaded: they remain dormant until triggered by a relevant request, then execute a complete structured workflow without the user needing to re-explain the method. Skills are the Layer 2 implementation mechanism in MAIK.
Subagent	A specialised AI agent operating within a larger multi-agent system, assigned a single well-defined task (e.g., research, synthesis, output formatting) with its own isolated context window. Subagents prevent context pollution and enable parallelisation. In MAIK Layer 3, connected Skills function as subagents running in sequence.
Tacit Knowledge	Implicit, experience-based knowledge that is difficult to articulate or formalise, often underlying intuition and aesthetic judgement in design decisions.
Three-Layer Prompting Canvas	The full conceptual name for the MAIK framework, describing its progressive structure: Layer 1 (Interaction: Cards + Canvas), Layer 2 (Agents: Claude Skills), and Layer 3 (Orchestration: multi-agent workflows with human review checkpoints).
Verification & Constraints (AI Capability)	An AI capability involving checking outputs against defined criteria, applying constraints, and ensuring consistency. Becomes especially critical at phase gates when the design direction must be defensible.
Zone of Proximal Development	The gap between what a learner can do alone and what they can do with support (Vygotsky, 1978). Applied in this thesis to frame MAIK's three layers as progressive scaffolding matched to growing AI literacy.

TABLE OF CONTENTS

CHAPTER 1

Introduction

Collins Aerospace published one graduation brief. It asked a student to explore AI integration across the aircraft interior design workflow. This chapter introduces the context, traces how the problem found its shape, presents the research questions, and describes the project approach.

1. Introduction

Collins Aerospace initially published one graduation brief focused on exploring AI in the aircraft interior design workflow. After scoping discussions, it was divided into two separate assignments. One student was assigned to the later stages: using AI for idea generation, concept development, and visual rendering. I was assigned to the early stage, where designers frame the problem, gather and structure research, and build alignment with stakeholders before committing to a direction.

Collins sees clear opportunities for AI to reduce time across their workflow. They directed me to the early stage specifically because this is where time pressure is most severe and the work is least structured. There is no solution yet. The designer must construct a shared understanding from fragmented inputs, satisfy multiple audiences simultaneously, and produce a defensible direction, often within weeks. Collins identified this as their harder problem. That decision shaped everything that followed.

1.1 The Context

Collins Aerospace is one of three business units of the RTX Group (formerly Raytheon Technologies), alongside Pratt & Whitney and Raytheon. RTX employs approximately 185.000 people worldwide (RTX Corporation, 2024), of whom roughly 80.000 work within Collins Aerospace across more than 300 locations globally (Collins Aerospace, n.d.). While Raytheon focuses on defence systems and Pratt & Whitney specialises in aircraft engines, Collins Aerospace covers a large portfolio of aviation systems and cabin interior solutions for the commercial aviation industry.

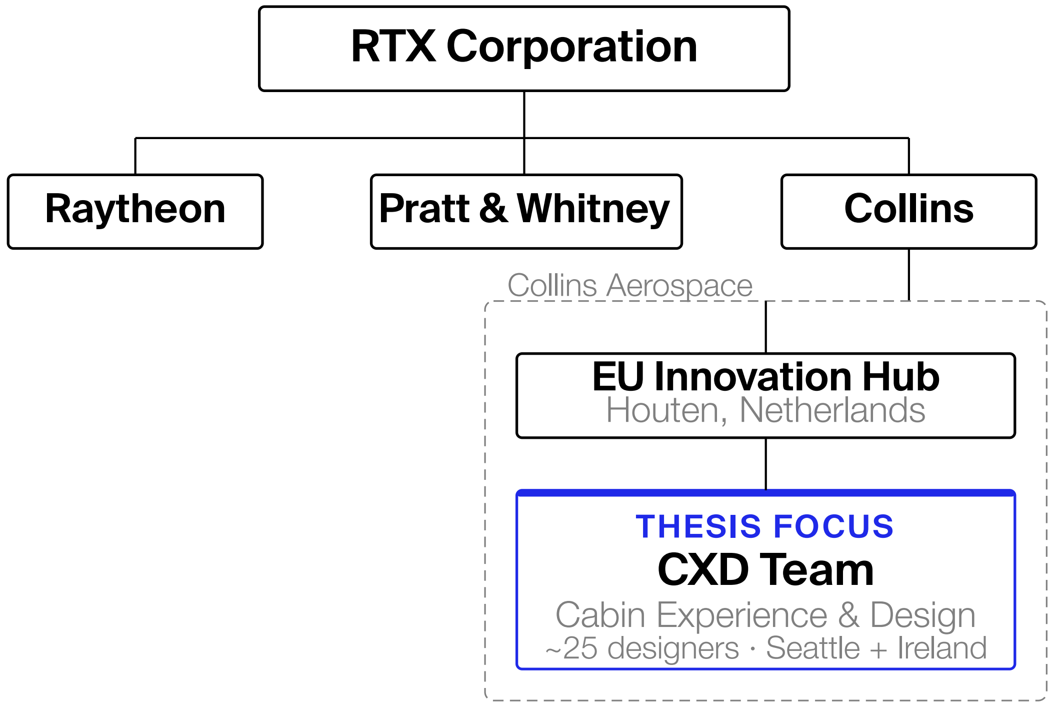


Figure 1.1 - The three business units of the RTX Group: Raytheon, Pratt & Whitney, and Collins Aerospace. Source: RTX Corporation (2024).

In the Netherlands, Collins operates an EU Innovation Hub in Houten, established in partnership with TU Delft and Pratt & Whitney's European Technology Centre in 2024. The hub focuses on sustainable advanced materials, intelligent systems, novel cabin experiences, and innovative manufacturing processes (TU Delft, 2024). This thesis was conducted in direct collaboration with this hub.

Domain

The domain of this thesis is specifically the Cabin Experience & Design (CXD) team. Although Collins' portfolio includes avionics, flight controls, and landing systems, the CXD team focuses on passenger-facing cabin solutions: galley systems, lighting concepts, seating configurations, and full interior experience visions. The team consists of approximately twenty-five industrial design engineers, with its largest group based in Seattle, USA, and a smaller group in Ireland. As a result, collaboration relies heavily on remote communication tools, making distributed workflows a fundamental aspect of daily design practice.

The CXD team collaborates across a dense network of stakeholders: airlines, OEMs, engineering teams, certification bodies, and internal business leadership. Each speaks a different language and evaluates work through a different lens. Chapter 3 describes this network and its implications in detail. What matters here is the structural consequence: every piece of design work must satisfy multiple audiences simultaneously, under time pressure, with limited shared vocabulary. That is the environment into which any AI support must fit.

1.2 Problem Framing

The original Collins brief positioned the research problem as one of tool integration. Which AI tools (Midjourney, DALL-E, Stable Diffusion, ChatGPT) could enhance the design workflow, and how? Before the project formally began, Collins refocused the assignment. Rather than mapping AI tools across the full workflow, this project would address the early stage specifically: the Discover and Define phases of the Double Diamond, where designers face their most communication-intensive and cognitively demanding challenges.

Communication Challenge

The initial framing of this narrowed problem positioned it as a communication challenge. Two areas appeared particularly relevant through early observations and discussions with the team.

First, concerned how designers communicate across disciplinary boundaries. Designers at Collins frequently need to translate their work into formats and terms that make sense to engineering, programme management, business, or certification teams, rather than using design-focused language. This translation takes time away from actual design work. Considerable effort goes into explaining decisions, research insights, and project progress to non-design stakeholders. Often in formats like slide decks, structured reports, and alignment presentations that require significant production time in their own right.

The second area that appeared particularly relevant through early observations and discussions was the space available for exploration during the early phases of a project. The combination of fragmented information, shifting requirements, and limited time constrains the breadth and depth of exploration possible within standard project timelines. Designers reported that they often felt they had to commit to a direction before they had adequately explored alternatives, because the deadline of the next phase gate left no margin for extended divergence.

Cycle 1 research, conducted through interviews and workflow mapping at Collins, revealed that both of these challenges (communication difficulty and limited exploration) were symptoms of a more fundamental problem.

Sensemaking

The underlying challenge was sensemaking: how designers construct shared understanding during the First Diamond, in conditions of ambiguity, fragmented information, and compressed timelines.

The communication difficulty was not primarily about translation. Designers often had a clear sense of direction, built on intuition, spatial judgement, and accumulated experience. But this knowledge is tacit. It resists the kind of measurable, evidence-based reasoning that stakeholders require. The challenge was converting design intuition into defensible arguments: giving engineering teams quantifiable rationale, giving business leaders traceable evidence, giving certification bodies structured proof. The designer was doing the thinking and the justifying simultaneously, in the same document, under the same deadline.

Reframing

This reframe (from visual tools, to stakeholder communication, to sensemaking in the First Diamond) shaped the direction of the entire project. Interviews also revealed a compounding structural problem: designers atCollins had no shared framework for using AI. Individuals experimented with tools like ChatGPT in isolation, developing their own inconsistent approaches without shared vocabulary, shared method, or any organisational structure to support adoption or learning. The absence of a framework was itself part of the gap.

Together, these findings pointed toward a specific design direction: not just a tool, and not just a communication template, but a structured framework for AI interaction. One that provides the shared vocabulary, staged architecture, and cognitive scaffolding that the team currently lacks.

1.3 Research Questions

This thesis is guided by three research questions, developed through the problemframing process described in chapter 1.2. These research questions are addressed sequentially across the three cycles, with each cycle contributing to one or more of the questions.

Research trough design

These questions are addressed through an iterative Research through Design approach in which understanding of the problem and exploration of possible interventions develop in parallel across multiple cycles.

RQ1 is addressed through Cycle 1 (workflow understanding and problem framing).

RQ2 is addressed through Cycle 2 (AI capability exploration and prototype development) and the literature review.

RQ3 is addressed through Cycle 3, which focuses on the development and evaluation of the concept

These research questions are deliberately sequenced, rather than parallel: it is not possible to evaluate what AI Capabilities might improve the workflow without first understanding what the workflow needs. It is not possible to design appropriate AI support without understanding which capabilities are actually relevant. The research question structure reflects this logic.

Cycle 1

RQ1 What are the current problems in the design workflow at Collins Aerospace, particularly in terms of communicationdemands and exploratory capacity?

Cycle 2

RQ2 Which capabilities of AI are relevant to these problems, and how might they be applied in the design workflow?

Cycle 3

RQ3 Do these AI applications improve the workflow in practice — and if so, how?

1.4 Project Approach

The chosen method for this research is Research through Design (RtD). In RtD, designing itself serves as a method of inquiry rather than a path toward a predetermined solution (Stappers & Giaccardi, 2017). The problem space and solution space develop together through iterative cycles of designing, prototyping, evaluating, and refining. Neither is fully understood before the other is explored (Dorst & Cross, 2001). This makes RtD particularly appropriate for this project. The problem itself was subject to reframing. The designed artefact needed to respond to a specific, partially tacit organisational context (Figure 1.2).

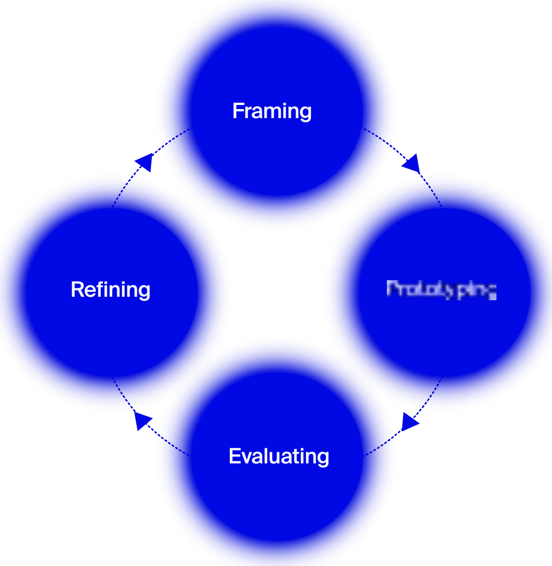


Figure 1.2 - Four phases, own interpretation, of the RtD approach: Framing, Prototyping, Evaluating, Refining

Artefacts to learn from

Prototypes in this approach are not finished products to be evaluated for performance. They are artefacts to learn from: they surface insights into how designers work, communicate, and explore. They make partially understood problems more visible by showing what a response to them looks and feels like in practice. The shift from problem space to solution space happens gradually through this process, as each cycle deepens understanding and narrows the design direction.

This project goes through three research cycles, each emphasising different phases:

Cycle 1: Framing

Corresponds to the Framing phase. The focus was developing a grounded understanding of the early-stage design workflow at Collins and articulating the problem space. Data is gathered through semi-structured interviews with CXD designers, workflow mapping sessions, and analysis of existing communication artefacts. Cycle 1 produces the sensemaking reframe and the eight structural challenges that guide the subsequent cycles. Chapters 2 and 3 present this work.

Cycle 2: Exploring

Aligns primarily with the Prototyping phase. Building on Cycle 1, this cycle mapped the eight challenges to five AI capability categories, scanned fifty existing AI tools, and explored relevant capabilities through eleven low-fidelity AI probes. Three concept directions emerged and were compared through the Harris Profile evaluation. A post-midterm client alignment meeting established five success criteria. Chapter 4 presents this work.

Cycle 3: Developing

It begins with a literature review grounding the chosen direction (Chapter 5), formalises the design brief and programme of demands (Chapter 6), presents the full MAIK framework: cards, canvas, Skills architecture, and orchestrated workflow (Chapter 7), and evaluates it through structured sessions with designers and the CXD team (Chapter 8). At last answer the research question personal reflection on the methods and learning objectives (Chapter 9)

1.5 Report Structure

This report follows the structural logic established by the RtD approach: it moves from context and theory toward the designed artefact, with each chapter building on the previous one.

Chapter 2 establishes the AI landscape: what AI systems can currently do, how the interaction paradigm has evolved from prompt engineering toward context engineering and agentic architectures, and why this evolution matters for design practice.

Chapter 3 documents the design workflow at Collins in depth, based on Cycle 1 research: the Double Diamond structure, the phase gate process, the challenges of the First Diamond, and a detailed case study of a single project that illustrates all of these challenges in concrete, measurable terms.

Chapter 4 synthesises Chapters 2 and 3 to identify where AI capabilities meet workflow needs: the opportunities for integration. It justifies the direction chosen for Cycle 3.

Chapter 5 provides the theoretical grounding for the Three-Layer Canvas through literature on canvas frameworks, context engineering, multi-agent systems, and organisational AI adoption.

Chapter 6 presents the design brief: the explicit statement of what is missing and what the canvas must achieve.

Chapter 7 presents MAIK in full: its architecture, agent specifications, visual design, organisational implications, and implementation pathway.

Chapter 8 presents the evaluation: study design, participant groups, session structure, and results from the evaluation sessions.

Chapter 9 concludes by answering the research questions, stating the contribution to knowledge, and providing recommendations for Collins and for the broader design field.

CHAPTER 2

AI Capabilities in Design Workflows

What AI can do today, what tools exist, and why designers cannot yet use them effectively. This chapter maps the technology landscape, identifies the post-hoc gap in current tools, and introduces the AI Capability Framework that structures the rest of the thesis.

2. AI Capabilities in Design Workflows

What AI can do today, what tools exist, and why designers cannot yet use them effectively

Before exploring how AI might support the Collins workflow, it is necessary to establish what AI systems can currently do, how they work, and how the nature of human-AI interaction has changed. This chapter approaches AI in three ways. First, the technical hierarchy that situates generative AI within the broader field. Second, the evolution of how designers interact with these systems, from typed commands to prompt engineering to something more architectural. Third, a framework for the specific capabilities that are most relevant to early-stage design work. Together, these sections build the conceptual vocabulary that later chapters draw on directly. It combines published research with a structured landscape scan of fifty AI tools conducted as part of this project, and answers three questions;

1. What can AI systems do today?

2. What do tools that designers encounter today look like?

3. And why does interaction between designer and AI so consistently fail, even when the designer knows exactly what they need?

Together, these question build the conceptual vocabulary that later chapters draw on directly.

2.1 The AI Landscape

Artificial intelligence is a layered field in which more specialised techniques are nested within broader ones (Figure 2.1). At the highest level, AI consists of computational methods that enable machines to simulate human cognitive functions. Like reasoning, learning, pattern recognition, and problem-solving (Russell & Norvig, 2016). Within this broad field, Machine Learning is the subset in which systems learn patterns from data rather than from pre-programmed rules. Within Machine Learning, Deep Learning uses multilayered neural networks to extract progressively more abstract features from raw inputs. This is the architecture that powers most of the AI tools designers encounter today. Generative AI sits within Deep Learning. It includes a family of model types (transformer models, diffusion models, generative adversarial networks) that have learned patterns from enormous datasets and can generate new content that is statistically consistent with what they have learned.

Transformer models underlie most large language models; diffusion models power image generators like Stable Diffusion and Midjourney. What distinguishes generative AI from earlier AI systems is that it does not recognise what it is given. It creates something new. Large Language Model Systems that simulate human cognitive functions, algorithms that learn patterns from data, and multi-layered neural networks. Models that create new content: Transformer, diffusion, GAN, Text-based generative models trained on big data sets.

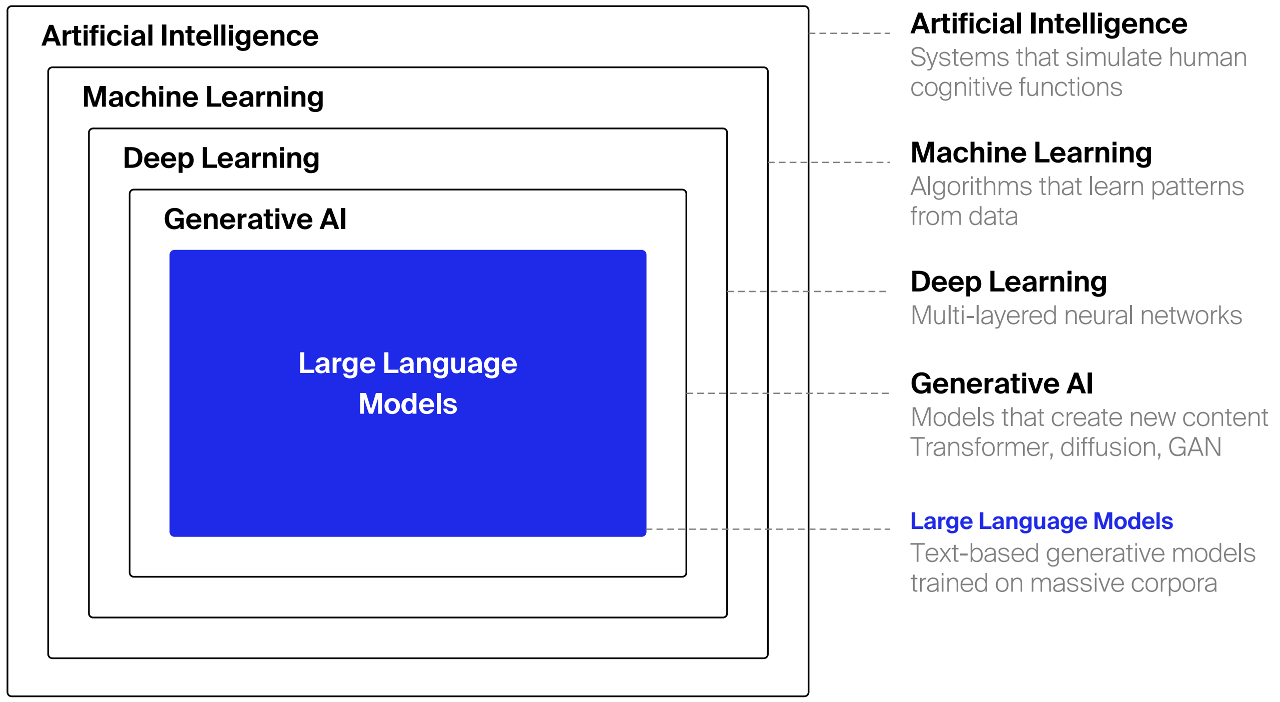


Figure 2.1 – The layered hierarchy of intelligent systems, situating Large Language Models within the broader field of Artificial Intelligence. Based on

Foundational Models
Foundational models (large-scale pre-trained architectures such as GPT-4, Claude, Gemini, and Llama) serve as the building blocks for specific AI tools (Figure 2.2). They determine what kinds of inputs the system can process (text, images, audio, structured data) and what kinds of outputs it can produce.

Multimodal models
Multimodal models combining multiple foundational architectures: a single system can receive an image, process additional text, and return a written analysis, a reformatted summary, or a new image derived from both inputs. For design practice, where sensemaking (structuring information, exploring alternatives, communicating across disciplines) is the relevant technology is large language models and the multimodal models built on them. The visual generation tools (Midjourney, DALL-E) that dominate the public conversation about "AI in design" address a different part of the workflow. This AI landscape has gone through evolution, especially in the last two years.

To understand what AI can do today in design practice, requires a closer look at where the technology stands right now. For the Collins context, the most directly relevant part of this landscape is generative AI. This includes large language models for language-mediated tasks such as summarising, structuring, translating, and reasoning, as well as multimodal systems that combine text with visual and formatted output.

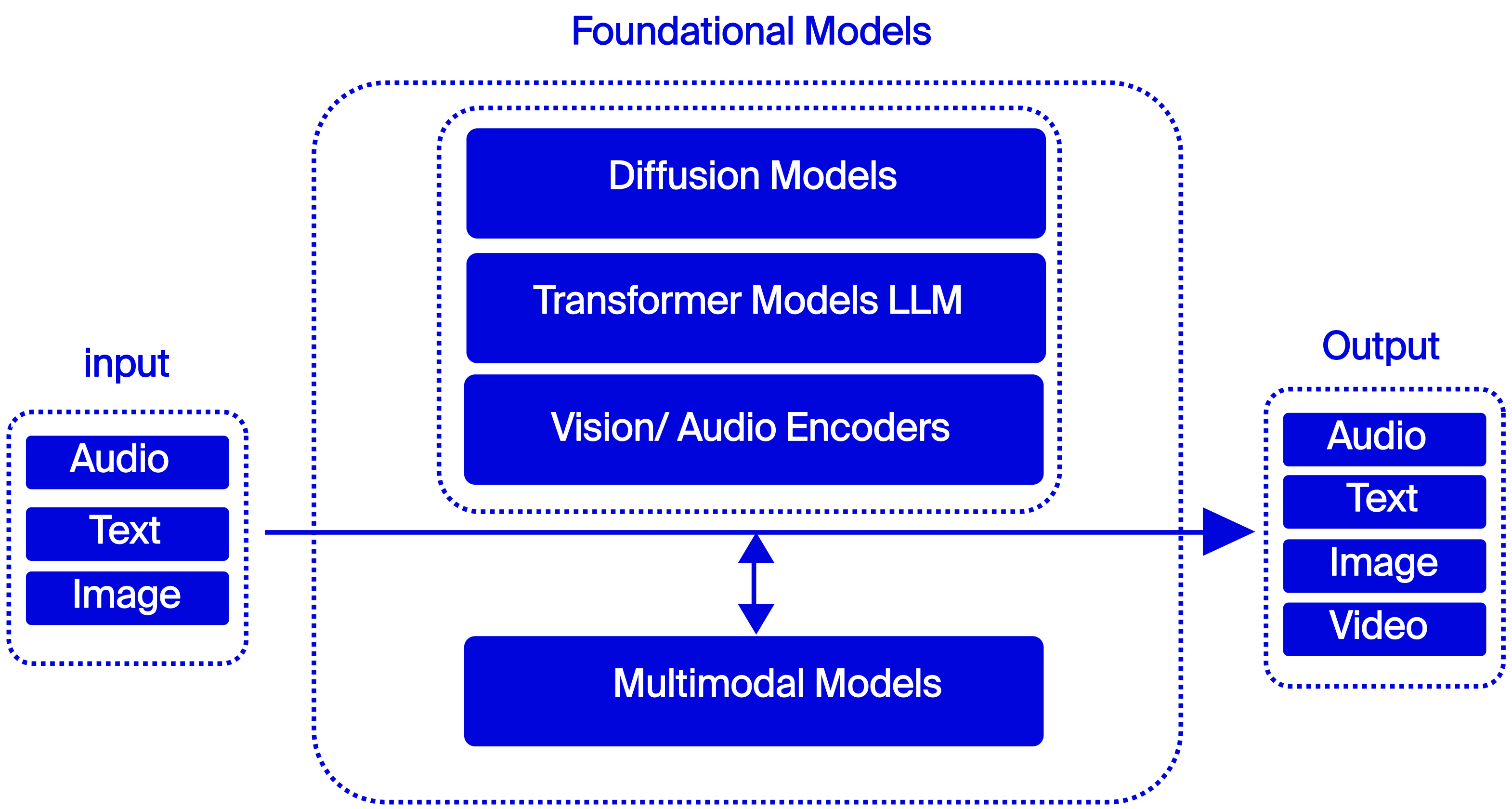


Figure 2.2 – Foundational model architectures and their role in multimodal input-output processing.

2.2 The State of AI in 2026

By 2026, AI systems will do more than generate text from a single instruction. They hold persistent roles, access external tools, maintain memory across sessions, and coordinate multi-step workflows. The terminology around these capabilities is still changing. Terms like agentic workflows and context engineering appear in industry faster than literature can standardise them.

The definitions below follow the usage in Huang (2025) and Mei et al. (2025). The most important shift in this thesis is from prompt engineering to context engineering: from writing a single instruction to designing the full environment the model operates in. Chapter 5 provides the deeper theoretical grounding for each concept introduced here.

From prompt engineering to context engineering
The context window is everything a model can see at once. The instruction, the conversation history, any provided documents, and the system configuration. By 2024, context windows have increased from a few thousand tokens to over 200,000 (Anthropic, 2023; Google DeepMind, 2024; OpenAI, 2023). This makes it possible to add entire reports or big datasets into a single conversation. This addition introduced a critical shift in the capabilities users were required to have. Previously: prompt engineering, the practice of crafting a single instruction, to now: context engineering—the broader discipline of designing the full informational environment the model operates in, including roles, memory, documents, and tool access (Huang, 2025; Mei et al., 2025).

Agents
An AI conversation only answers or does something when asked. An agent is an AI system that can take a sequence of actions toward a goal, using tools and saved memory (Huang, 2025). Unlike a conversational AI, an agent can reach out to external services, read and write files, browse the web, or trigger other processes. The user sets the goal and the constraint. Then the agent determines the steps and executes them. What matters for practice is the boundary between what the agent decides autonomously and what requires human input: the human-in-the-loop approach.

Subagents
A subagent is a specialised agent operating within a larger workflow, assigned a single well-scoped task. Each operates with an isolated context window: it sees only what is relevant to its task, not the full history of the workflow. This isolation prevents context pollutionand reduces tokens.

The Model Context Protocol (MCP)
An agent on its own can only work with what is already in its context window. The Model Context Protocol (MCP) is an open standard that connects agents to external tools, databases, APIs, and other services like GitHub or Figma (Anthropic, 2024). The MCP works through a server-client architecture. An MCP server exposes tools that the agent can use. Protected by API keys and permissions, making access to the cost and usage data. MCP turns a language model into a system that can act and access several applications.

Where MCP gives an agent access to tools (the what), a Skill gives it expertise (the how).

Skill
A Skill is a reusable prompt or instruction: the steps to follow, the output format to produce, and the constraints to respect. It is a saved instruction set that tells the agent what steps to take, what format the output should have, and what constraints to follow. It sits between a one-off prompt (which disappears after one session) and a hardcoded program (which cannot adapt). Without a Skill, the same domain knowledge has to be typed out again every time a new conversation starts.

Limitations and failure modes
None of this works perfectly. Language models are probabilistic, which means they guess based on patterns. Sometimes those guesses are wrong, and the model does not flag them. That is a hallucination: output that reads confidently but has no factual basis. Models easily misstep and misread what is being asked. This produces outputs that sound relevant, but are based on non-credible sources (Shneiderman, 2020; Liu et al., 2023). Here the quality of context plays an important part in this. If the model is fed with too much irrelevant material, it starts latching onto the wrong things. Mei et al. (2025) call this context distraction. Start with a badly structured input, and the reasoning goes sideways from the beginning, regardless of what comes after. They call that prompt pollution. Problems mean that humans must check everything.

Human verification remains necessary at every stage.

Takeaways

1. Context engineering means designing the full environment the model works in, not just writing a good prompt. What limits the quality of the output is usually the context.

2. An agent works toward a goal in steps, using tools and memory. The designer sets the goal. The agent figures out how to get there.

3. A subagent takes one piece of a larger task and works on it in isolation. It only sees what it needs to see. That keeps the context clean.

4. MCP connects agents to external services like databases, file systems, and APIs. Access is controlled through API keys, so usage stays trackable.

5. A Skill is a saved instruction set. MCP gives the agent access to tools. A Skill tells it how to use them. Without one, domain knowledge has to be re-explained every session.

6. Language models get things wrong. They produce output that sounds confident but can be factually incorrect. Checking every output is part of the process.

2.3 The Current AI Tool Landscape

To map the current state of AI tools available to designers, a scan of fifty AI applications was conducted as part of this project. The selection was based on the a16z AI Application Spending Report (Andreessen Horowitz, 2025), which provides an empirically grounded view of where AI development and investment are concentrated. The scan was exploratory rather than systematic: not a comprehensive survey, but a structured search for patterns in how AI capabilities are currently expressed in available tools. The detailed methodology and capability-challenge mapping are presented in Chapter 4. Cross-domain Translation (reframing information from one disciplinary register to another) appeared less frequently and less consistently. Figure 2.3 shows the full set of fifty applications analysed.

Four dominant capability patterns

Four types of applications were prominent during the tool scan. The most common was Semantic Transformation: tools that transcribe, summarise, rewrite, or reformat content. Sensemaking & Structuring was the second most frequent. These tools take scattered research inputs and try to organise them: clustering notes, grouping themes across documents, sorting through messy data. Externalisation & Visualisation appeared in image generators, dashboard builders, and tools that auto-format reports or presentations. The rarest was Cross-domain Translation. Only a handful of tools attempt to take information written for one discipline and reframe it for another. The ones that do are inconsistent in quality.

Figure 2.3 – Overview of fifty AI applications analysed, selected based on the a16z AI Application Spending Report (Andreessen Horowitz, 2025).

The post-hoc gap

A critical finding emerged from this scan: the vast majority of tools operate in a reactive, post-hoc mode. They help designers after information has already been produced – transcribing a meeting that already happened, summarising a document that already exists, reformatting a slide deck that has already been built. They reduce documentation effort. They do not support sensemaking in the moment: the active, exploratory, ambiguity-managing work that characterises early-stage design.

> 'Existing tools are largely post-hoc. They arrive after the thinking has already happened. What designers need is support during the thinking.'
>
> – Synthesis from tool landscape scan

Divergent Exploration and active reframing, the capabilities most relevant to managing early-stage ambiguity, were the least consistently supported across the analysed tools. When idea generation was present, it was typically ad hoc, surface-level, and not designed as a sustained or structured design activity

Five interface types

About twenty of the scanned tools were also analysed on their interaction. The question: Does the way a tool presents itself shape what a designer can actually do with it? Five types kept coming back.

ChatGPT and Claude are initially text-only. The designer has to describe what they want, explain the constraints, and define what a good result looks like, all by typing and adding documents (PDF images, sheets). There are no buttons or menus to fall back on. Midjourney and DeeVid AI add sliders and parameter fields on top of the prompt. That helps, but only if you already know which settings matter.

Vizcom and Runway accept images as input, which opens up visual workflows. But they push toward producing a finished output rather than supporting open-ended exploration. Workflow tools like N8N and Zapier bury AI inside automated chains of steps. Smoother to use, but the designer loses sight of what the AI is actually doing behind the scenes.

Then there are template tools like Canva and NotebookLM. They make the first step easy and keep outputs consistent. The trade-off is that they cap what you can express. If your question does not fit the template, the tool cannot help.

None of these types handles everything well. The ones optimised for producing polished output are good at Semantic Transformation but poor at Divergent Exploration.

Takeaways

1. Most of the available tools arrive after the design thinking is done. They help with formatting, summarising, and documenting. They do not help during the exploration itself.

2. Divergent Exploration barely shows up in the current market. It is the capability the First Diamond needs most, and the one designers have least access to.

3. Which interface a designer uses constrains what they can do with AI. No single type covers all capabilities. Picking a tool means picking a set of limitations.

2.4 The Interaction Problem

People who have worked with AI recognise this. The output is almost right. You try to fix it, but there is no obvious way to do it. You rephrase the prompt a few times. The output gets different but not better. Eventually, you close the tab. The interaction with AI has barriers, and improving prompting does not fix it.

The invisible interface

With conventional software, the interface tells you what is possible. Buttons, menus, and controls externalise a system's capabilities: you can see what you can do. Generative AI collapses this entirely. The interface has no buttons, and everything the system can do is invisible (see Figure 2.5).

This places the full responsibility for directing system behaviour on the esigner, expressed through language alone, without any of the visual cues that conventional tools use to guide interaction. Most designers are not trained for this (Nielsen Norman Group, 2023), and the tools themselves offer no help.

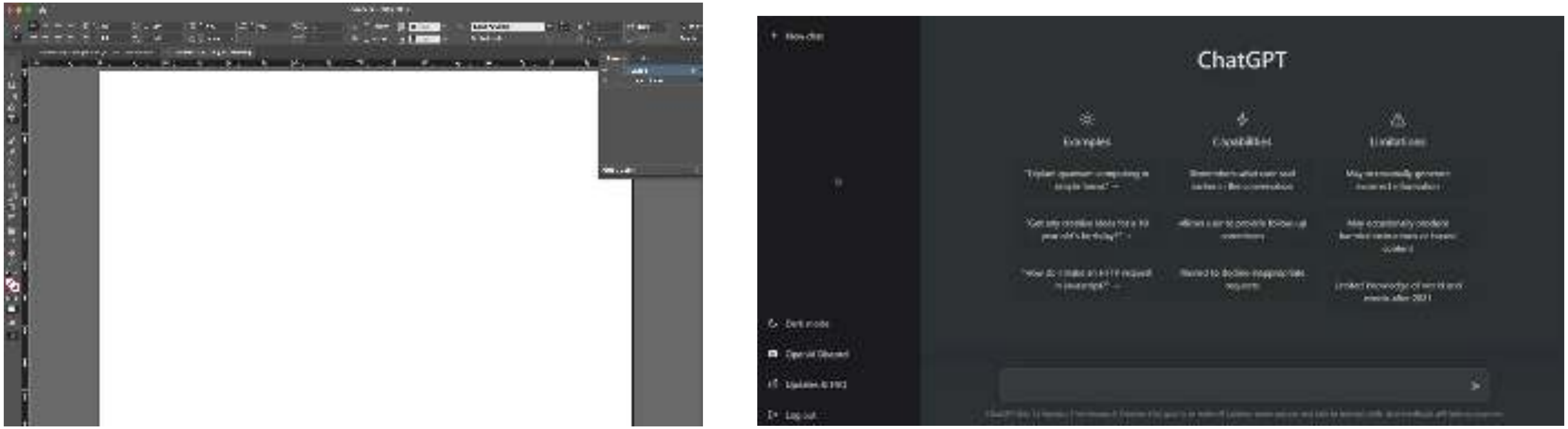


Figure 2.5 — Adobe InDesign (left) and ChatGPT (right). InDesign externalises system capabilities through visible controls; ChatGPT presents an undifferentiated text input regardless of what the model can do.

Framework by Abowd and Beale

This research analyzes what occurs when a designer collaborates with AI, using the Interaction Framework by Abowd and Beale (1991). Their original model describes a continuous cycle in which a User articulates Input, the System processes it, the Interface presents the result, and Output returns to the user. For generative AI, the Interface layer is no longer informative: a prompt box reveals nothing about model capabilities or constraints.

The adapted framework replaces the Interface layer with a Capability layer, asking: "What cognitive tasks can this system support, and under what conditions?" Figure 2.6 shows the original Abowd and Beale model. Figure 2.7 shows the adapted Capability layer developed in this thesis. Interface System User

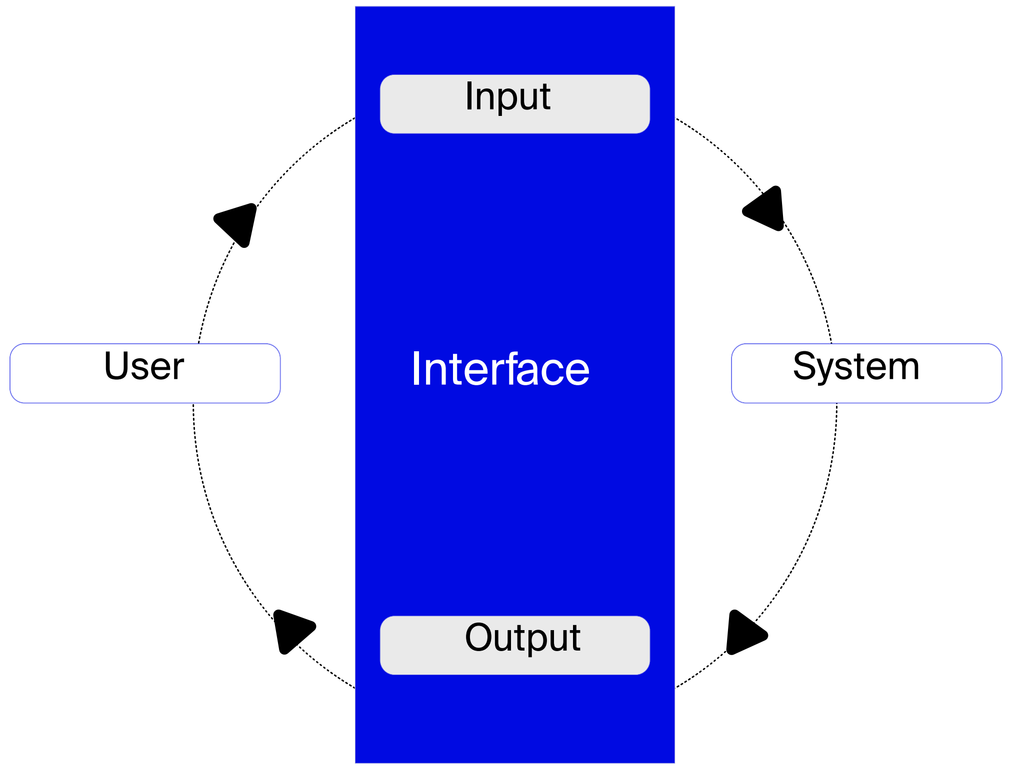


Figure 2.6 – Simplified representation of Abowd and Beale's (1991) Interaction Framework.

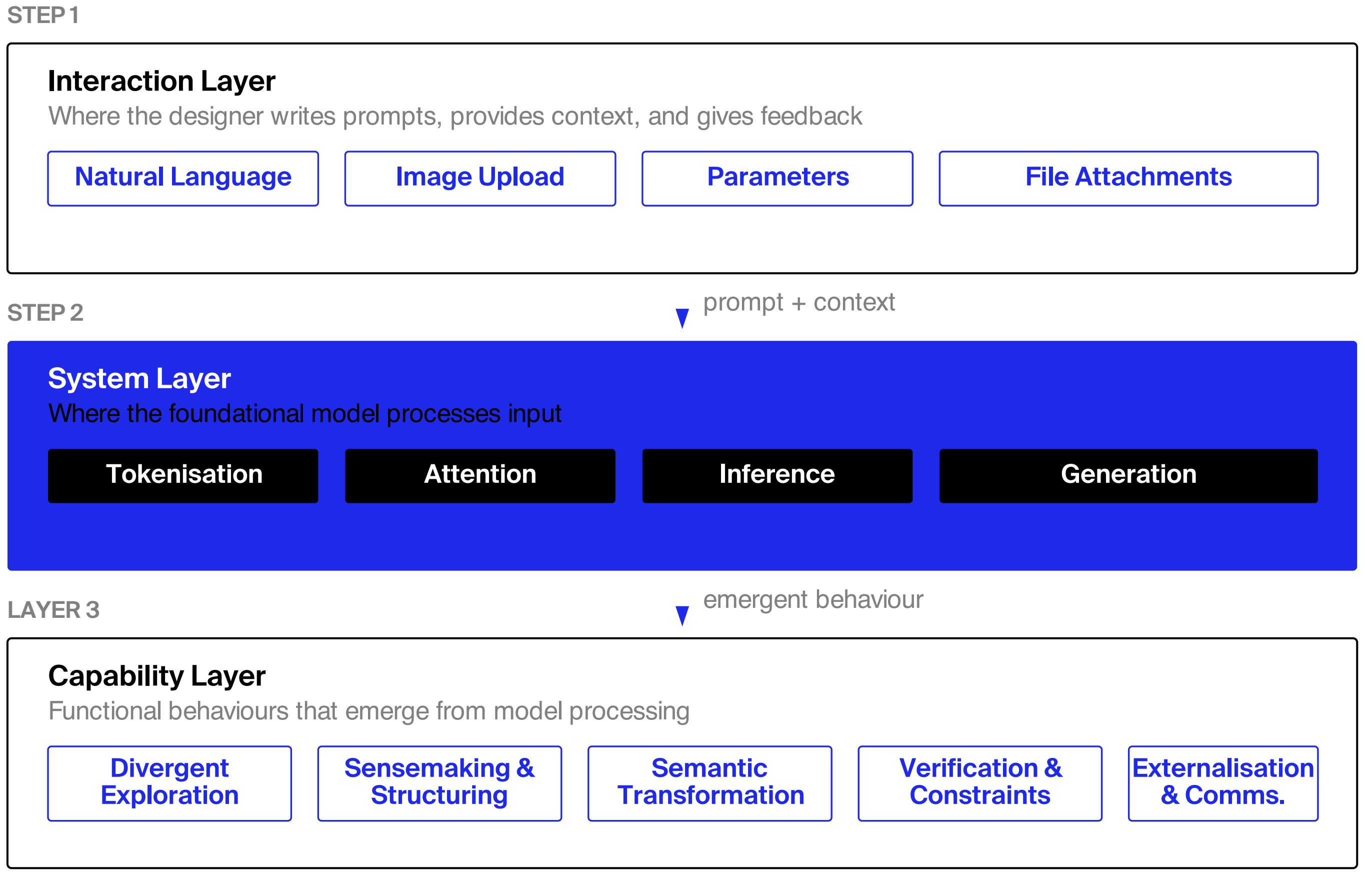


Figure 2.7 – Adapted Capability Framework based on Abowd & Beale (1991). Interface replaced by Capability.

The Gulf of Envisioning: three gaps
The Gulf of Envisioning describes the cognitive distance between what a user can imagine and what they can actually produce through unstructured interaction with an AI system (Subramonyam et al, 2024). Three gaps occur simultaneously: the Capability gap, the Instruction gap, and the Intentionality Gap (Figure 2.8).

Together, these three gaps explain why a designer can know exactly what they need, believe they have asked for it clearly, receive output that looks plausible, and still walk away with something that misses the point. The gulf is important in light of Section 2.2: the technology now supports agents, subagents, persistent memory, and tool access, but none of that matters if the designer cannot bridge the gap between what they need and what they ask for.

Gap	Definition	Examples
Capability Gap	Users cannot accurately model what the system can and cannot do	“I don’t know what to ask for”
Instruction Gap	Users cannot translate a goal into a prompt that reliably produces the intended output	“I know what I want but can’t explain it”
Intentionality Gap	Users cannot evaluate whether the output genuinely met their underlying need	“I’m not sure if this answer is actually right”

Figure 2.8 -The Gulf of Envisioning: three cognitive gaps in unstructured LLM interaction (Subramonyam et al., 2024).

The Prompt Canvas: a structured response
Hewing and Leinhos (2024) respond to this condition directly. Through a systematic review of 718 sources, they synthesise fragmented prompt engineering knowledge into a single practitioner-facing framework: the Prompt Canvas. It externalises the tacit knowledge of effective prompting into eight defined dimensions: Persona, Audience, Task and Intent, Step-by-Step, Context, References, Output format, and Tonality. Together, these dimensions address all three envisioning gaps: giving the designer language for what they want, structure for how to express it, and criteria for evaluating whether the output is right.

The Prompt Canvas works for individual, text-to-text prompt construction. Its authors are explicit about this boundary: agent-based and workflow-level interaction are a different paradigm, deliberately set outside the scope. In December 2024, when it was published, this was a reasonable constraint. By 2026, it had become the most important gap in the available tooling. As Section 2.2 showed, AI systems now operate with persistent agents, subagents, MCP tool access, and Skills, none of which the Prompt Canvas Hewing and Leinhos addresses.

The Prompt Canvas solves one problem precisely. But what counts as “input” to an AI system has expanded far beyond the single prompt, and the design challenge has changed with it.

Takeaways

1. Invisible interface. Generative AI collapses the affordance layer. Every capability has to be discovered through prompting.

2. Three gaps. The Gulf of Envisioning (Subramonyam et al., 2024) names three: Capability, Instruction, and Intentionality.

3. Prompt Canvas. A structured response to those gaps (Hewing & Leinhos, 2024). It is scoped to single-prompt, text-to-text use. agents, orchestrating agents are not included.

Three Layers of Interaction

In practice, all of these components come together in three layers. These three layers are not a new technology. They are a way of organising existing and emerging AI capabilities into a progression that reflects how interaction complexity increases, from a single conversation to a fully configured workflow.

Layer 1 is a conversation. The designer types something, the model responds, the designer refines. Nothing is saved between sessions. Nothing is pre-configured. This is how most people use ChatGPT, and it is how most designers at Collins currently work with AI. It is also the only layer for which the Prompt Canvas (Section 2.4) was designed.

Layer 2 is the configured agent. Here, the designer has set up an agent ahead of time with specific Skills and MCP connections. The agent takes a task, runs through it on its own, and comes back with results. A Layer 2 agent does not need to be rebuilt for each new project. The setup happens once. The returns come back across multiple uses.

Layer 3 is an orchestrated workflow. Multiple subagents, each with its own configuration, handle different parts of a larger task. Some run one after another. Some run at the same time. The output from one becomes the input for the next. This is the most powerful mode but also the most technically demanding to build.

2.5 Chapter Takeaways

1. AI models produces different results depending on how the input is set up. Which role it plays, which documents it can see, and whether it has access to external tools all change the output. Getting the context right matters more than getting the prompt right

2. AI systems in 2026 are assembled from specific parts: agents that work toward goals, subagents that handle isolated tasks, MCP connections to external tools, and Skills that store working methods. A framework for designers needs to account for all of these.

3. Current tools mostly help once the thinking is done. Formatting, summarising, documenting. The capability designers need most; Divergent Exploration, is the one the market covers least.

4. When designers prompt without structure, the same three things go wrong. They do not know what the model is capable of. They cannot get what they need into words that produce the right output. And once the output is there, they have no way to check if it actually answered the question. Subramonyam et al. (2024) mention that the gap is rather cognitive is than skill based. Experience might help reduce friction, but trial and error alone does not close them. Structure does.

5. Hewing and Leinhos (2024) built exactly that for single prompts: the Prompt Canvas. Eight dimensions that make the prompting process visible and repeatable. It works well for text-in, text-out. It does not cover agents, multi-step workflows, or any of the system-level configuration described in Section 2.2.

6. The core challenge for designers has moved. Writing a good prompt is no longer enough. Designing the full environment the model operates in determines output quality. Most designers cannot do that on their own yet. A framework needs to make it possible without turning designers into engineers.

CHAPTER 3

Design Practice at Collins Aerospace

Cycle 1 findings: the workflow, its five phases, the First Diamond challenges, and the post-midterm client alignment. This chapter documents how design work actually unfolds at Collins Aerospace, identifying eight recurring challenges rooted in sensemaking under pressure.

3. Design Practice at Collins Aerospace

The workflow, its five phases, the First Diamond challenges, and the criteria for any intervention

This chapter explains the context at Collins and builds a grounded understanding of what designing at Collins Aerospace entails.

This chapter has been built over approximately four to five months through ten to fifteen conversations with three members of Customer Experience & Design (CXD). The following designers from the CXD team were interviewed: P1 (junior designer), P2 (senior designer and design manager), and P3 (programme manager). These conversations were supplemented by an analysis of internal workflow documents and earlier project materials, which were reviewed together with the fellow graduate students. All contact was remote, as the team is distributed across Seattle and Ireland. To identify patterns, a four-hour context-mapping session was conducted with fellow master students, during which all collected materials were clustered on a physical whiteboard using sticky notes (Figure 3.1).

Gaining access to certain (proprietary) data was a slow process. Many internal documents could not be shared due to the proprietary regulations. Many gaps had to be filled through inference and repeated questioning across separate conversations. This constraint shaped the research throughout and reflects the reality of conducting design research within a regulated organisation (remotely)

.

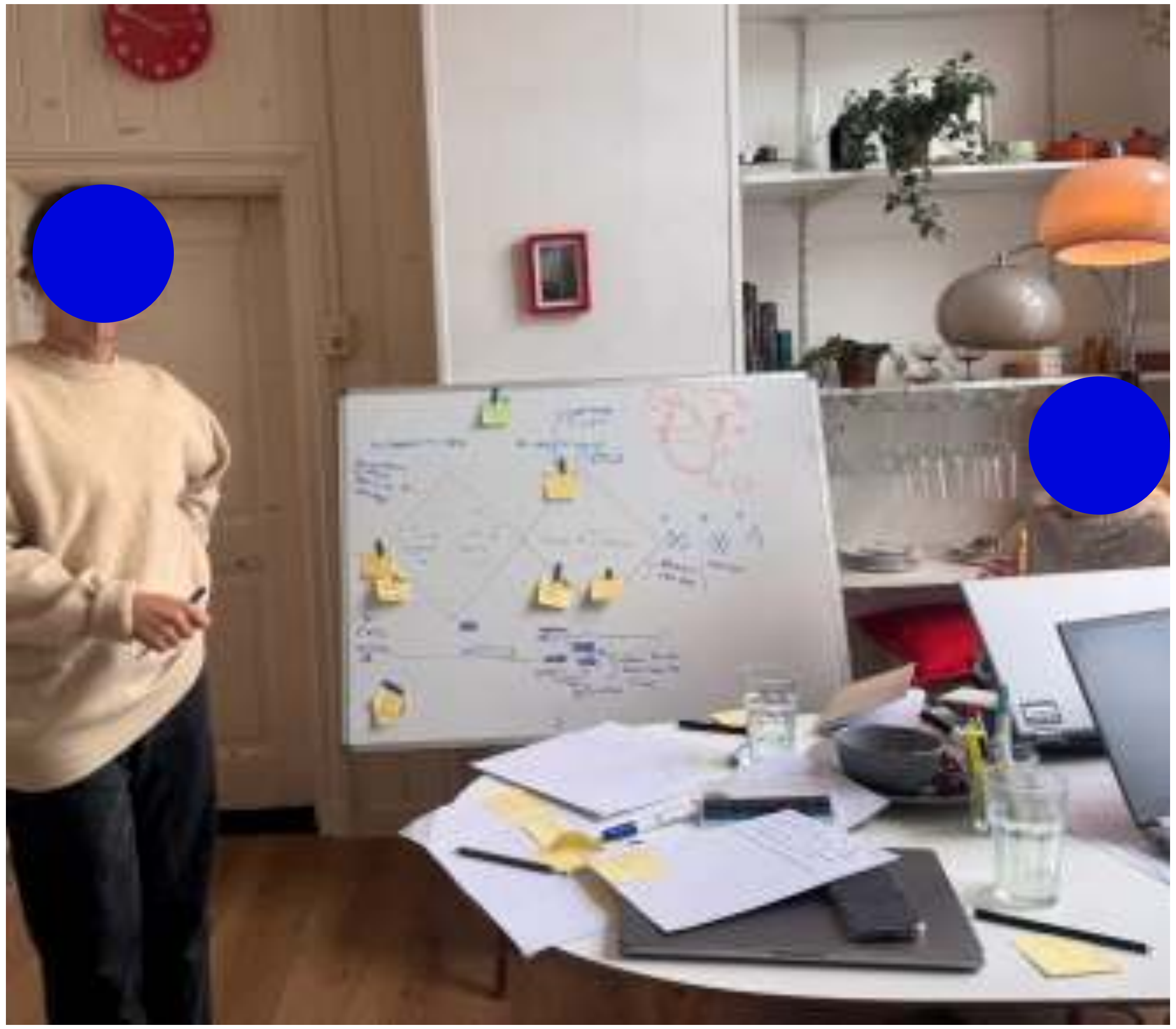


Figure 3.1. Left: context-mapping session conducted with fellow master students to analyse challenges across the Double Diamond. Right: assumptions visualised and used as interview prompts with CXD team members.

3.1 The Aviation Market Context

Collins Aerospace operates in a B2B aviation market, serving airlines and OEMs (Original Equipment Manufacturers) such as Boeing and Airbus. The CXD team gets projects through two procurement channels (Figure 3.2). In the Supplier-Furnished Equipment (SFE) channel, the OEM is the primary client. The designer must follow their specifications. In the Buyer-Furnished Equipment (BFE) channel, the airline leads the design specifications.

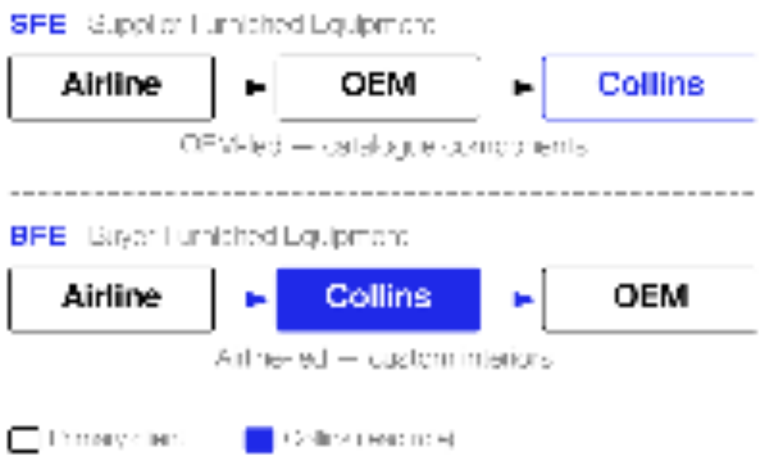


Figure 3.2 – SFE and BFE procurement pathways in commercial aviation.

Before the project reaches the CXD team, it passes through Business Development, which evaluates briefs, assesses strategic fit, and hands the project over once scope is approved (Figure 3.3). This market structure means designers must simultaneously satisfy the airlines OEMs, or the internal business leadership.

Each speaks a different language and evaluates work through a different lens. Carlile (2004) describes this as a knowledge boundary problem: actors from different disciplines do not simply transfer information. They must translate it across fundamentally different frameworks of relevance.

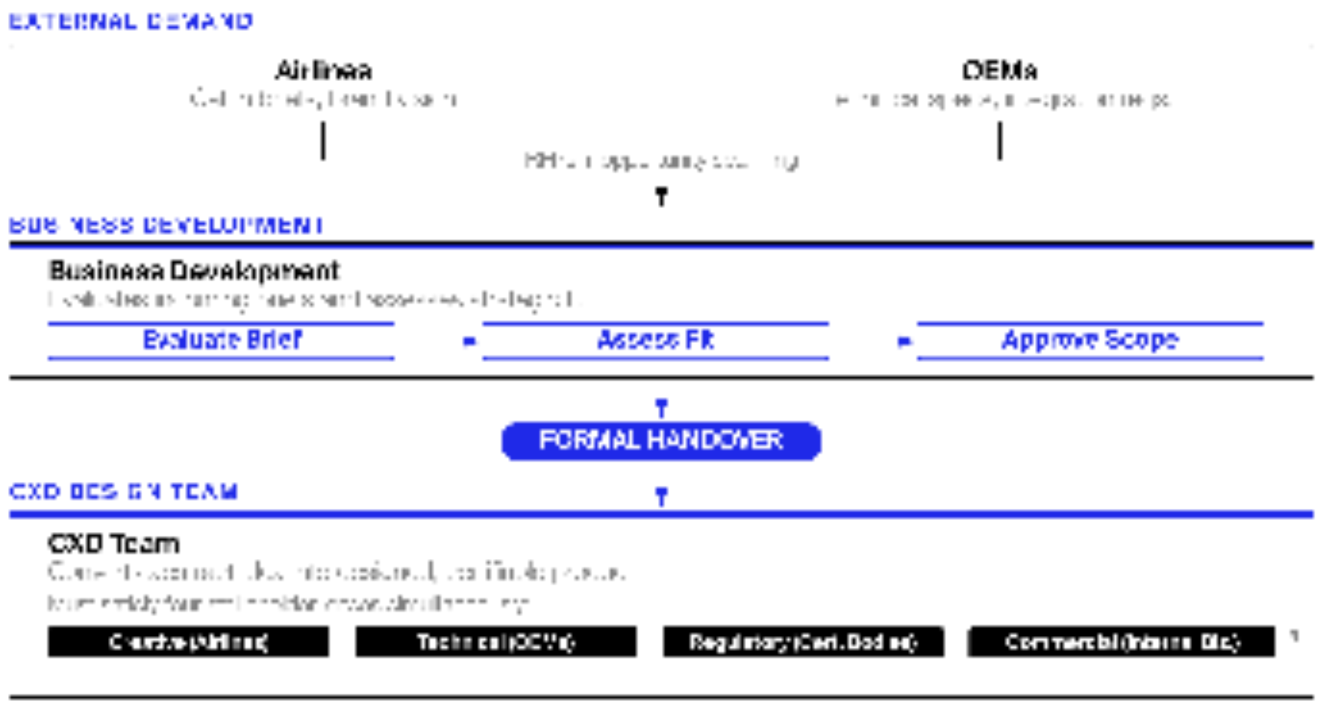


Figure 3.3 – Internal communication process from market demand to CXD project assignment.

3.2 The Five-Phase Development Process

Projects at Collins are structured into five phases, from abstract idea to a flying, certifiable product (Figure 3.4). Each phase has a defined goal, gate review, and formal output. The CXD is most active in Phases 1 and 2, the First Diamond, where the design team sets the creative direction. Phases 3–5 progressively increase the engineering constraints, moving from concept to production and certification of the end product.

Phase 1 (Question to Idea) is the most open and exploratory part of the double diamond: trend research, competitive benchmarking, passenger analysis, and stakeholder conversations all happen in the first diamond. Part 2 shifts to synthesis: translating insights into a Design Brief that must satisfy different stakeholders. e.g., a design lead and a client simultaneously. Phase 3 evaluates the concept based on the criteria of engineers and the CWC hierarchy (Cost, Weight, Comfort/ Differentiation). The CXD team remains involved to prevent what the team calls ‘losing design intent, ‘ pragmatic engineering adjustments that individually make sense but collectively undermine the design vision. Phases 4–5 translate the concept into a manufacturable product, making it ready for certification.

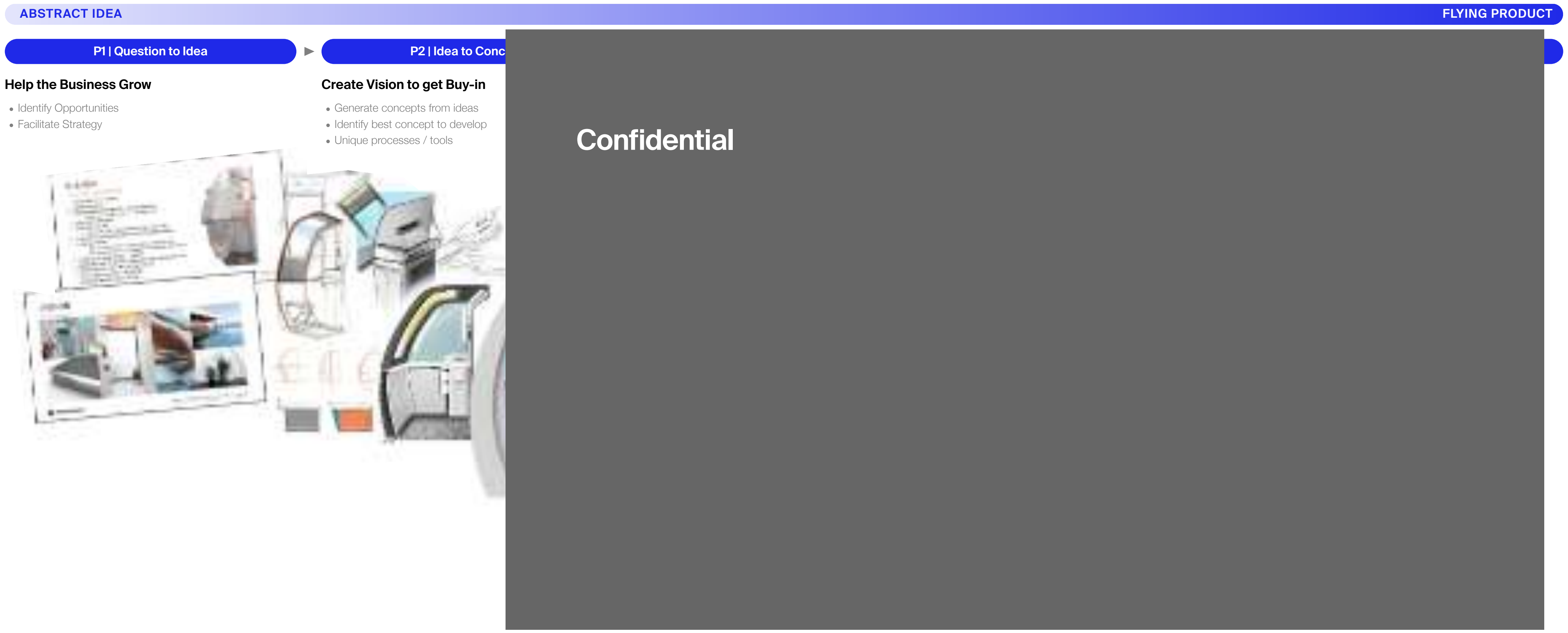


Figure 3.4 – The five-phase development process at Collins Aerospace, from abstract idea to flying product. Visual outputs progress from mood boards and sketches (P1–P2) to engineering models and installed cabin products (P3–P5).

3.3 The First Diamond

As Figure 3.4 shows, the CXD team is most involved in Phases 1 and 2. In these phases, the team works according to the Double Diamond model (Design Council, 2024): they move from divergent (exploration) to convergent (synthesis) (Figure 3.5). Phase 1 (Question to Idea) aligns with the First Diamond, covering both the Discover and Define stages, while Phase 2 (Idea to Concept) aligns with the Second Diamond. Figure 3.5 outlines the activities, gate outputs, and how the Double Diamond maps onto the CXD workflow.

For the remainder of this thesis, the primary focus is on Phase 1, where the design direction is established. Phase 1 contains both the Discover stage (opening up the problem space) and the Define stage (narrowing toward a committed direction. In the following section, we look at the phases in detail.

Discover

The Discover stage is about building a broad understanding before committing to any direction. Designers do trend research, benchmarking, product audits of existing Collins products, flyer forum analysis, stakeholder conversations, and requirements gathering. Each stakeholder group (airlines, OEMs, certification authorities, Business Development) brings different knowledge. None of it arrives in a format directly usable for design. The designer has to extract, translate, and synthesise across what Carlile (2004) calls pragmatic knowledge boundaries. This is cognitively demanding: managing loosely structured information while resisting premature convergence (Cross, 2004).

Figure 3.5 — Phases 1 and 2 at Collins Aerospace: goals, activities, gate outputs, and the Double Diamond structure underlying the First Diamond.

The main tools are MS Teams Whiteboard and MS PowerPoint. All three team members named these without being prompted. No one mentioned a purpose built research synthesis tool. Both tools force tidiness on information that is not yet tidy. That creates extra formatting work and premature convergence

Define

The Define stage shifts to convergence: synthesising the research into a direction that can pass the gate review and allowing us to make a final direction visible to the client. The gate output includes a Statement of Work, research findings, recommendations and a vision. In this part of the workflow, the distinctive friction is the pressure to converge early. The CXD designers explain that they are forced to choose one direction despite having doubts, since they cannot explore alternative options.

> 'We spend as much time formatting the deck as we spend thinking about the design. And by the time the deck is done, there is no time left to question whether we are going in the right direction.'
>
> — Collins CXD designer

3.4 The First Diamond in Practice

The Double Diamond model (Design Council, 2024) describes design as alternating divergence and convergence (Figure 3.6). At Collins, Phases 1 and 2 constitute the 'First Diamond': the arc from initial research to committed design direction (Figure 3.7). In practice, this arc is far messier than the diagram suggests. Designers move through rapid cycles of investigation, interpretation, and reframing (Dorst & Cross, 2001), amplified at Collins by continuously arriving information, shifting stakeholder priorities, and phase gate deadlines that impose convergence before sufficient exploration (Figure 3.7).

There is also a sales-cycle pressure in the BFE pathway, where the airline leads the brief. Airlines send RFPs to multiple suppliers at once, and the team has about a week to respond. That is enough time for a few hours of research and a day of sketching, and not much more. The team is racing to win the work before a competitor does. On top of that, phase gate deadlines force convergence before the team has explored enough (Figure 3.7). The real process is faster and more constrained than the diagram shows.

In literature, the first part of the Double Diamond is also called the Fuzzy Front End (Figure3.8_. Sanders and Stappers (2008) describe the front end as the activities that explore open-ended questions before formal design begins. They call it fuzzy because of its ambiguity and chaotic nature. At this stage, it is often not yet known whether the result will be a product, a service, an interface, or something else. Dorst and Cross (2001) add that, in this phase, problem and solution co-evolve: designers do not start with a fixed brief; they build the brief and the answer at the same time. The goal is to figure out what is worth designing in the first place.

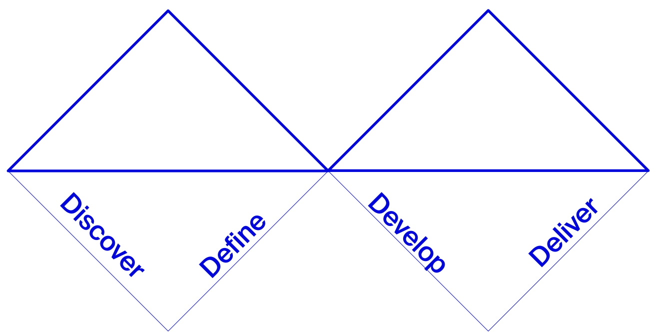


Figure 3.6 — The Double Diamond model showing the four phases of Discover, Define, Develop, and Deliver, with the First Diamond highlighted as the focus of this thesis.

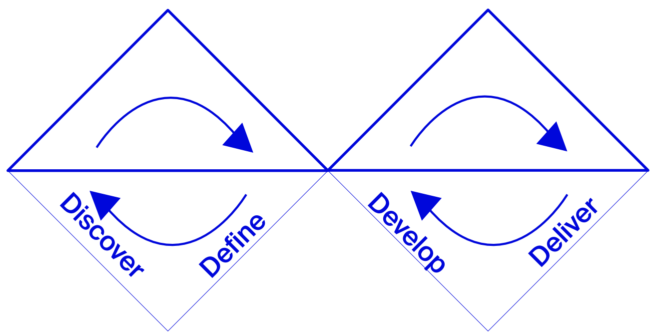


Figure 3.7 — The Double Diamond as it operates in practice at Collins, showing the non-linear, iterative character of early-stage design work as teams cycle through investigation, interpretation, and reframing.

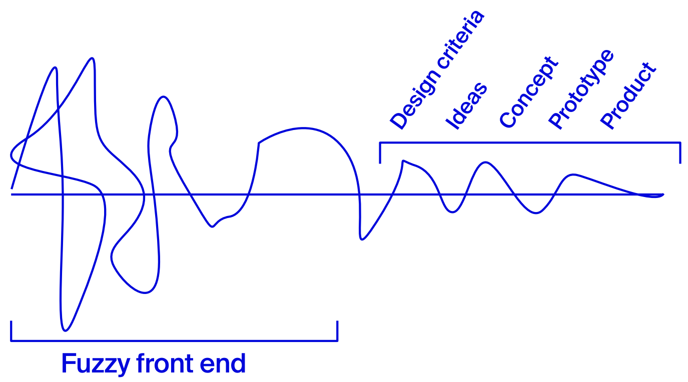


Figure 3.8 — The Fuzzy Front End of the design process, showing the transition from vague context and fragmented inputs toward defined concepts. Adapted from Sanders and Stappers (2008).

3.5 Eight Challenges in the First Diamond

My starting point was Contextual Design (Beyer & Holtzblatt, 1998): workflow mapping, interviews, and looking at what the team produces. The classic version of the method is built on sitting with people while they work. I could not do that. Collins has strict confidentiality rules, and the team is split between Seattle and Ireland. So everything was remote. I conducted interviews, shared my reading with peers, and kept returning to the team as I learned more.

The data came from three places. The brief and the internal docs that the team shared. Interviews with P1, P2, and P3, sometimes together and sometimes one-on-one. Team calls I sat in on. I was unable to record most interviews or take screenshots of the tools (Section 3.1). So I wrote notes and cross-checked the conversation with the interviewees.

Early on I showed what I was seeing to fellow master students. I drew the tensions I thought were there between the people at Collins and the things they make and asked if any of it rang true from their own research (Figure 3.1). Most of it did. They also caught assumptions I had missed. That session cleaned up what I needed to ask next.

During my interview with p1 I went over my finding and asked if it matched. It mostly did. That felt different from the open replies earlier. It was someone finding themselves in a description I had built up elsewhere.

The eight challenges came from accumulating and synthesizing all information. I went through the transcripts, the notes, and the docs, and grouped the patterns that kept showing up. Figure 3.9 places them on the Discover and Define phases.

Figure 3.9 — The eight challenges mapped onto the Discover–Define phases, showing how they compound across the workflow.

Relations between challenges
These challenges form a structure that is systemal. (Figure 3.9). Limited time (C1) amplifies scattered information (C2): no time to organise means inputs remain fragmented. Siloed inputs (C3) combined with communication load (C4) means designers spend time translating rather than synthesising. Time pressure (C1) meeting hidden knowledge (C5) produces premature lock-in (C6): when there is no time and the reasoning is tacit, convergence happens by default. Weak intent transfer (C8) is the downstream result of all upstream challenges compounding.

Reframing Sensemaking
Initially, I thought the problem was information. It was either too scattered, too late, too much, not enough. Some version of that. The eight challenges together point somewhere else. The work that actually matters is sensemaking: building a shared picture of the project within a team with different priorities, under time pressure.

Weick (1995) calls this organisational sensemaking. People take ambiguous situations and build plausible meanings from them so they can act. At Collins, that process runs on a tight clock. The team is remote. Stakeholders hold different priorities. Whatever the designers end up with has to convince people who do not share the designer's frame of reference.

The actual challenge is to start with sensemaking, not communicating. The information is already there. The hard part is turning it into something that convinces other disciplines, under pressure, in time for the next gate.

Takeaways

1. Sensemaking, not information transfer. The eight challenges describe one structural difficulty in eight forms: constructing shared understanding from ambiguous inputs under time pressure (Weick, 1995).

2. Reinforcing system. The challenges compound: time pressure amplifies fragmentation, siloed inputs increase communication load, tacit knowledge under deadline pressure produces premature lock-in.

3.6 Case study

To get a better understanding of what the first phase looks like, I reviewed P1's work. P1 worked alone on a cabin experience brief. On paper, the project had a two-month timeline. In practice, the Discover and Define phase lasted about two weeks. Other projects, client availability, and a fixed gate review took up most of the other obligations (Figure 3.10).

Starting under pressure

The project began with a brief handed down from Business Development. P1 had no involvement in the initial scoping or RFP process. The timeline and deliverables were already fixed.

> 'We basically have two months and the first phase of the double diamond encompasses both research and the early stages of definition ... honestly just like two weeks in between each phase for this kind of two-month phase.'
>
> — P1

Two weeks for Discover and Define together. P2, who manages the team's project intake, confirmed that this is not unusual. RFP-driven projects can compress the entire First Diamond into days::

> 'Sometimes an airline will say, "Hey, we like that you do labs. We have an RFP for a new lab design ... and we need your response within a week."

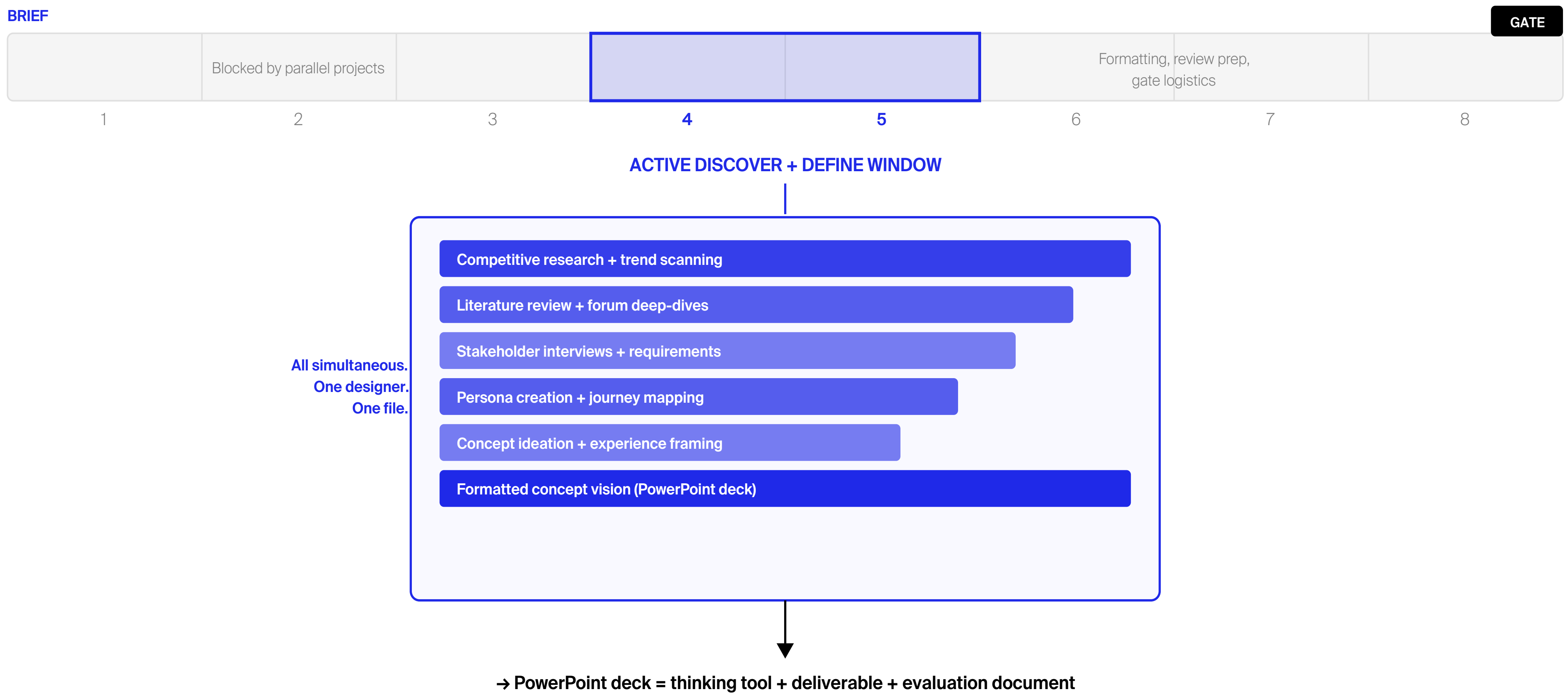


Figure 3.10 — P1's two-month project compressed into a two-week active working window for Discover and Define.

> That gives us probably ten minutes to figure out what research we want to do, about four hours of research, maybe a day of sketching, and then we just jump right into CAD.'
>
> — P2

This is Challenge 1, Limited Exploration Time, and it sets the beginning for the other challenges

Researching alone

Within that two-week window, P1 was solely responsible for competitive research, trend scanning, literature review, requirements gathering, and a fully formatted concept vision. There was no centralised research repository. The designer can go back to previous projects, but these are not always stored in one place, and materials get deleted in the team's environment due to company policy.

> 'A lot of the research is just purely manual — going through vlogs, reading up scientific articles, basically doing a deep dive of the interwebs myself, and it can be very tedious and time-consuming.'
>
> — P1

Because there was no time for firsthand interviews, P1 had to source passenger insights from earlier, unrelated projects or online vlogs:

> With this truncated timeline we didn't have time to do firsthand interviews, so I had to go and source some interviews from passengers with disabilities from a much earlier project.'
>
> — P1

Inputs distributed across vlogs, Reddit forums, scientific articles, old project files, and personal notes. None in a shared system, all reconstructed manually. This is Challenge 2: Scattered Information.

Working across silos
The brief required input from engineering, regulatory, market, and client stakeholders. Each operates independently. When asked about cross-functional collaboration, P1 stated it flatly:

> 'The team operates in silos.'
>
> — P1

P2 shared this perspective: the CXD team's involvement across business units is 'very inconsistent', sometimes they join in at the start of a phase, sometimes only to 'clean up the tail end'. Stakeholder inputs arrive at different times, in different formats, through different channels. The designer becomes the sole integration point for knowledge that no one else is combining.

Building on assumptions
Without firsthand data, P1's research was built on indirect sources: online forums, vlogs, and fragments from older projects. The personas were a fake representation of potential passengers:

> 'Personas are made up, but it's sort of an aggregate of online forums like Reddit, vlogs of people ... So there are assumptions being made, but it is based on real-world experience.'
>
> — P1

P1 was aware that the evidence base was thin, but the timeline left no alternative. Translating this into something stakeholders would accept as rigorous was itself a challenge:

> 'The challenge is taking all these anecdotal pieces of data and translating them into more objective visualisations for people to understand like, why this matters.'
>
> — P1

This is not unique to P1's project. P3, who oversees programme-level decisions, framed the evidence problem from the business side:

> 'Business now is risk-averse and they want to understand: how do you know that for sure? Can you give me some point of evidence?'
>
> — P3

The data is not wrong, but it is not the kind of data the organisation recognises as evidence. The gap between design intuition and proof is where decisions stall or get overruled.

Synthesising into the deck
All of P1's research was synthesised directly into a single PowerPoint deck. This was the same document that would serve as the gate deliverable. There was no separate thinking space.

> 'It's usually pretty low tech. I collect all of my sources and then make a giant research deck and then eventually I end up taking out the best bits, synthesising that into a more condensed version for internal share-out.'
>
> — P1

In Star and Griesemer's (1989) definition, this deck functions as a boundary object: simultaneously a thinking tool, a coordination artefact, and an evaluation document. The design reasoning for why certain directions were pursued, which alternatives were discarded, and why lives in the designer's head, not in the artefact. When the deck is handed over, that reasoning disappears. P2 named the core difficulty:

> 'Design is hard to quantify because it is oftentimes subjective. Emotion in design is not a utilised term within a business structure — because it's hard to quantify.'
>
> — P2

The knowledge that matters most, aesthetic judgement, spatial intuition, the tacit sense that one direction is stronger than another, is precisely the knowledge that resists externalisation.

Justifying the work
As the project moved from Discover into Define, a growing share of P1's time shifted from research to communication. The same findings had to be reframed for different audiences. Engineering, programme management, and the client each require a different register.

> 'It's a lot of framing for why this matters. Preemptively answering the

> questions that our stakeholders will ask, because sometimes our value-stream leaders don't actually know why we want to do something ... They're just funding it, so we have to justify over and over again why we're doing it.'
>
> — P1

P3 summarised what all this communication ultimately reduces to:

> 'At the end of the day, it is just: here is a design and this is why it is important.'
>
> — P3

But the time spent reformatting to reach that simplicity was time not spent thinking. By the time the deck was ready for each audience, there was no time left to question whether the direction was right direction.

Committing before ready
The gate review date was fixed. As the two-week window closed, P1 had to commit to a direction. Because the deadline was approaching. By default, alternatives that had not been fully explored were discarded. P2 described the broader pattern:

> 'It's occurred multiple times where we say it's not ready, but we have to meet the timeline. We continue down that road and then we have to redirect later on ... And so that is the urgency to get it right the first time.'
>
> — P2

The phase gate does not ask whether exploration was sufficient. It asks whether the deadline was met.

After the handover
Once the concept passed the gate, it moved into Phase 3. Engineering department starts making changes when they see a technical difficulty. Each change would be defensible on its own. But over time the original design vision is reduced. The hard part is not the research itself; it is getting the reasoning behind design decisions to survive the handover. P2 has watched the same pattern over the last few years:

> 'When we're not involved, the line drops. Often we're communicated to a challenge after they've already modified, versus we're at the table to collaborate.'
>
> — P2

P3 identified what would be needed to hold the line. and why it is so difficult:

> 'What we're trying to do is enable designers to be able to have developed that evidence that sometimes is hard to develop, but without having loads of time and loads of money.'
>
> — P3

The design reasoning is not fully captured in the deck; this becomes progressively harder to defend as the concept moves further from its origin. Without all the traceable evidence, every engineering modification is a negotiation; the designer is not equipped to win because they cannot communicate their original intent.

Negative effects
We see that the current workflow does not leave room to document alternative directions or run a synthesis session with colleagues. This gap persists.

Figure 3.11 reconstructs the research deck P1 produced for this project, based on interview data and workflow documentation. It shows the scope of work attempted within the two-week window: background framing, user personas built from indirect sources, user journey mapping, industry benchmarking, synthesis into design guidelines, and an initial opportunity prioritisation. All authored by a single designer and delivered as a single PowerPoint file.

A representative case
What happened to P1 is normal. It is what the whole team deals with. Short timelines. One designer owns the deck. Inputs coming from all over. A deliverable that has to work for engineering, the programme manager, and the client all at once.

P1 knew about ChatGPT but had not used it on this project. Nobody on the team had worked out when to use AI or what good AI output looks like, and there was no time to figure it out

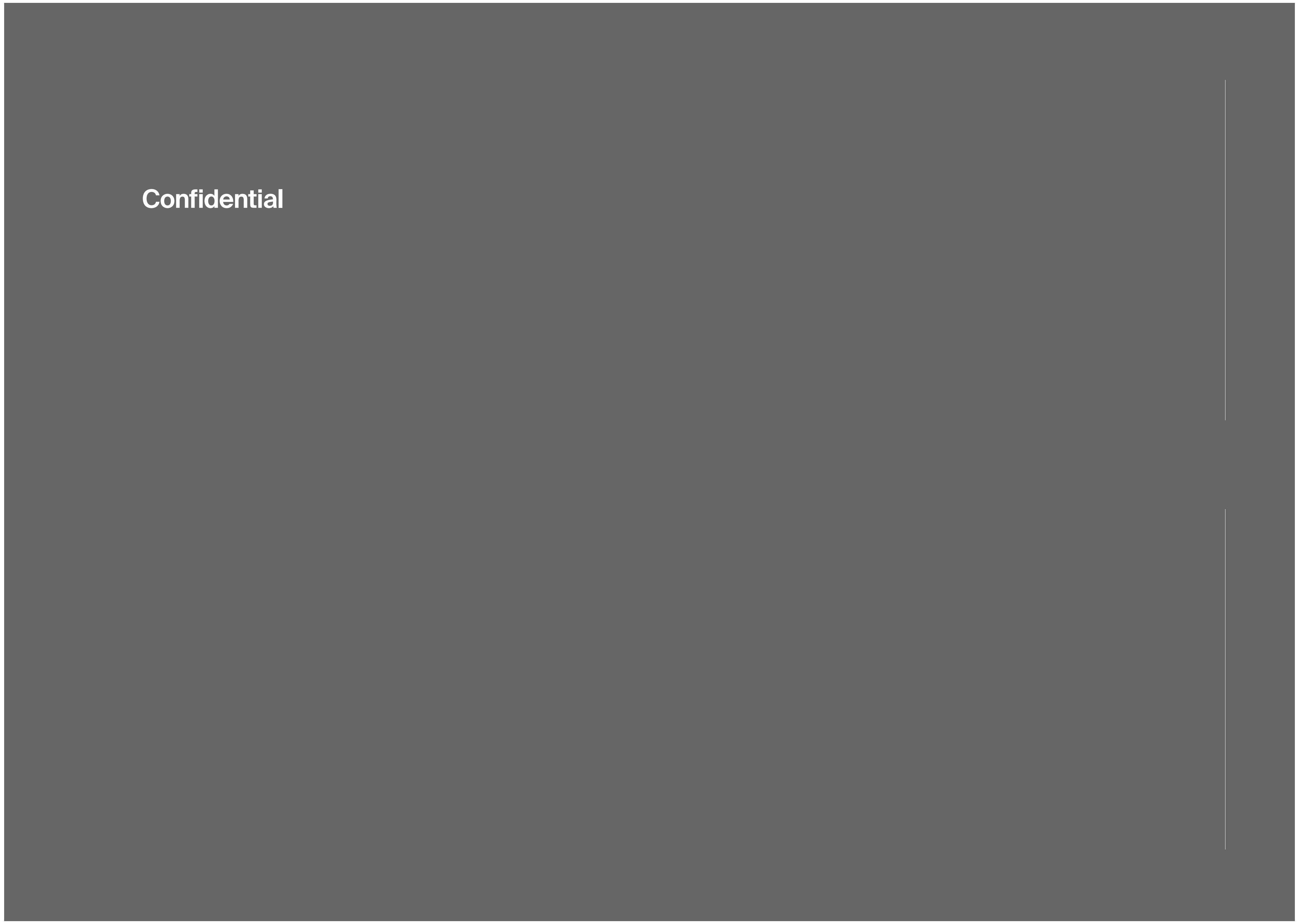


Figure 3.11 — Reconstructed research deck produced by P1: table of contents and milestone timeline showing , user personas, Design guidelines and cross sector research the scope of Discover–Define work attempted in two weeks.

Takeaways

1. All eight challenges in one project. P1's case shows every challenge from Section 3.5. The challenges compound. Time pressure pushes designers into shallow research. Shallow research produces evidence that does evidently convince the design vision to stakeholders

2. The workspace is the deliverable. The PowerPoint deck acts as a boundary object (Star & Griesemer, 1989). Designers think inside it, coordinate through it, and are evaluated on it.

3.Designers do not lack tools. They know AI tools exist. What is missing is a shared way to use them, a standard for what good output looks like, and time to try alternatives.

3.7 Chapter Takeaways

1. The CXD team is mostly active in the First Diamond. Phases 1 and 2 are where the creative direction gets set, with the most ambiguity and the least time.

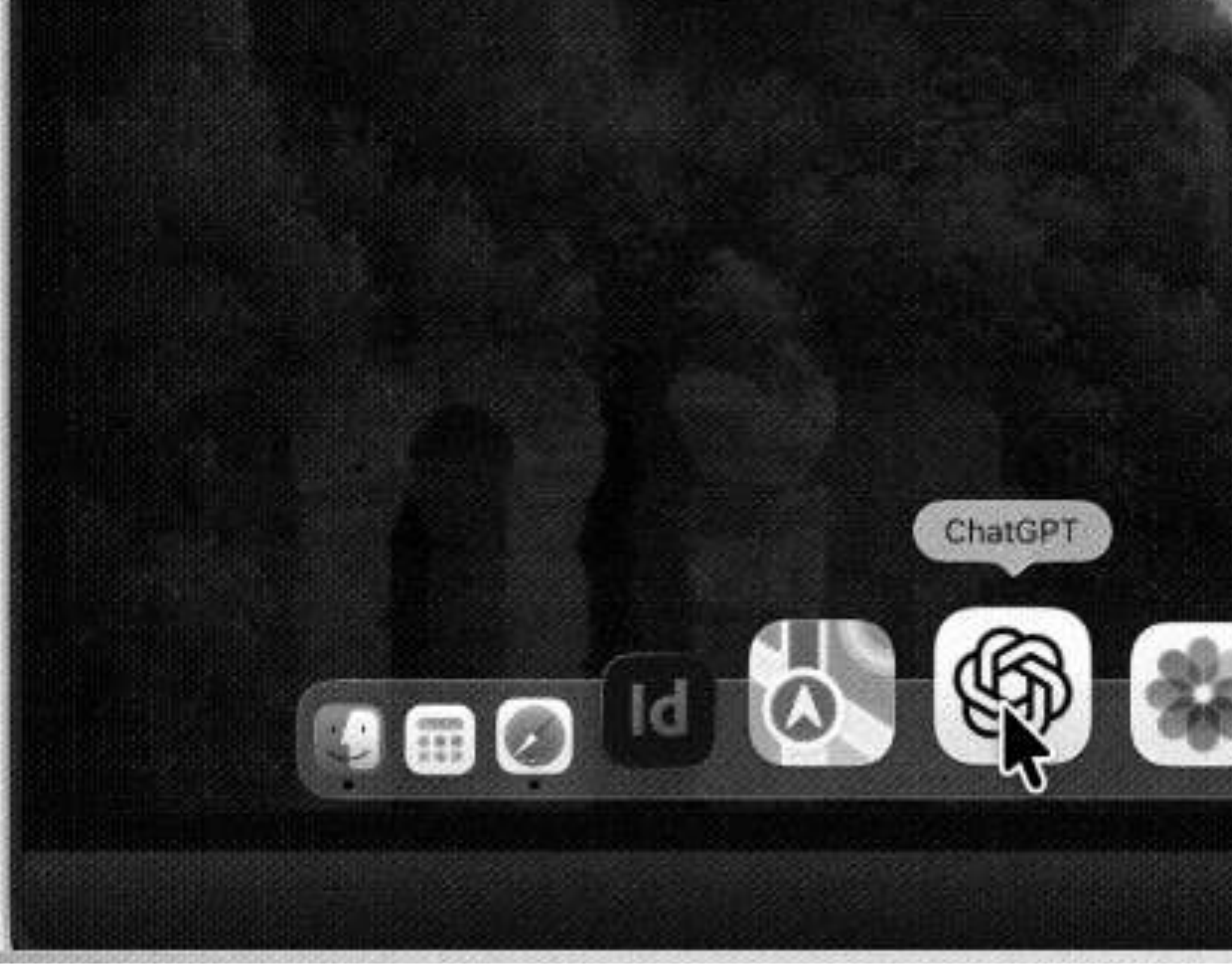
2. The double diamond is in reality messier then the model shows. Phase gates force the CXD team to converge before they have really explored.

3. The eight challenges are related. Altogether, they point to the challenge of sensemaking? building a shared picture of the project before time runs out.

4. The PowerPoint deck does three jobs at once. It is how designers think, how they coordinate, and how they get evaluated.

CHAPTER 4

Opportunities for AI Integration



Cycle 2: mapping challenges to capabilities, probing AI as a design material, and selecting a direction. Eleven probes explored how AI capabilities relate to the eight workflow challenges. The Prompting Canvas was selected through Harris Profile evaluation and client alignment.

4. Opportunities for AI integration

Cycle 2: mapping challenges to capabilities, probing AI as a design material, and selecting a direction

In this chapter, we look at the challenges identified in Chapter 1 and explore how the discovered capabilities can address them. In chapter 2, we established what AI systems can do today and identified an interaction problem: designers face a Gulf of Envisioning (Subramonyam et al., 2024). Here we see that the capability, instruction, and intentionality gaps prevent effective use of AI. A response to these barriers is the Prompt Canvas by Hewing & Leinhos (2024). Chapter 3 grounded the challenges for the CXD designers: sensemaking under limited time, a PowerPoint deck as a boundary object, and a gap (not a necessary tool gap) as the central barrier to AI adoption.

In Cycle 2, we address RQ2: Which AI capabilities are relevant to these problems, and how might they be applied in the design workflow?

Three sub-question guide cycle 2.

1. Which AI capability types are relevant to the Cycle 1 challenges?

2. How do different interaction structures shape designers' sensemaking during early-stage work?

3. And what requirements are relevant for the solution?

4.1 Mapping Challenges to AI Capabilities

The first step in Cycle 2 was to connect the eight challenges identified in Cycle 1 to the five AI capability categories found in Chapter 2: Divergent Exploration, Sensemaking & Structuring, Semantic Transformation, Externalisation & Visualisation, and Verification & Constraints.

This mapping serves two purposes: it creates a common language for design exploration and encourages thinking in terms of capabilities rather than tools. Instead of asking “which AI product fits here?”, the mapping asks “what type of AI function does this challenge need?”

The mapping reveals two important patterns (Figure 4.1). First, most challenges require multiple AI capabilities. Not a single AI function, but combinations of functions working in sequence or in parallel. For example, limited exploration time calls for both Divergent Exploration (generating more directions faster) and Sensemaking & Structuring

Second, the relevance of different capability types shifts across the Double Diamond phases. Divergent

In the diverging stage of the design process, the aim is exploration. What we see here is that capabilities like Exploration and Sensemaking & Structuring are the most relevant. In the define phase, we aim to synthesise the work to communicate this. Then we see that Semantic Transformation and Externalisation become relevant. At last, Verification & Constraints is relevant throughout but becomes especially critical at phase gates, when the design direction must be defensible.

Figure 4.1 – Challenge-to-capability mapping across the Discover-Define phases, showing how capability needs shift over time and how multiple layers are often simultaneously relevant.

This shift across phases matters for the structure of AI support. Most challenges do not map to a single capability. They require combinations, and those combinations change as the work moves from divergence to convergence. A tool built around one function will fit some moments but miss others.

This mapping raised the question of whether existing AI tools already address these stacked capability needs. The scan of 50 AI tools presented in Chapter 2 (Section 2.3) answered this definitively: most of the available AI tools work after the fact, supporting documentation and reformatting rather than active sensemaking. Divergent Exploration and active capabilities, most needed in the First Diamond, were the least consistently supported. No single interface type served all capability categories. This gap could not be closed by selecting a better existing tool; it required a different approach entirely.

Takeaways

1. Challenges need combined capabilities.
Most challenges require two or more AI capabilities.

2. The mix changes across phases.
Divergent Exploration and Sensemaking are most active in Discover. Semantic Transformation and Externalisation become more relevant through Define. Verification & Constraints matter most at phase gates.

3. Existing AI tools miss what matters most.
The tool scan in Chapter 2 showed that most available tools support documentation and reformatting, not the active thinking that comes before it. The capabilities and challenges that need the most are the ones that are least covered.

The gap between what existing AI tools provide and what is needed to support the designers in the First Diamond motivated a shift from analysis to making. The next section describes how AI was explored as a design material through low-fidelity probes.

4.2 Low-Fidelity Probing with AI

The gap confirmed in Section 4.1 cannot be closed by selecting or combining existing tools. It requires a different approach: exploring AI as a design material. In line with the co-evolution of problem and solution spaces (Dorst & Cross, 2001), low-fidelity AI probes are built using Gemini to test whether the mapped capabilities actually solve the right challenges. AI outputs are unpredictable. What a capability does on paper is different from what it does in practice. The only way to find out is to build something and experience the interaction firsthand.

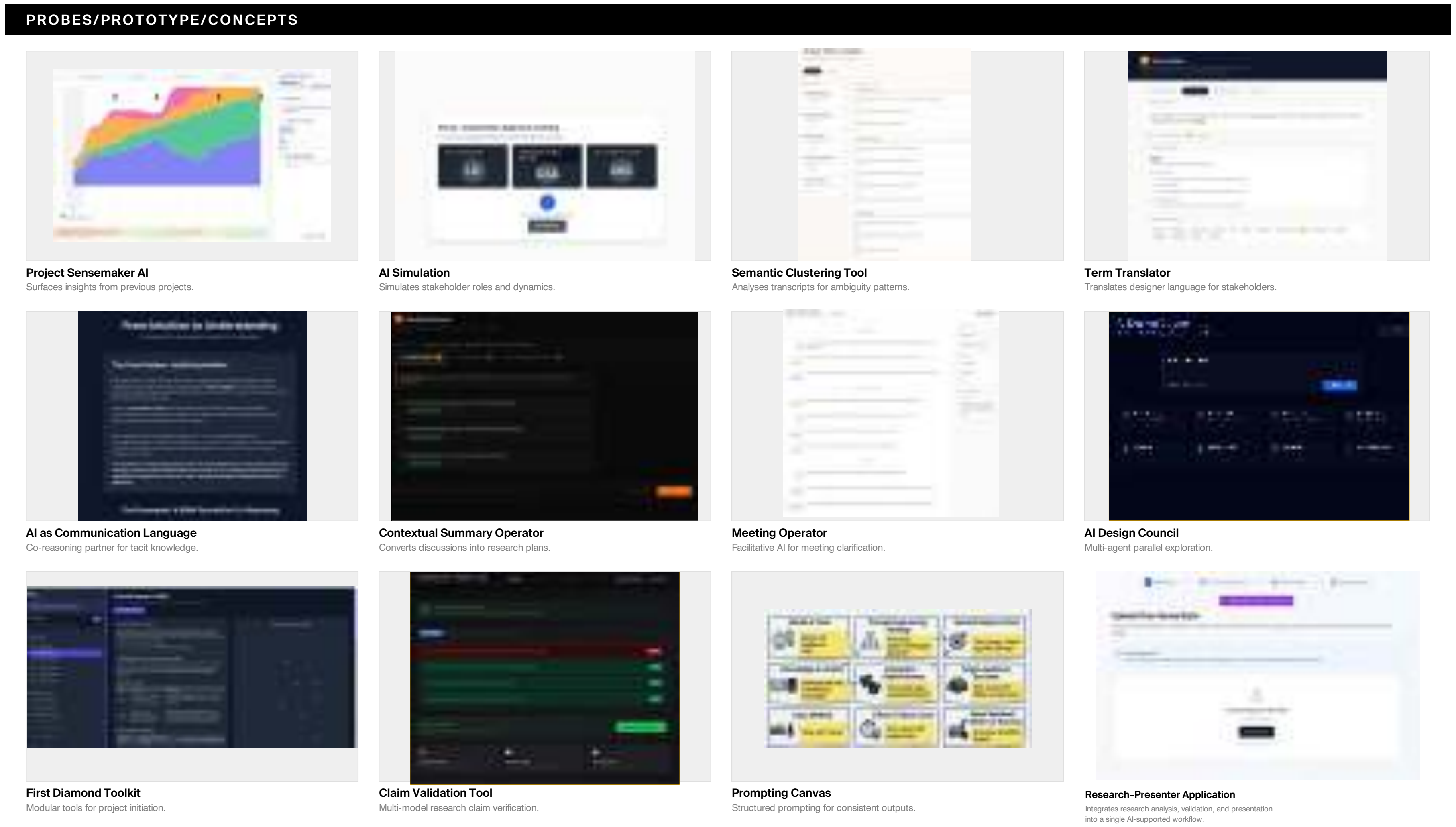


Figure 4.2 — Probes, Prototypes and concepts for research and exploration

Probes

In this context, probes are not tools to be tested for performance. They are instruments for generating design knowledge. Each probe tests a specific question: Does this capability address the challenge it is mapped to? By building a probe, running it with members of the CXD team, and reflecting on what happens, it becomes possible to learn things about AI that analysis alone would miss.

From that experience, both the problem and the proposed solution can be reframed (Dorst, 2011). For example, the challenge in the original brief was framed as “communication with stakeholders”. At that point, what this actually meant was unclear. After a probe session, the real problem becomes clearer. The designers spend a lot of time translating their design choices. They need measurable, quantifiable evidence to make their design choices convincing to stakeholders. Without that evidence, every handover requires extra effort to translate design reasoning into language that stakeholders would accept. This reframing happened through prototyping and evaluation.

Online explorations with CXD team

The CXD team is distributed across Seattle and Ireland. Direct observation is not feasible. Probes are presented and demonstrated during online meetings with team members, who give feedback on whether the probes feel useful for their workflow. This remote format shapes the probes in two ways. It pushes them toward configurations that can be meaningfully demonstrated on screen. And it reinforces the requirement for any eventual framework to be self-explanatory and usable without in-person support.

The probe set

Twelve probes and prototypes are developed across the Discover and Define phases. Each targets a specific capability-challenge combination. The distinction here matters: probes are designed to learn about the interaction, and prototypes are designed to be closer to a possible solution. These two types do not follow a clean sequence. Building a prototype sometimes changes the understanding of the problem, which leads back to a new probe. Problem and solution develop together. The probes vary in interaction type and in the phase of the first diamond they address. Figure 4.2 shows the full set.

Selection and convergence
After the sessions, a few things became clearer. The simulation-based ones (AI Simulation, Meeting Operator) and claim validation add too much overhead for where the designer is in their process. These get deprioritised. Then, two directions were further explored. The Contextual Summary Operator works because it has access to previous projects and raises open questions. The team finds this helpful. The AI Design Council also had an important insight: the real value is not in automating the process, but in cutting down the hours designers spend turning insights into slide decks.

Key insights from the probes
Four insights come out of the probes consistently. They shape the concept directions in Section 4.3 and the agent architecture in Chapter 7.

1. The problem is structure, not tools. The problem is structure, not tools. Designers at Collins use ChatGPT regularly for exploration and synthesis. The tool itself is fine. However, each designer prompts in their own way, based on what they are used to. That works individually. At the team level, it breaks down because one designer's AI summary looks nothing like another's. There is no shared structure for interacting with AI, and without it, the outputs cannot be compared or combined.

The absence of a shared structure for interacting with AI reduced the effectiveness of these tools in collaborative settings.

2. Designers already think in roles for AI, but lack a shared language for it. During the probes session, it became clear that designers talk about AI as if it takes on roles. At times, they want an active collaborator: someone who helps build a research plan or proposes a direction. In others, they want AI to remain passive and respond only when asked. The deciding factor is oversight. When making a design decision, human judgment takes priority—this distinction between active and passive AI roles. The instinct is there. But they have little to no time to explore alternative AI tools. The designers point out that they stick with ChatGPT even when it does not fit. Within the CXD team, the mental models exist. What is missing is a way to name them and put them to use.

3. Splitting tasks across agents beats a single assistant. Building the AI Design Council probe shows that splitting a more complex task into smaller subtasks produces more coherent results than one assistant handling all three. Each subtask needs a different configuration. This finding shaped the design decision for the final architecture

4. AI takes on a certain attitude. When probes put AI in an active role, like running a meeting (Meeting Operator) or interpreting stakeholder reactions (AI Simulation), the CXD team pushes back. They do not see it as useful for early-stage work. What they want is AI that organises and surfaces information quietly or when asked, without making decisions on its own.

The preference was for AI as background infrastructure: structuring and surfacing information without including itself in the decision-making space.

From tools to design material
The probes also change how I think about AI itself. Comparing different AI tools and interaction types makes one thing clear: the same AI capability can lead to different outcomes depending on how the interaction is configured. There are many things behind that prompting interface that influence whether AI supports exploration. In this way, AI is not a tool to be selected. It is a design material to be configured.

This reframing directly resulted in the three concept directions described next: the Prompting Canvas, focusing on structured interaction and consistency; the First Diamond Toolkit, focusing on shared sensemaking and overview; and the Research-Presenter Application, focusing on synthesising and communicating design outcomes.

4.3 Three Concept Directions

These insights point in a clear direction: AI support for the First Diamond must be structural. The next section describes how these insight and prototypes turned into three concept directions. The probe insights within the Collins context (see Figure 4.3). They were developed iteratively, with feedback from the CXD team shaping which directions felt most relevant and realistic.

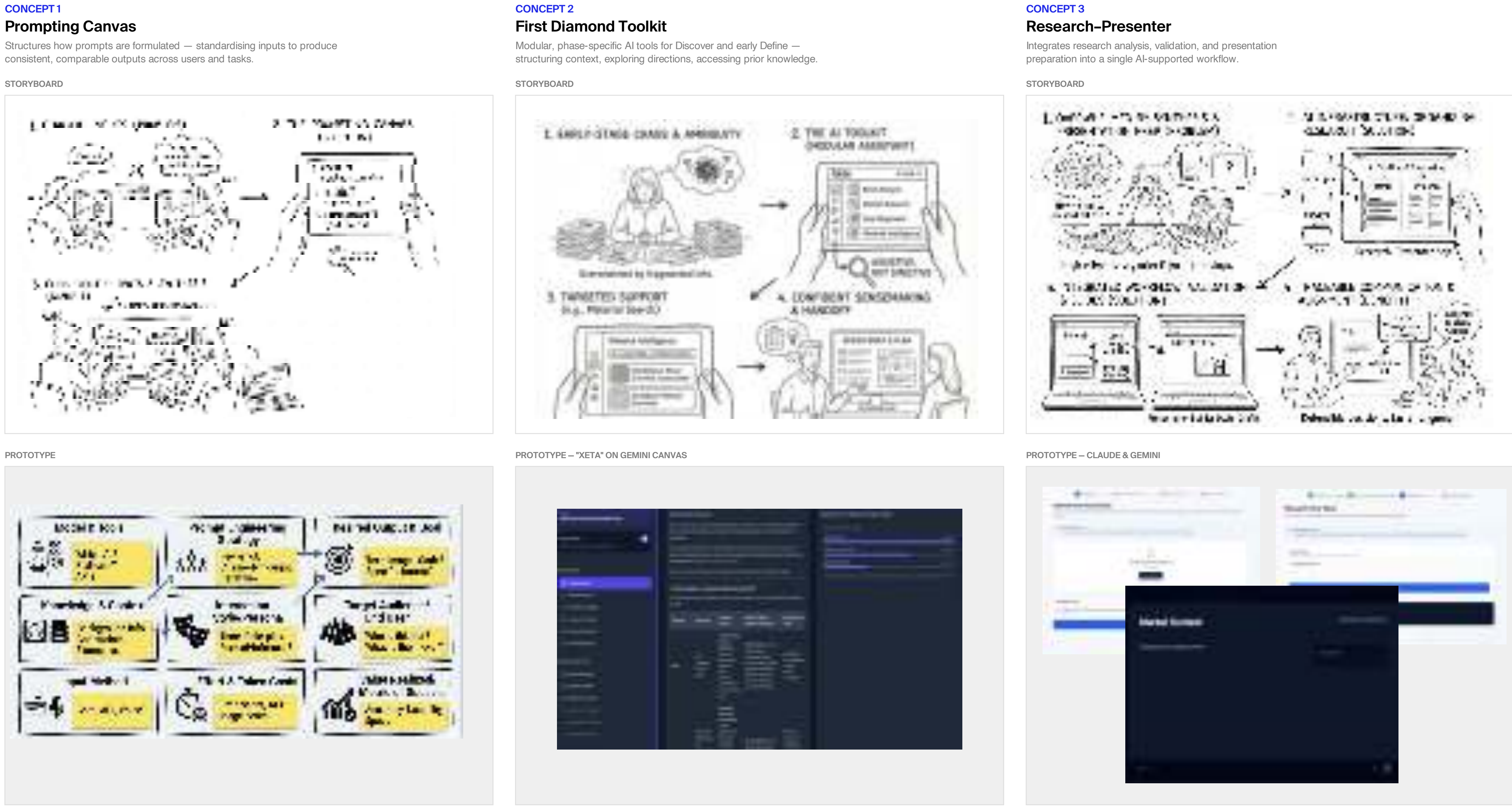


Figure 4.3 — Three concept directions with storyboards and prototype explorations. (a) Prompting Canvas: chaotic individual inputs → canvas-structured interaction → consistent outputs. (b) First Diamond Toolkit: the "Xeta" modular interface tested with Gemini Canvas. (c) Research-Presenter: prototype screenshots from Claude and Gemini.

Concept 1 — The Prompting Canvas

The probes revealed that the CXD team already uses tools like ChatGPT frequently, but each designer prompts differently based on personal preferences. This individualised use creates problems when AI outputs need to be compared or synthesised at a team level: inconsistent inputs produce inconsistent outputs. The Prompting Canvas addresses this by standardising how prompts are structured, drawing inspiration from tools like the Business Model Canvas. Rather than focusing on the AI model itself, the canvas helps designers define the context, intent, and constraints of a prompt before interacting with an LLM. It is intentionally model-agnostic and designed to work alongside existing AI tools. The full canvas architecture and its dimensions are developed in Chapter 5.

Concept 2 — The First Diamond Toolkit

The First Diamond Toolkit is a set of AI tools. The AI tools are tied to a specific moment in the Discover and early Define phases. Instead of a single assistant handling everything, designers can engage with different chats. The brief analysis, goal alignment, and material intelligence are separate AI tools that contain information from previous projects. The AI is assistive, not directive. Designers decide when the AI tool is active and when it stays passive.

The toolkit is built to be shared across the design team and to grow over time as new modules are added. A working prototype ("Xeta") is built and tested using Gemini Canvas.

Concept 3 — The Research-Presenter Application

The Research-Presenter combines research analysis, validation, and presentation preparation into one AI-supported workflow. Designers at Collins spend much of their time turning findings into stakeholder presentations, often under time pressure. This concept handles that in one place, rather than switching between separate tools for research and slide-making.

After testing with Gamma AI and Canva, there was a clear gap; they either generate slides without house-style control or allow styling without research integration. Neither does it. The Research-Presenter is designed to close that gap.

Choosing a direction

The three concepts have tradeoffs and organisational constraints. The Prompting Canvas has a low adaptation barrier but addresses only the interaction layer. The First Diamond Toolkit covers more ground but requires building custom tools. The Research-Presenter solves an immediate pain point, but it depends on integration with existing systems and lacks feasibility. The next section describes how these trade-offs are evaluated and which direction is selected.

4.4 Harris Profile Evaluation

To evaluate the three concept directions, the Harris Profile from the Delft Design Guide was used. This method supports comparing concepts by making their strengths and weaknesses visible.

Eight evaluation criteria were defined, derived from the Cycle 1 challenge analysis, probe findings, and the client-validated criteria from the post-midterm alignment (Section 4.5). The criteria were reviewed with the lead designer of the CXD team to verify if they reflected the priorities of the Collins context. The evaluation was based on accumulated probe findings and CXD team feedback. The three concept directions were scored across the eight criteria using the Harris Profile's four-point scale (--, -, +, ++) (Figure 4.4).Based on this evaluation, the Prompting Canvas was selected. The choice was made on a few: it offers the best combination of sensemaking support, designer control, workflow compatibility, and low adoption effort. Moreover, it teaches how to interact with AI, not just what to ask. Therefore, it also allows the designer to approach interacting with AI in a standardised format.

From the clients' feedback, the Prompting Canvas was also the most interesting direction. The client envisioned how their team could structure its own context and adapt.

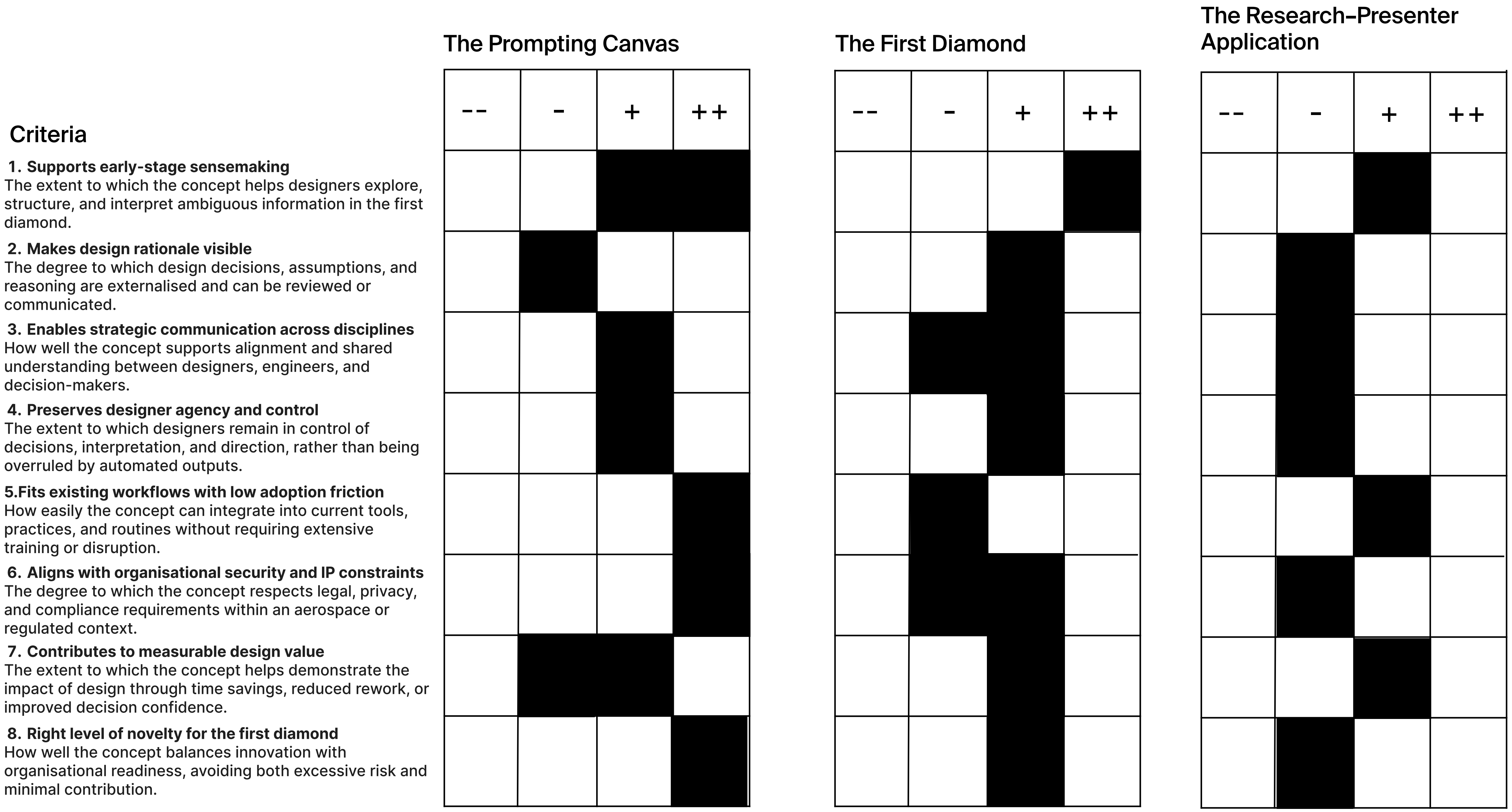


Figure 4.4 — Harris Profile evaluation comparing the three concept directions across eight criteria derived from Cycle 1 challenges, probe findings, and client-validated constraints.

Takeaways

1 The Harris Profile method was selected for selecting the concept

2. Prompting Canvas is the selected design direction. Best for sensemaking support, autonomy, workflow compatibility, and low adaptation barrier.

3. Trade-offs of the other concepts. The First Diamond Toolkit and Research-Presenter Application each addressed specific challenges well, but required greater adoption effort or had a limited scope within the First Diamond.

4.5 Post-Midterm Client Alignment

Following the midterm presentation (at which the probe findings, three concept directions, and the Harris Profile evaluation were presented), a structured alignment meeting was held with the Collins client to validate the direction and establish criteria that any intervention must meet.

Growing pressure

In companies it is appearant that design team have the ability to get smaller. This can be due to various reasons. In practice, the core of the team's work is research, sensemaking, and translation of design knowledge across disciplinary boundaries, sometimes this is not recognized within bcompanies. Meaning that the if a team become smaller there is less time for meaningful design work.

Five criteria for any intervention

Five criteria came from the meeting with the client (Figure 4.5): relevance over polish, multiple interaction modes, speed, user ownership, and structure that reduces risk.

These are more like a set of constraints. A system that fits the workflow but hallucinates fails to solve any problem. One that gives accurate results but needs its own separate workflow fails on the fourth. All five should be satisfied with the final concept.

In discussion with the client, the Prompting Canvas is confirmed as the direction. However, this also changes how I understand the problem. The client is not asking for a chatbot, a slide generator, or one more AI tool. What they need is shared infrastructure: a structured AI approach that the team can own and adapt.

01
Relevance over polish
Accurate, decision-relevant output matters more than visual presentation. AI-generated content must be verifiable against source material.

02
Multiple interaction modes
Designers need both ambient, ongoing support and targeted, on-demand queries. These are structurally different patterns.

03
Speed as primary metric
Must demonstrably compress research timelines — weeks into minutes — without sacrificing relevance.

04
User ownership
The team must be able to build, adapt, and maintain the system independently. Enable autonomy, not dependency.

05
Structure reduces risk
Formalised, shared protocols are lower-risk than unstructured individual experimentation. Traceability matters in regulated industries.

Figure 4.5 – Five client-validated criteria for any intervention at Collins.

TAKEAWAYS

1. Five constraints that work as a system. If one of these criteria is not met, the concept is not useful.

2. The problem reframes itself. The client does not need a tool. They need shared infrastructure that the team can own.

4.6 Requirements for AI Integration

Four conditions define what the Prompting Canvas must do. These come from Cycle 1 and 2: the challenge analysis, the sensemaking reframe, the five client constraints, and the probe findings.

Autonomy only works if the designer understands what AI gives them. Someone who cannot evaluate the output can also not decide when to use it. The Prompting Canvas closes this gap. It structures the interaction so that understanding builds through use.

Control is different from autonomy. Autonomy is deciding to act. Control is being able to steer what happens. Outputs need to be reviewable. Guardrails need to be adjustable. And because AI hallucinates, human review before any output enters the workflow is a requirement.

Cost matters too. Collins has no formal AI budget. However, designers who scope their prompts and use agents well are more cost-efficient and therefore easier to justify to compliance. That is a practical constraint that shapes design decisions for creating the concept later.

Takeaways

1. The canvas builds that understanding through structured use.

2. Control and autonomy are not the same thing. Autonomy is choosing to act. Control is steering the result. Both need to be there.

3. Human review is necessary

4. Token efficiency matters. No AI budget exists in Collins. Scoped, intentional use of AI is more cost-efficient.

4.7 Chapter Takeaways

1. Most First Diamond challenges require multiple AI capabilities. The relevant capability mixes shift across the Discover-Define phase.

2. The AI tools that exist today mostly help after the design thinking is done. The capabilities designers need most, like divergent exploration, are the least covered (Section 2.3).

3. The real problem is not a lack of tools. It is a lack of shared structure. Without it, the same AI capability gives inconsistent results that cannot be compared across a team.

4. Using the subagents architecture is a must. Splitting research, justification, and synthesis across dedicated agents produces better results than asking one assistant to do all three. Each task needs its own setup.

5. The Prompting Canvas was selected through Harris Profile evaluation as the direction with the best based on criteria of sensemaking support, designer control, organisational feasibility, and easy adoptability.

CHAPTER 5

Literature Review

Theoretical foundations for the three-layer canvas. From canvas frameworks and context engineering to scaffolding theory and the automation paradox. This chapter takes design decisions already made and shows they align with established knowledge in HCI, organisational learning, and human factors.

5. Literature Review

Theoretical foundations: canvas 5. Literature Review

The design decisions so far have been based on identifying, through iterative probing and evaluation (Chapter 4), that the Prompting Canvas was the direction, also based on what the client saw as desirable and feasible for their team. This chapter provides the theoretical foundation for the developed concept.

5.1 Canvas

The closest prior work is the Prompt Canvas by Hewing and Leinhos (2024). They spotted a real gap: people interact with LLMs without any shared structure for doing so. Their canvas “provides a practical approach for leveraging the potential of Large Language Models” and “is primarily designed as a learning resource for pupils, students and employees, offering a structured introduction to prompt engineering” (Hewing & Leinhos, 2024, p. 1). It organises the interaction into eight dimensions, guiding the user through writing a well-structured prompt. It addresses the Gulf of Envisioning (Subramonyam et al., 2024), but only at the level of one prompt in one session.

As Chapter 2 discussed, what counts as “good prompting” has shifted since then. The focus now is less about writing better prompts and more about designing the context the model works in. That said, what the Prompt Canvas got right still holds. A structured, visual format helps people interact with AI more effectively. The open question is whether that format can stretch to cover more than a single prompt.

FIGURE 5.1 — CANVAS FRAMEWORK EVOLUTION: FROM FLAT TO LAYERED

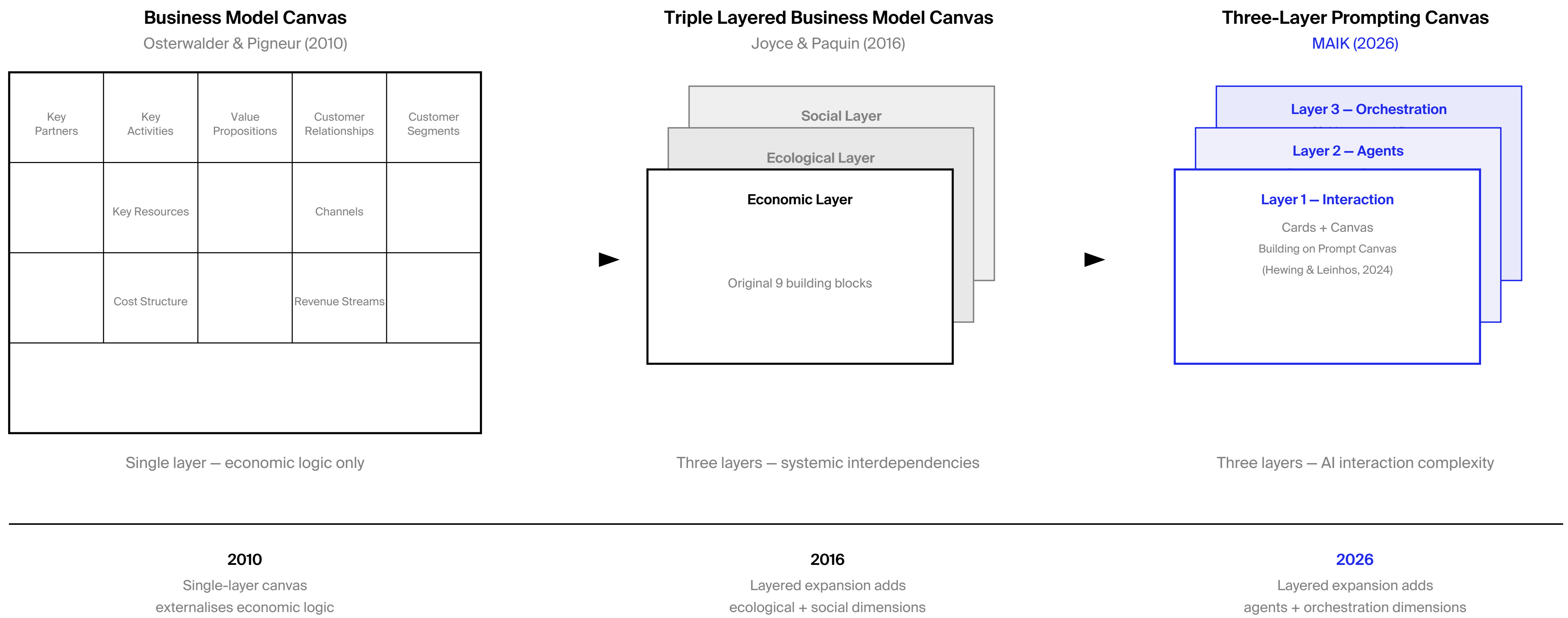


Figure 5.1 - Canvas framework evolution: BMC (2010) -> TLBMC (2016) -> Three-Layer Prompting Canvas (2026).

Canvas formats work because they take complicated systems and make them usable. The Business Model Canvas (Osterwalder & Pigneur, 2010) showed this for strategy. Before it existed, business models were fragmented across the entire company. Osterwalder and Pigneur turned that into a one-page document that a team can fill in collaboratively. It can be shared across departments and revisited if something changes. The guiding questions on the canvas stop users from filling it in too quickly. The visual layout makes it something collaborative.

Over time, the Business Model Canvas itself expanded. Joyce and Paquin (2016) added an ecological layer and a social layer on top of the original economic one, creating the Triple Layered Business Model Canvas (Figure 5.1). The original stayed intact. The scope grew around it. That is what happens when the system behind a canvas gets more complex: you add layers rather than starting from scratch. One thing to note here is that the TLBMC’s layers sit next to each other. Economic, ecological, and social are parallel views. MAIK’s layers work differently. They stack, with each one building on the previous.

The same move applies to the Prompt Canvas. It works well for structuring a single AI interaction. But agents, memory, and orchestrated workflows are outside its current scope, something Hewing and Leinhos state. For MAIK, the Prompt Canvas is the starting point. It stays as the foundation, and the layers the current AI landscape demands are added on top.

5.2 From Conversation to Orchestration

In Chapter 2, three levels of AI interaction are introduced: a single conversation, a configured agent, and an orchestrated workflow. This progression is named in the current literature. I drew it from how AI systems have been built over the last few years. But what we do see in current literature is the barrier that each layer brings. This can be a potential bottleneck for non-technical users, like designers (Figure 5.2).

Subramonyam et al. (2024) name the problem in theory: the Gulf of Envisioning, three cognitive gaps between what a user wants and what they can get an AI to produce. Zamfirescu-Pereira et al. (2023) show what this looks like in practice. They asked non-experts to write prompts for LLMs. Participants over-generalised from one attempt, assumed the AI would understand them the way a person would, and had no way to tell whether a bad result came from a bad prompt or a bad model. A blank text box gives no guidance. What is important to note is that the development of AI has addressed the challenges that were found in 2023, which could potentially be solved by the better-performing LLM. But Gulf of Envision may persist for non-technical users.

At the second and third levels, the problem changes. Schömbs et al. (2025) describe the shift from conversation to orchestration as an open HCI problem. A user can talk to one agent fine. But supervising a group of specialised agents, each handling a different part of the work, is something no current interaction model supports for non-technical users. The paradigm that works at one level does not carry over to the next.

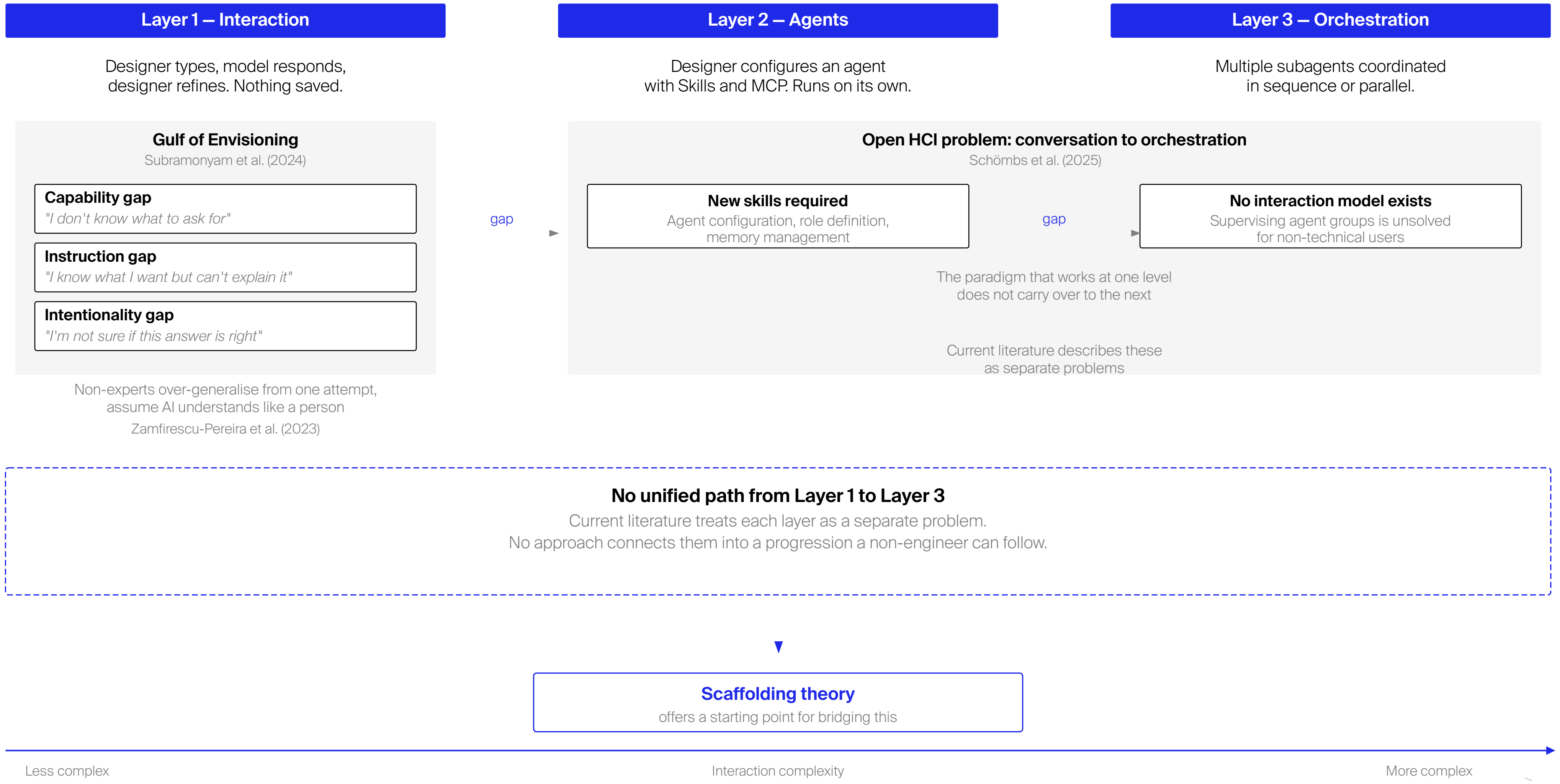


Figure 5.2 - Three layers of AI interaction and their barriers for non-technical users

The current literature describes these levels as separate problems. There is no unified approach that connects them into a single path that a non-engineer can follow. How to do that is a design question, and scaffolding theory offers a starting point

5.3 Making the Layers Learnable

AI has a quickly growing complexity. The proposed layers can still easily become three barriers. If each layer feels like starting over, the framework fails quickly. Here, the scaffolding theory explains how to prevent this.

Vygotsky (1978) described the Zone of Proximal Development: the gap between what someone can do on their own and what they can do with the right support. Wood, Bruner, and Ross (1976) developed the concept of scaffolding and identified three factors that make it work. The support needs to match the learner's current level. It needs to step back as competence grows. And the learning should happen through doing the actual work, not through a separate training session. What makes scaffolding different from instruction is that last point. The person learns by doing the actual activity with support.

Originally, scaffolding was developed for individual learning with a (human) tutor. But in team settings, the structure of a shared tool can serve the same function. Vogel et al. (2017) ran a meta-analysis on collaboration scripts, structured formats that guide how groups interact, and found they improve both individual learning and shared understanding within the group. When people work through a shared structure together, alignment happens as a side effect.

A framework built on scaffolding teaches the thinking, not just the doing. Its goal is to make itself less necessary over time.

Two principles come out of this. First, each level should be reachable based on what the previous level taught. Learning happens through use, not through documentation. Second, a framework built on scaffolding does not need to be the place where all the work happens. It needs to teach thinking. Once someone understands how to structure an interaction, they can take that understanding anywhere. But scaffolding only addresses how people learn. It does not address what happens when AI takes on more, and the person pays less attention. The next section looks at that risk.

5.4 The Automation Paradox

More capable AI creates a specific risk that the Learning to use AI effectively comes with a risk. As people become more comfortable, they start to trust the output more. They check less. And the moments when checking matters most are exactly the moments when they are least likely to do it.

Endsley (1995) breaks down what gets lost. Her model of situational awareness has three levels.

1. Perception: noticing what information is there.

2.Comprehension: understanding what it means.

3. Projection: anticipating what comes next.

Automation can worsen all three. If an AI retrieves information, the user may stop looking at what other information exists. If it interprets data, the user may stop questioning the reasoning. If it suggests a next step, the user may stop considering alternatives. Endsley developed this model for dynamic systems in general, but aviation is one of its main application domains. At Collins Aerospace, the consequences of lost situational awareness are not abstract. An unchecked hallucination could pass a gate review and end up in engineering commitments.

Shneiderman (2022) reframes the question. The goal is not to pick between giving people control or giving AI autonomy. It should be designed for both. Systems should make people more capable without dulling their ability to judge what is happening.

Amershi et al. (2019) turn this into practice. Their 18 guidelines for human-AI interaction, published at CHI 2019, are organised by when they apply: at the start, during use, when things go wrong, and over time. Three are directly relevant here. Make clear what the system can do. Make clear why it did what it did. And give the user controls that work across the whole system, not just in one place.

The design principle that needs to be considered here: every output needs to be visible to the user(s). It needs to be reviewable and changeable at any time before it enters the workflow. Human review is not a preference. It is a structural requirement. This addresses what happens when AI acts. The next section addresses how people first encounter it.

5.5 Physical Artefacts

Currently, AI interaction is individual. One person sits at a screen, types into a box, and works with whatever comes back. Dang et al. (2023) ran a CHI study with 129 participants to see what happens when you offer people two options: write your own prompt or pick from a set of suggestions. Most people picked the selection because writing a prompt from scratch takes real effort.

Sanders and Stappers (2008) write about how tangible objects work as boundary objects in collaborative design settings. Put something physical on a table, and people gather around it. They point at it. They move it. They put something else next to it, and the conversation shifts. That does not happen the same way with a card on a screen.

Smith et al. (2024) brought these threads together specifically in the context of AI. They developed a set of 16 printed cards based on Long and Magerko's AI literacy framework. They tested them with 50 participants, none of whom had a technical background. The effect was measurable: participants asked sharper questions about AI, gave more varied feedback, and were more engaged. In a smaller co-design session with six people, three of whom had low AI experience, the cards helped build common ground between the technical and non-technical halves of the group. The researchers noted that the physical version performed better than the digital version for group sessions. Being able to spread cards out, group them, and rearrange them supports a kind of conversation that screens do not.

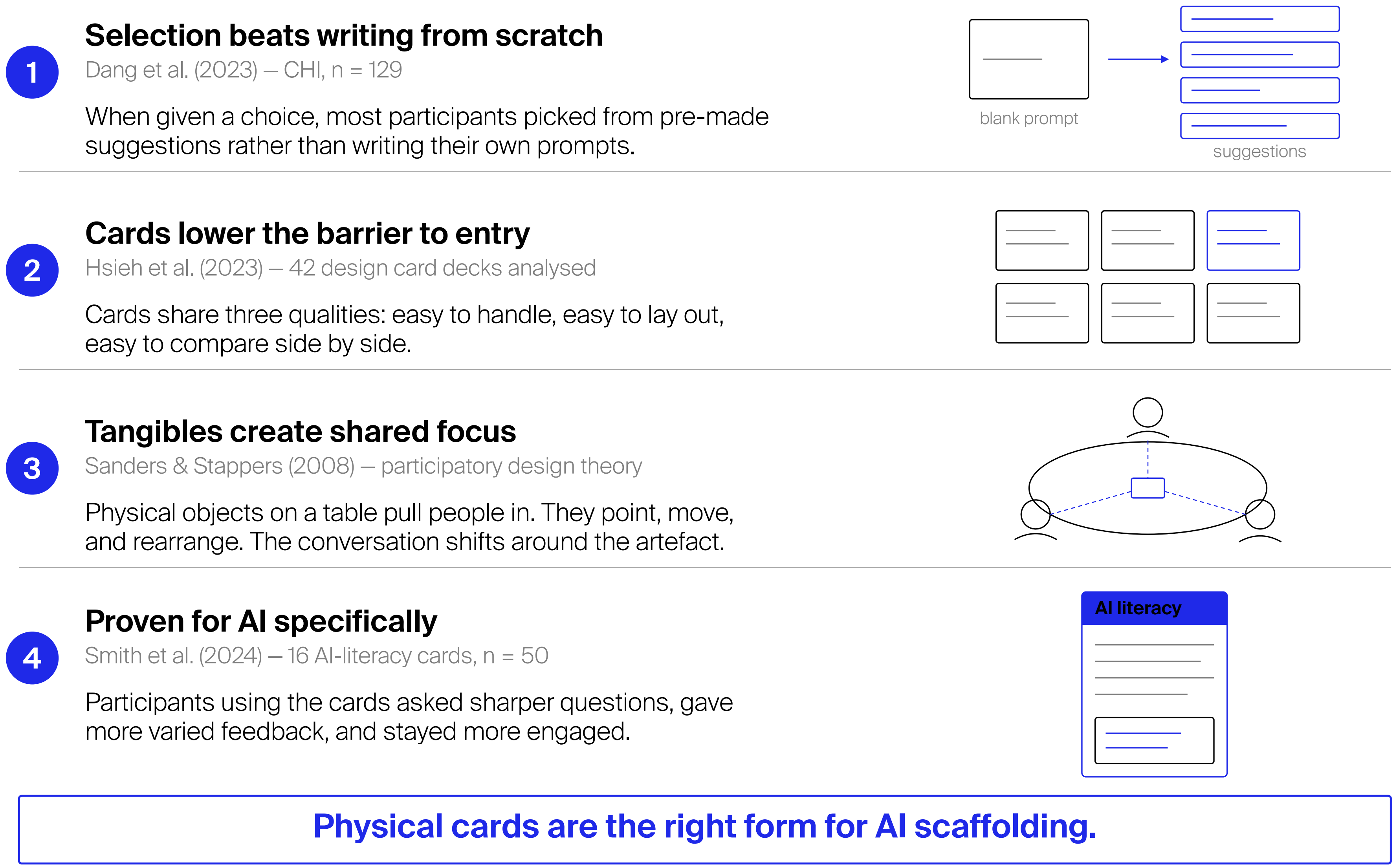


Figure 5.3 - Physical cards as artefact to interact with AI.

The point here is about where AI interaction begins. If it starts with a blank text box, it stays individual. If it starts with a physical set of options that a team can browse and discuss together, the activity is shared before anyone opens a laptop. That shared starting point becomes more valuable for the team.

5.6 Consistency over Individual skills

AI use in teams tends to stay individual. Xiao et al. (2025) followed a software development team for two years. Everyone adopted AI tools. But the tools stayed personal: coding, writing, documentation. The team expected AI to help coordinate their work. Instead, collaboration gaps persisted. Worth noting software development is already fairly individual work. In a design team like the one at Collins, where most of the work happens together, the disconnect between shared projects and individual AI use is harder to ignore.

Nonaka and Takeuchi (1995) have a name for why this keeps happening. They describe four ways organisations move knowledge around: socialisation, externalisation, combination, and internalisation. The one that usually gets stuck is externalisation (Figure 5.4). That is when someone has to take what they know from experience and turn it into something other people can actually use. A designer who stumbles on a prompting approach that gives good results has learned something real. But if nothing in their workflow asks them to write it down or talk it through with the team, that knowledge stays with them. The next person who faces the same task starts over.

The principle here is a choice. A framework for AI interaction can try to make each person better at AI. Or it can try to make the team more consistent. These are different goals. Individual skill is uneven and invisible. Consistency means that what one person learns becomes available to everyone. For a distributed team working across locations and projects, consistency might be worth more.

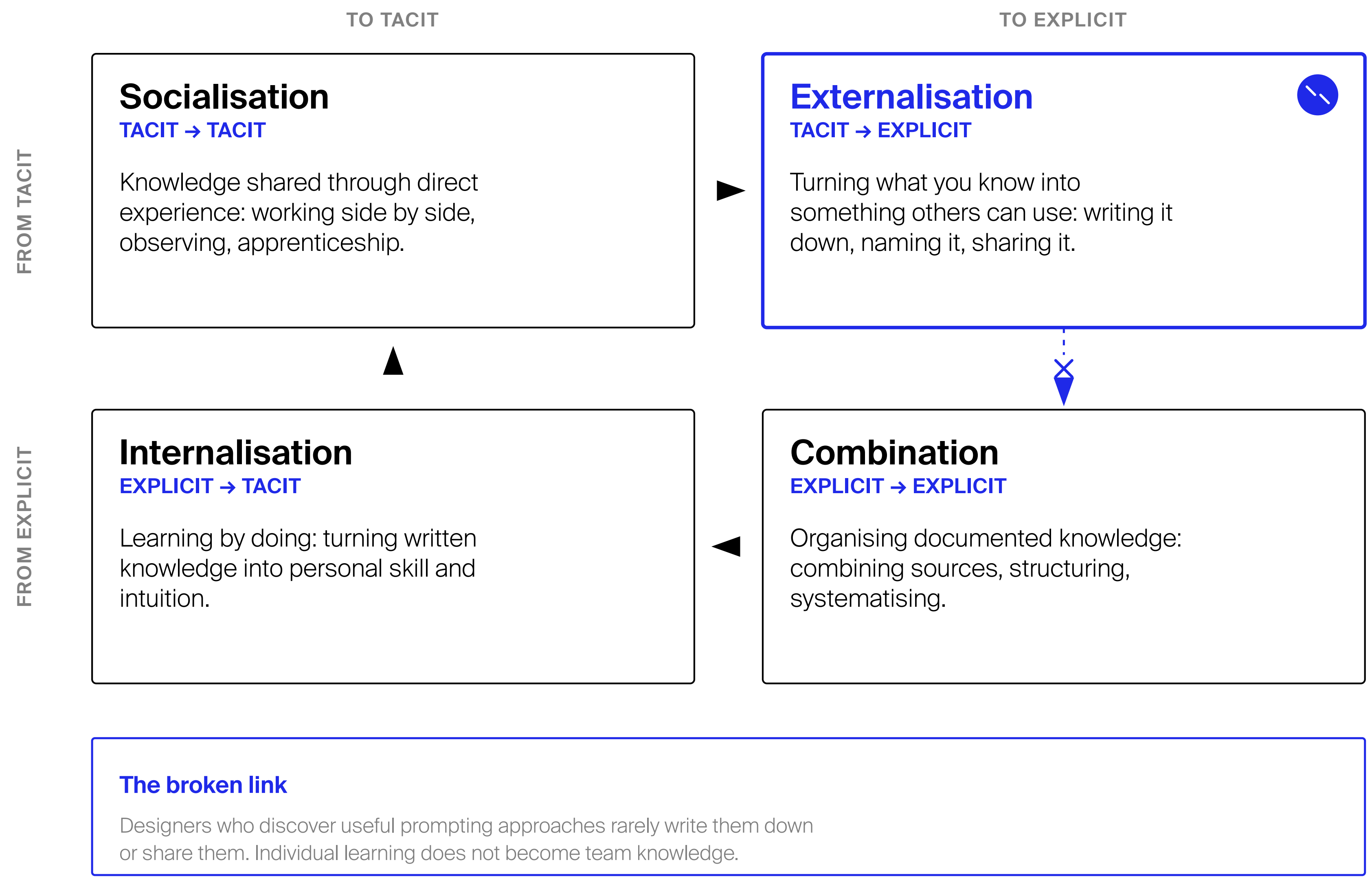


Figure 5.4 - The SECI knowledge cycle: where individual AI learning gets stuck

AI use should not remain an individual skill. It should become a team capability.'

5.7 Chapter Takeaways

1. When the system behind a canvas gets more complex, canvas frameworks respond by adding layers. The Business Model Canvas became the Triple Layered BMC. The same logic applies to AI interaction. Single-prompt frameworks do not cover what AI systems can do now.

2. AI interaction has gone from typing a prompt to configuring agents to coordinating multi-step workflows. The research community treats these as separate problems. There is no unified path through them that someone without an engineering background can follow.

3. Scaffolding makes growing complexity manageable. It works by matching support to where the learner is, pulling back as they get better, and embedding the learning in real tasks. When that scaffolding occurs through a shared structure, it also brings team members closer together in their understanding of the work.

4. The better AI gets, the fewer people check what it produces. Situational awareness drops at all three levels: noticing what is there, understanding what it means, and anticipating what comes next. Any output that comes from automation needs to be visible, reviewable before it goes anywhere, and changeable after the fact.

5. Physical objects change how people engage with AI. Given the choice between writing a prompt and picking from a set of options, most people choose the latter. When those options are cards on a table rather than items on a screen, the interaction stops being something one person does and becomes something a group does.

CHAPTER 6

Problem Framing & Design Brief

Three cycles converge into one brief. Workflow challenges, probe findings, and client constraints compile into a single statement of what the canvas must achieve. The chapter sets the design goal and lists the requirements in a Programme of Demands.

6. Problem Framing & Design Brief

This chapter synthesises the outcomes of Cycles 1 and 2 into a problem framing and design brief. It establishes what the design intervention is meant to solve, why a direction was selected, and what requirements it must meet.

The previous chapters looked at the organisation, the AI tools available to it, and the theory that connects the two. This chapter takes those findings and turns them into a design brief: what needs to be solved, what any solution has to do, and how it will be tested. The full programme of demands is in Appendix B. Chapter 7 builds on it.

6.1 Synthesising Research Takeaways

This chapter synthesises takeaways from three sources: the Collins workflow research, the AI capability explorations, and the literature on human-AI interaction. Together they lead to one conclusion and to the design brief that follows.

01 **Designers at Collins Aerospace spend a large part of their time translating design choices into measurable, evidence-based arguments for stakeholders.** They follow the Double Diamond model. In the first diamond, the goal is to explore, make sense of what they find, and turn it into a direction the client can agree on so the project moves forward. Designers experience roughly eight challenges in this phase. They are all connected to the same thing: having to converge before all alternatives have been explored (Chapter 3). AI could support designers in this phase. Several designers at Collins already use AI tools in their daily work. However, everyone uses a different tool and approaches it differently. For the same task, this produces fragmented outputs. That fragmentation costs more time. Current AI tools are built for synthesising, not for the exploration and sensemaking that designers need at the start of the process (Chapter 2). Interacting with AI in ways that use its full capabilities is a separate problem. Subramonyam et al. (2024) call it the Gulf of Envisioning. Designers do not know what AI can do, how to ask for it, or what they are actually asking for.

02 **The AI landscape evolves quickly.** New tools and capabilities appear regularly, and the knowledge required to adopt them is difficult to build. This is compounded by the fact that whatever one designer learns about using AI stays with that person. Prompting habits that produce good results individually fall apart when outputs need to be compared or combined at the team level. Nonaka and Takeuchi (1995) describe the mechanism: knowledge about how to work with AI is tacit. It lives in personal routines and habits. Current AI interfaces do not support sharing what was learned or what worked. Tacit knowledge does not automatically become team knowledge. It needs a structure that makes it explicit and shareable. At Collins, this structure is not in place. Currently, there are no moments for designers to compare their approaches to AI or to evaluate what has produced useful results. What the team needs is not training on how to use AI. They need infrastructure that they own and can adapt over time. The post-midterm alignment with the client (Chapter 4) produced five criteria for what that infrastructure must deliver.

First, relevance over polish: the accuracy of AI output matters more than its presentation.

Second, two interaction modes: AI must be able to guide a conversation actively or respond to users' inquiries and instructions.

Third, speed: if AI does not produce results faster than manual effort, its adoption loses justification.

Fourth, user ownership: both the knowledge generated and the decisions made with AI remain with the designer.

Fifth, structure as risk reduction: AI tools are operated by external corporations, and structured use allows the team to work within Collins' proprietary data policies.

03 **In the current AI evolution, different types of AI can be categorised into three layers: single conversations, configured agents, and orchestrated workflows (Chapter 5).** Most designers at Collins use the first. The second and third would allow them to save setups and connect agents into larger workflows. However, moving from the conversational level to the workflow level requires technical knowledge that most designers lack. Research confirms that no established path exists from conversation to orchestration for non-engineers (Schömbs et al., 2025) and that non-experts get stuck when left to figure out the progression on their own (Zamfirescu-Pereira et al., 2023). From the literature, two opportunities emerge to present this progressive model for introducing AI.

The first is scaffolding. Scaffolding works through graduated support, where each level is reachable from the one before, withdrawal as the designer gains competence, and embedding the learning in the task itself rather than in separate training (Wood, Bruner & Ross, 1976). However, as AI takes on more of the work, a risk appears. Endsley (1995) showed that as automation increases, people stop checking the system's output. Each level of AI use requires review moments, when the designer evaluates the output. The design has to satisfy both: making each level reachable while ensuring the designer stays aware of what the AI produces at every step.

The second opportunity is physical artefacts. Research showed that physical cards help people engage with complex systems. They make it easier to choose from options rather than starting from a blank page (Dang et al., 2023).

6.2 Problem Framing

PROBLEM

Designers working in the first diamond juggle many tasks simultaneously exploring information, synthesising inputs, and building the evidence needed to move forward. Time is limited. AI has the potential to take on parts of this work and free designers to focus on what requires their judgment.

While designers already use AI individually, current interactions are unstructured and inconsistent shaped by personal habit rather than shared practice. This defeats the potential of AI to become a shared team capability that grows through collective use.

The gap

The CXD team has no shared way of working with AI. There is no common approach for deciding when to use, how to configure or how to learn from what a colleague already tried. The field moves fast. What counted as good AI use a year ago looks different today. Without a shared structure, every designer starts from scratch every time. AI at Collins stays something people do on their own. It does not become something the team does together. While this brief is written for Collins, the problem it addresses exists in any organisation where designers work with AI without shared infrastructure.

THREE CONSTRAINTS

- **Organisational**
  The team works across locations. The solution has to work without everyone being in the same room.
- **Technical**
  AI access is limited to public tools. Nothing can touch proprietary Collins systems.
- **Human**
  Designers have limited time. Anything that needs a lot of setup before it becomes useful will not get used.

6.3 Design Goal and Programme of Demands

CXD designers at Collins Aerospace need a shared, collaborative approach to interacting with AI during early-stage design work.

The goal is to design an intervention that gives the team that structure supporting sensemaking, turning individual learning into team knowledge, preserving designer agency, and fitting within an complex organisational environment.

TARGET USERS

Primary
CXD designers in Discover and Define phases. Already use AI informally. Not AI specialists.

Secondary
Design leads and project managers who review, build on, or present outputs made with AI support.

OBJECTIVES

- Structure AI at three levels: prompting, agents, workflows.
  scaffolding — §5.3
- Make AI use visible and shareable across the team.
  knowledge conversion — §5.6
- Keep the designer in control. Outputs checkable at each step.
  automation paradox — §5.4
- Make it easy to start. Pick options, not a blank field.
  physical artefacts — §5.5

CONSTRAINTS

- No access to internal proprietary systems. Public AI tools only.
- Must fit into existing tools and routines. No long onboarding.
- Must work across locations. Team cannot depend on co-presence.

SUCCESS METRICS

- Designers use it on their own, without a walkthrough.
- AI outputs more consistent and comparable across the team.
- Designers stay in charge. Outputs need human judgement at every step.
- Used in a real project during evaluation.

The evaluation criteria below restate the objectives, constraints, and success metrics as testable questions.
Each criterion maps to one or more requirements in the full Programme of Demands (Appendix B).

EVALUATION CRITERIA

- Does it help designers deal with complexity and build shared understanding early in a project?
- Are the outputs fast enough to be useful and clear enough to share?
- Do designers stay in control of what they interpret and decide?
- Can someone get useful results on their own, without prior training?
- Does it fit into how Collins already works, without getting in the way?
- Does it help the team work with AI together?

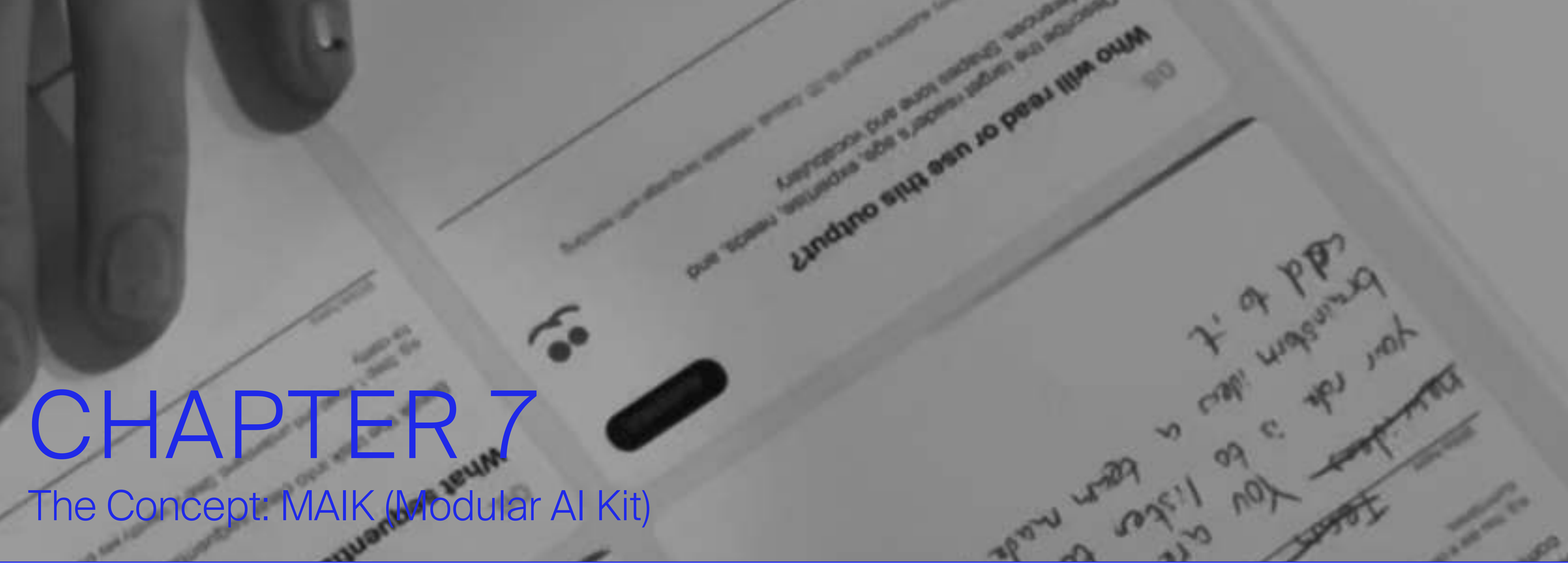
6.4 Chapter Takeaways

1. Designers at Collins use AI independently. Without a shared approach, everyone gets different results, and the kind of support most needed early in a project is the least available in current tools.

2. What one designer learns about AI stays with that person. Without something that makes personal knowledge visible and reusable, the team does not get better. Only individuals do.

3. Three levels of AI use exist, but there is no way to move from one to the next without the technical skills most designers lack. Scaffolding can build that path. The automation paradox means every step toward more AI needs a matching step in human checking. Physical artefacts and layered canvases make the starting point easier.

4. Collins set five criteria for any intervention. They describe a way of working, not a product.



CHAPTER 7

The Concept: MAIK (Modular AI Kit)

Cycle 3: from design direction to concept. MAIK Cards, MAIK Canvas, and an infrastructure layer deployable as a web or mobile application. This chapter presents the architecture, agent specifications, visual design, organisational implications, and implementation pathway of the concept: MAIK.

7. The Concept: MAIK (Modular AI Kit)

Cycle 3: from exploration to concept. A failed Zapier direction, a supervisor reframe, and Claude's Skills architecture clarified what Layer 2 and Layer 3 mean in practice.

Chapter 6 defined what the solution has to do: a structured way for the CXD team to configure AI at three levels: individual prompting, configured agents, and orchestrated workflows. Five criteria from the Programme of Demands set the bar. This chapter shows what was built.

MAIK is a Modular AI Kit. It has two physical artefacts and a digital infrastructure layer. The physical artefacts are the MAIK Cards and the MAIK Canvas. The digital layer is a web application that connects them to Claude Skills: configured AI workflows that run without the designer having to rebuild the setup each time.

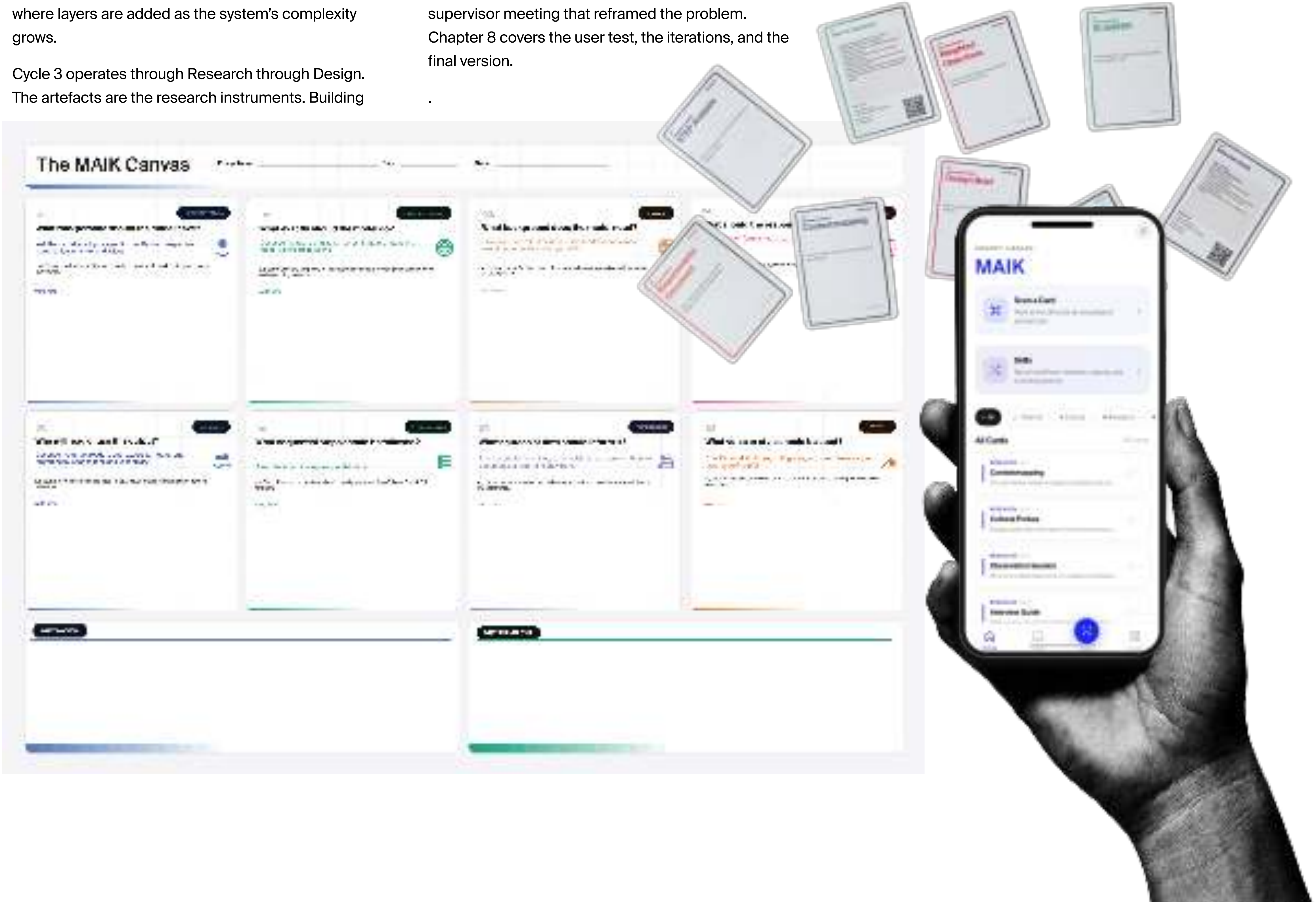
The three layers from Chapter 5 give MAIK its structure. Layer 1 is individual and team prompting through cards and canvas. Layer 2 is configured for agents through Skills. Layer 3 connects Skills into a sequence with human review between each step. Each layer builds on what the designer learned in the one before.

The scaffolding principle from Section 5.3 is built into the architecture. The cards teach prompting structure. The canvas turns that structure into something shareable. The Skill is what that shared structure becomes when it gets formalised and made reusable.

The canvas format has precedent. The Business Model Canvas became the Triple Layered BMC when new dimensions were added. The Prompt Canvas (Hewing & Leinhos, 2024) covers a single AI prompt. MAIK takes that same idea further: a growing canvas where layers are added as the system's complexity grows.

Cycle 3 operates through Research through Design. The artefacts are the research instruments. Building them is how the research questions get investigated. The dead ends are part of the story: the N8N and Zapier explorations that did not work, and the supervisor meeting that reframed the problem. Chapter 8 covers the user test, the iterations, and the final version.

.



7.1 Layer 1: MAIK Cards and the MAIK Canvas

Layer 1 is where a designer starts using AI with a clear, consistently formatted prompt. Two artefacts serve this layer. The MAIK Cards introduce structured prompting on an individual basis. The MAIK Canvas makes it a team activity.

MAIK Cards

The cards are the entry point. Each card presents one ready-to-use AI prompt built around a design method that designers already know. The methods come from the Delft Design Guide (Boeijen et al., 2014) and from tasks the CXD team performs in the First Diamond. A designer picks up a card, fills in one bracketed variable, and gets a structured output. No configuration knowledge is needed.

48 general cards applicable across design practice, covering recurring research and synthesis tasks

8 Collins-specific cards developed directly from the First Diamond workflow

Every card is two sided and follows the same strutcure

Figure 7.1 - MAIK: The Prompting Cards artefact

Figure 7.2 - Prompting Cards Front and Back

Card ID and category tag

Title and a one-line description of what the prompt produces

A pre-written prompt template with a single [TOPIC] variable

Prompt

A QR code — scan it, and the prompt copies directly to the clipboard of the device: pick a AI model, paste, edit, and go

Why are the cards physical?
AI has an almost infinite input space. Most designers do not know where to start. Given the choice between writing a prompt from scratch and picking from a set of options, most people pick the options (Dang et al., 2023). The cards collapse that space into 56 discrete, named actions.

The format matters as much as the content. Cards are easy to handle. You can lay out and compare several at once. They have the potential to lower the pressure to engage with something complex When tested for AI use, physical cards outperformed digital ones in group settings (Smith et al., 2024). People asked sharper questions and were more engaged.

At Collins, the PowerPoint deck works as a boundary object because the whole team can see it and react to it (Section 3.4, Sanders & Stappers, 2008). Cards work the same way. A card on the table changes what happens in the room. It turns AI exploration into a shared activity rather than having one person type into a text box. If AI starts with a blank field, it stays individual. If it starts with physical options that a team can browse together, the activity is shared before anyone opens a laptop.

The cards are not a training programme. They are an invitation. You flip through them, lay them out, sort them, and debate which ones fit the brief. That act of handling them is the exploration. The designer does not need to understand AI. They need to recognise a situation. Recognition first, then adoption.

What does picking up a card do
When a designer picks up a card and fills in the [TOPIC] variable, they take what they tacitly know about the task, the project, the client, the current phase, and put it into a structured form. In Nonaka and Takeuchi's (1995) terms, that is externalisation: making implicit knowledge explicit. When two designers do this together, comparing their choices and discussing which card fits the brief, it becomes a combination: merging individual knowledge into shared understanding. The card format turns a solo prompting action into a knowledge conversion event (Section 5.6).

Other AI card decks teach AI literacy (Smith et al., 2024) or catalogue design methods (Hsieh et al., 2023). The MAIK Cards are structured AI prompts mapped to a professional design workflow through the capability analysis in Section 4.1. They are physical artefacts that trigger knowledge conversion when used in a team. They are the first scaffolding layer (Section 5.3). The support is built into the card. The designer learns structured prompting by doing it, not by studying it. As competence grows, the card becomes less necessary. That is what scaffolding does.

7.3 - MAIK card being scanned by the MAIK app.

How the card set was built

Section 4.1 mapped eight recurring challenges in the First Diamond onto five AI capability clusters: Divergent Exploration, Sensemaking & Structuring, Semantic Transformation, Externalisation & Visualisation, and Verification & Constraints. Most challenges need more than one capability, and change over the entire span of the double diamond (Figure 7.4). The card categories respond to that mapping. Each category corresponds to a capability cluster and a phase of work:

- R — Research (R01–R08): methods for early-stage contextual inquiry and desk research
- A — Analysis (A01–A08): tools for structuring and interpreting design findings
- I — Ideation (I01–I08): methods for generating and diverging on concepts
- C — Critique (C01–C08): frameworks for evaluating and stress-testing ideas
- S — Synthesis (S01–S08): tools for clustering insights and defining direction
- W — Writing (W01–W08): templates for communicating design work to stakeholders
- X — Collins-specific (X01–X08): methods tailored to the CXD First Diamond workflow

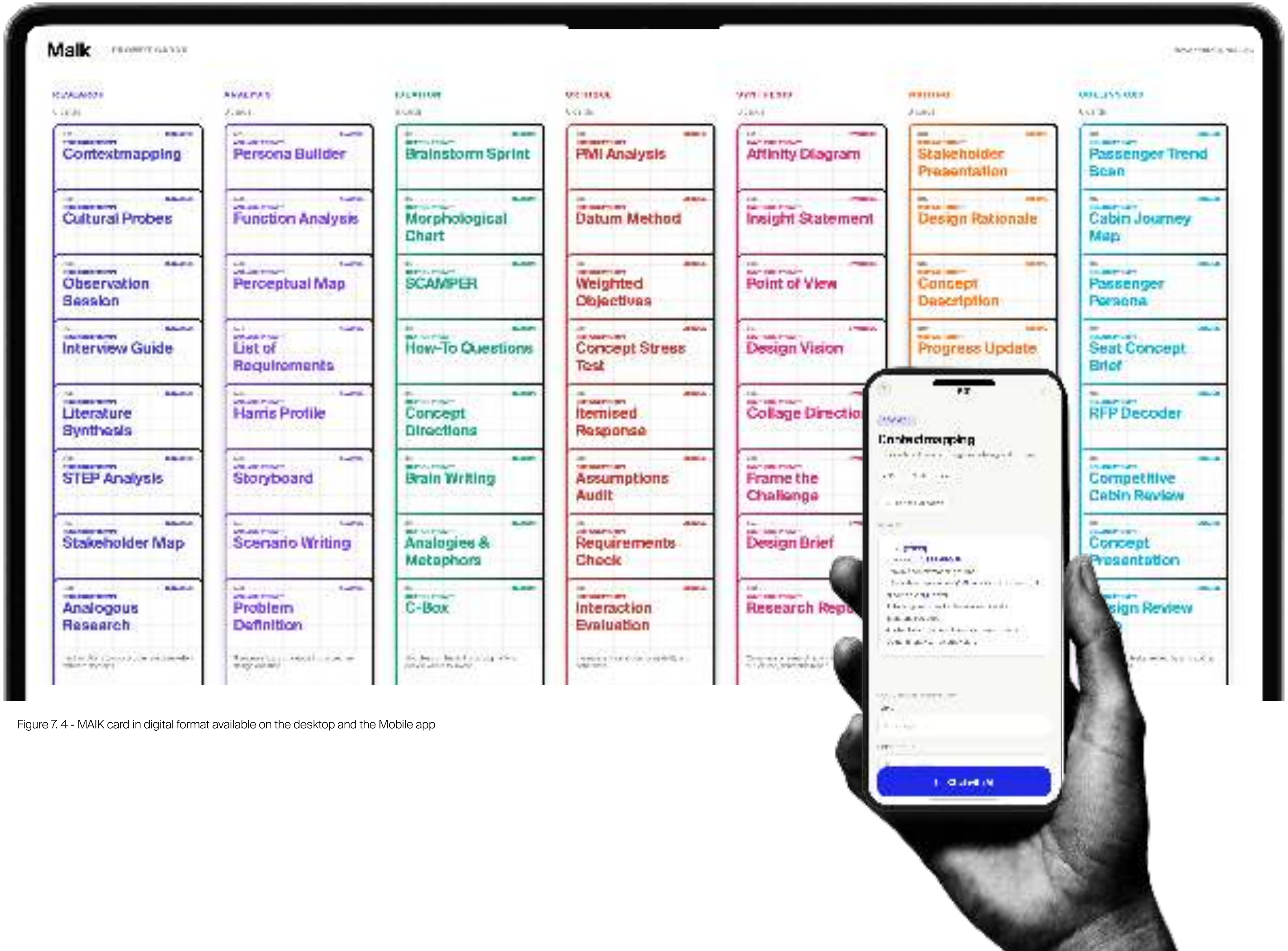


Figure 7. 4 - MAIK card in digital format available on the desktop and the Mobile app

The MAIK App

The MAIK App is the digital infrastructure for the toolkit. The Cards and Canvas produce structured prompts and team configurations. Without somewhere to store them, each session starts from zero. The app captures what designers create and organises it into a prompt library that builds up over time.

The app was built with the help of Claude Desktop on a Pro account. Due to limited context, new chats were opened when an iteration stopped working. The application was hosted on Vercel. The full backend found in Appendix D. The architecture of the app can be seen in Figure 7.5.

The app has four functions. Home presents all 56 cards by category, with a personal section for configurations a designer returns to. Scan connects the physical cards to the digital layer. Pointing the phone at a QR code copies the prompt and opens a model selector: Claude, Gemini, Perplexity, or ChatGPT. That keeps the cards model-agnostic. The designer chooses the tool that fits the task.

The library stores all saved cards in one searchable collection. The Canvas function is the digital counterpart of the physical A3 worksheet. A designer fills in the poster by hand, photographs it through the app, and Claude processes the handwriting. It assembles the eight fields into a structured prompt card and saves it to the library. What starts as a handwritten team exercise becomes a reusable digital artefact.

Designers can also create custom cards. If someone develops a configuration through the Canvas that is not in the standard set, they add it through a form: title, tagline, category, AI tool, and prompt text. The new card appears in the library next to the pre-built ones. That is how team knowledge from the Canvas enters the shared library. As designers build their own cards, they rely less on the pre-built set.

The following pages are dedicated to the full overview of each function.

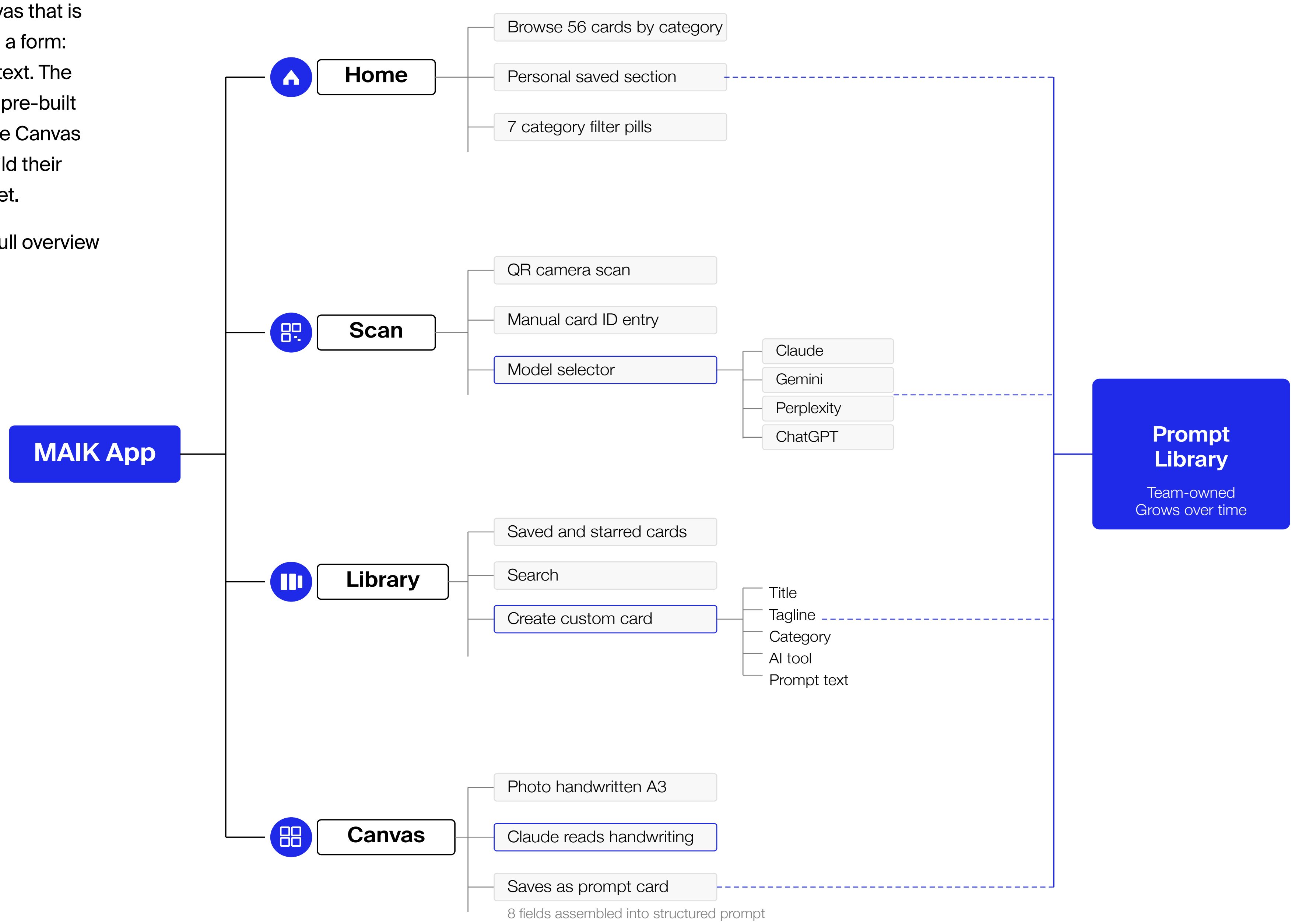


Figure 7.5 – Information architecture of the MAIK App, showing the four functions and how they feed into the shared Prompt Library.

Figure 7.X. MAIK application Wireframe

Home screen

The user opens the Home screen and browses the card library, filtered by the different categories (Figure 7.6). Tapping the Scan tab activates the phone camera. The user points it at a physical MAIK card and the QR code is detected automatically. The app then shows the card detail with its pre-written prompt. From here, the user taps “Chat with AI” to start a conversation where the prompt is already loaded as context.

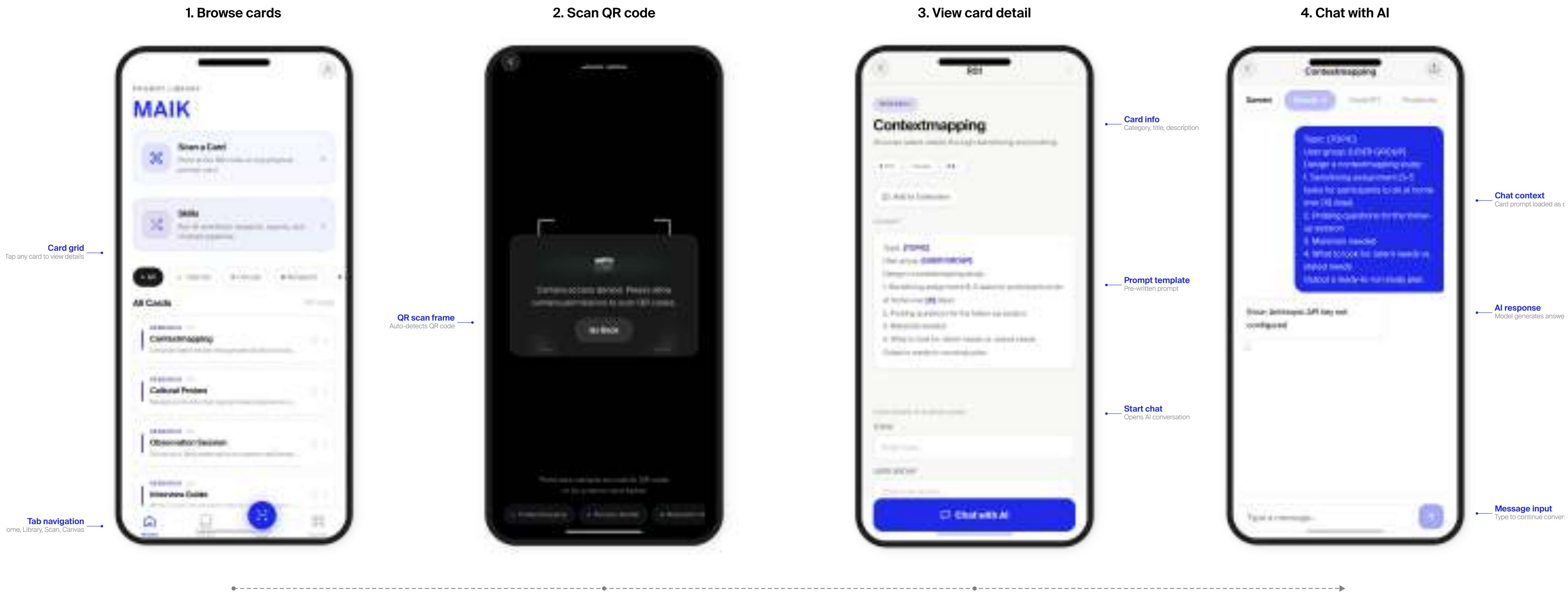


Figure 7.6 - Scan-a-card user flow in the MAIK App

Library

The Library stores all prompts the user has saved from previous sessions (Figure 7.7). Cards appear in a scrollable list, organised by thhe different categories with colour-coded labels. Tapping a card opens the full detail view, showing the prompt text, a copy button, and the option to start a new chat. Users can also star frequently used prompts for quick access.

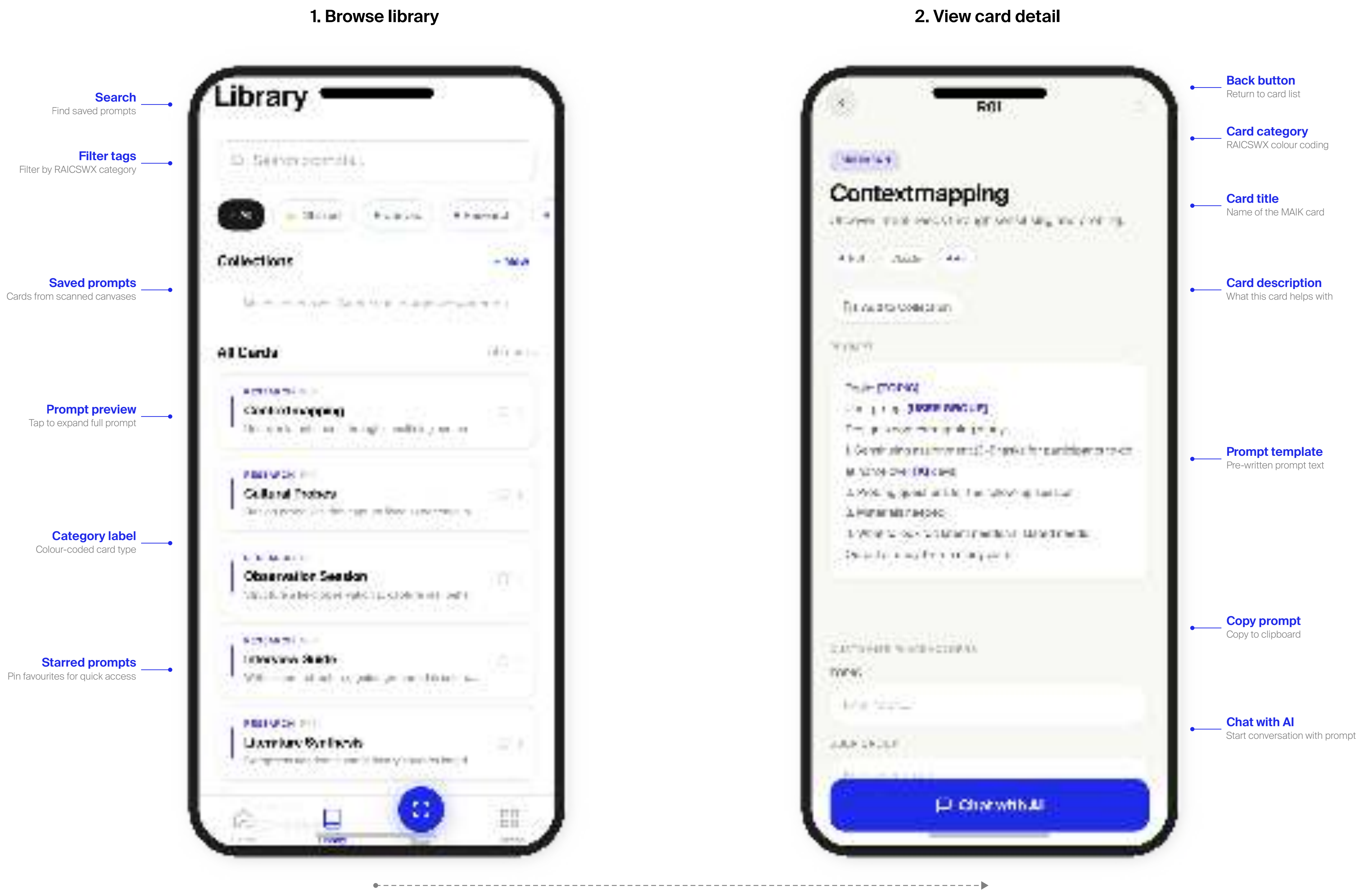


Figure 7.7 – Prompt Library: browsing and viewing saved cards

Scanning Canvas

The Canvas screen offers two ways to input the canvas (Figure 7.8) . In Scan Canvas mode, the user photographs a handwritten A3 Prompt Canvas poster. The app sends the image to Claude, which reads the handwriting and extracts the eight canvas sections. It then assembles these into a structured prompt card with a title, description, and full prompt text. The user reviews the result and saves it to the library. In Build Digitally mode, the user fills in the same eight sections through form fields: Persona, Task, Context, Output, Audience, Step-by-Step, References, and Tonality.

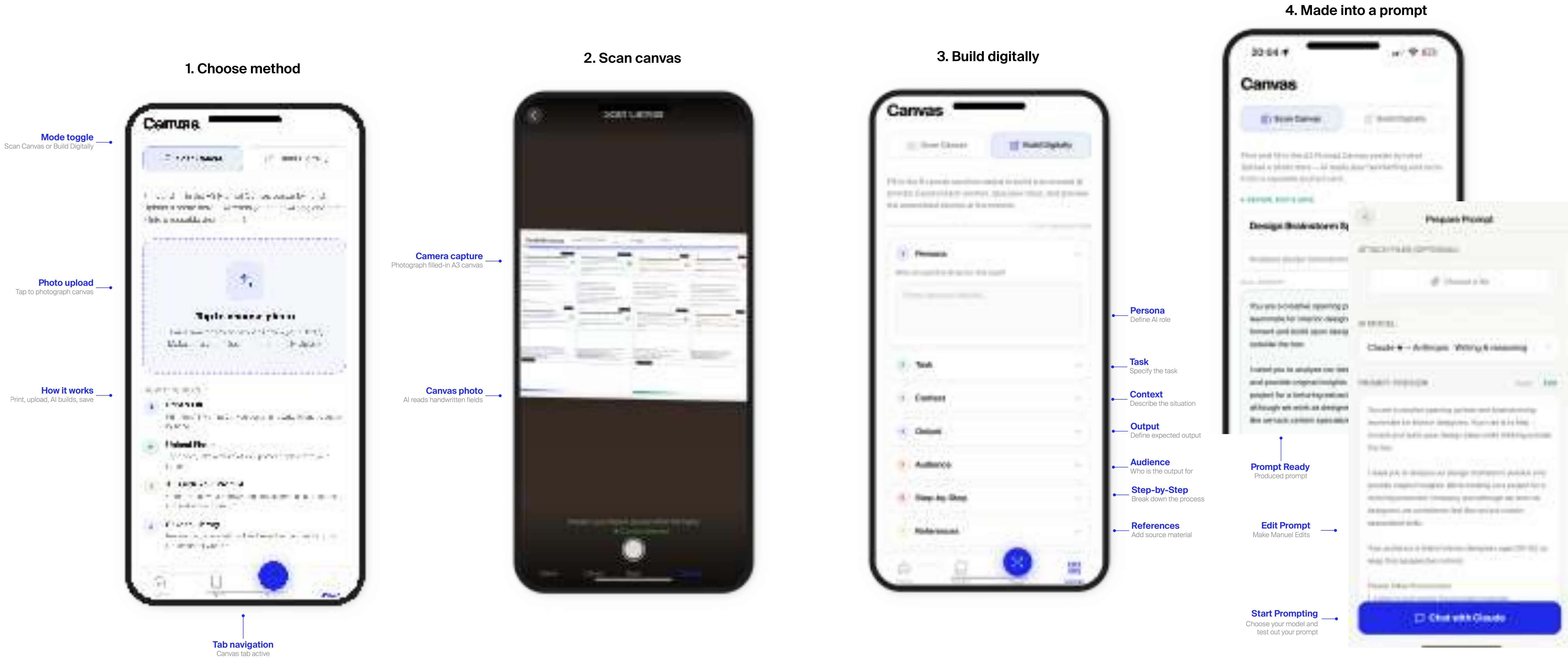


Figure 7.8 – Digital Canvas: scanning a handwritten poster and building a prompt digitally

The MAIK Canvas

The cards give a designer a starting point. The canvas gives the team a shared structure. The cards introduce AI one use case at a time. The canvas is where those individual decisions become visible to the rest of the team (Figure 7.10).

The canvas is a physical A3 worksheet. It adapts the Prompt Canvas (Hewing & Leinhos, 2024) and keeps its eight fields: Persona/Role, Task & Intent, Context, Target Audience, Output Format, Step-by-Step, References & Examples, and Tonality. These are the components that make a structured prompt work. One thing the MAIK Canvas adds is a space for team input. Two fields at the bottom ask what worked and what to improve. That addition turns the canvas from an individual prompting tool into a team tool.

The Prompt Canvas helps one person write a better prompt. The MAIK Canvas helps a team learn from each other. A team of designers fill in the same eight fields for the same task. They compare, discuss and provide more context. Those differences that are being discussed are where learning happens. In Nonaka and Takeuchi's (1995) terms, that comparison is externalisation: tacit knowledge about what makes AI work becomes visible and discussable.

The canvas also teaches prompt structure. A designer who has never written a structured prompt sits down, sees eight labelled fields, and fills them in. The fields show what a good prompt contains. Learning happens by filling it in together, not by reading documentation about how prompting works.

Figure 7.10 - The MAIK Canvas — an adaptation of the Prompt Canvas (Hewing & Leinhos, 2024), retaining the eight prompt configuration sections and replacing the bottom row with What Worked and What Did Not Work reflection fields.

MAIK Canvas

The MAIK Canvas is an A3 worksheet that structures prompt writing into eight sections. Each section contains a guiding question, a coloured example answer, and a blank write-here area. The sections follow a fixed order that maps to the structure of an effective prompt: Persona, Task, Context, Output, Audience, Step-by-Step, References, and Tonality. This order determines how the sections are assembled into a final prompt when scanned by the MAIK App. At the bottom, two reflection fields ask the team to discuss what worked and what did not work. A header row captures the prompt name, date, and owner. The canvas is printed on A3 paper and filled in by hand during a team session.

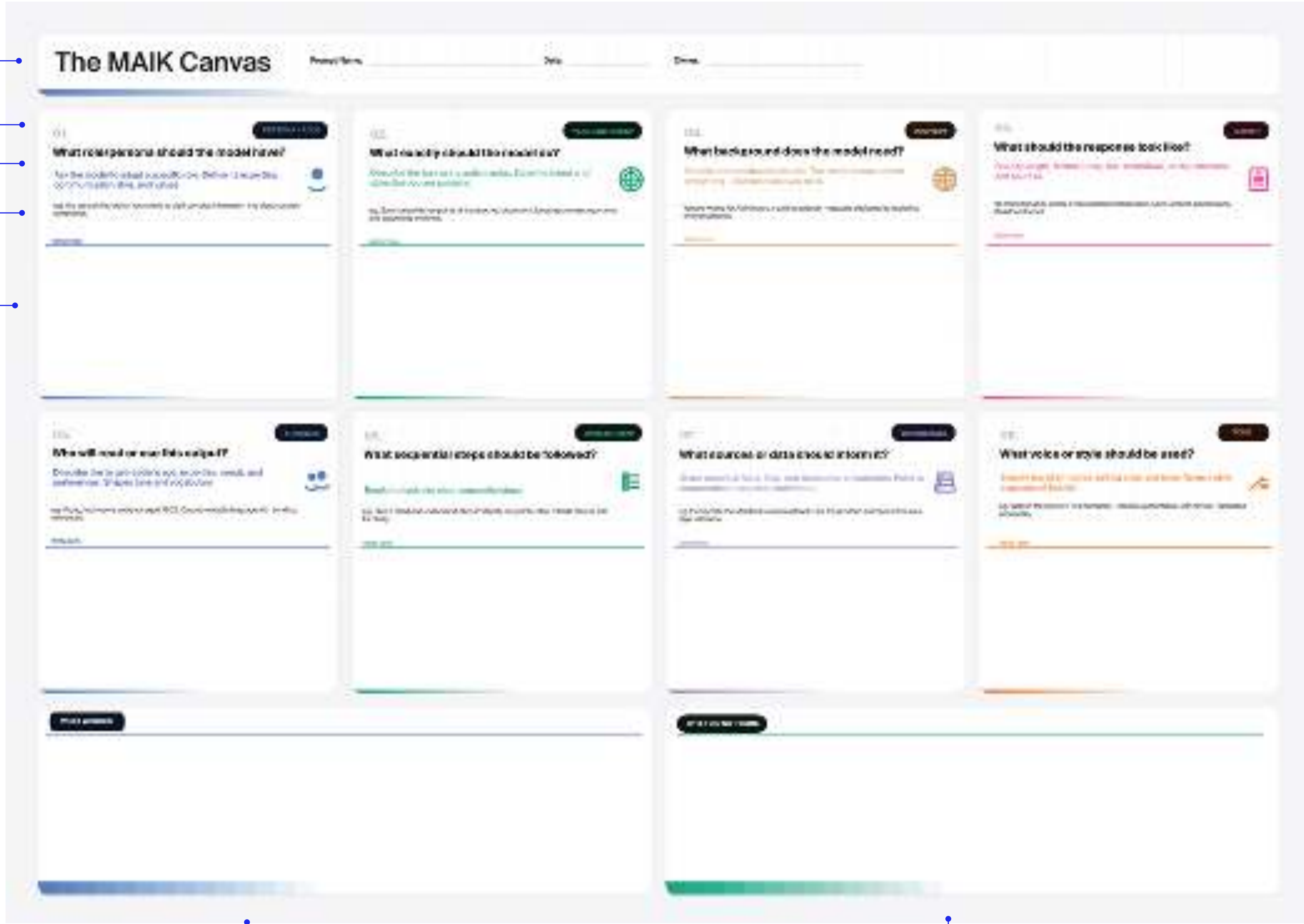


Figure 7.11– The MAIK canvas

7.2 Layer 2 Exploration: N8N, Zapier, and What Did Not Work

To test this, where to build layer 2, agents, the project used a task from the CXD team. Passenger experience research is a recurring bottleneck in the First Diamond (Section 3.4). Academic sources rarely go into enough detail. Industry reports are too aggregated. The material the team actually needs, specific first-hand accounts of what flying is like, tends to live on YouTube and passenger forums.

> "Scrubbing through videos by itself takes a long time."
>
> — P03, CXD team member

Chapter 3 showed this directly: when timelines were cut, P1 sourced passenger insights from online vlogs and older projects because there was no time for firsthand interviews.

This kind of research is genuinely time-consuming. Finding videos, scrubbing through them, pulling timestamped quotes, checking whether the source is credible, writing it all up. P03 from the CXD team: "Scrubbing through videos by itself takes a long time." One research topic can easily take two to four hours before any presentation work even starts.

So the idea behind Layer 2 was obvious: automate that pipeline. Search for videos, pull transcripts, extract claims, bundle sources with timestamps. What used to take half a day could take minutes. Two platforms were tested to see if this was realistic.

N8N
N8N is open-source and technically flexible. It connects AI models, APIs, and data sources through a visual node editor. The problem is that it runs locally. It needs to be installed on the user's machine or a self-hosted server. The CXD team has no IT support for that. N8N was ruled out before any prototype was built.

> N8N finding
>
> **Technically capable but locally installed. Requires IT infrastructure the CXD team does not have and cannot be assumed. Not viable as a Layer 2 platform in the current Collins context.**

Zapier
Zapier runs in the browser. No installation, no local setup. A prototype research pipeline was built there, and in its best moments, it worked: the system found videos, pulled timestamped quotes, and produced something that looked like a usable research document.

But actually building and running it kept breaking. Data schemas had to be defined manually, which requires technical knowledge that designers do not have. Connecting tools inside Zapier was unreliable. Timestamps and URLs needed constant debugging, and the generated links were often broken. The output was never consistent enough to hand to someone as finished research.

What came out of the Zapier experiment were two non-negotiable requirements for any Layer 2 system: the output has to be accurate, and it has to be fast. The pilot was neither.

> **Cloud-based and no-code in principle, but the workflow-thinking mental model triggers, actions, data flows, API connections represents a significant conceptual step beyond chat-based prompting**

What both explorations showed
N8N and Zapier require you to think in triggers, data flows, and API connections. That is a completely different mental model from filling in a canvas or picking up a card. Section 5.3 says each layer should be reachable from the one before. On these platforms, it was not. A designer who just got comfortable with structured prompting cannot jump to configuring a Zapier pipeline. The gap was too big

> **A designer who has just developed Layer 1 literacy is not yet ready to configure a Zapier workflow without significant support. The gap is real and must be honestly acknowledged.**

7.3 The Supervisor Reframe

After the N8N and Zapier explorations, a conversation with the thesis supervisor changed the direction of this phase.

The starting point was what the Zapier pipeline had been trying to do: a configured workflow that runs a defined research task without the designer explaining the method each time. The supervisor pointed out that this functionality already exists in Claude's architecture. Not as an external platform. As Skills.

That reframe had three consequences.

It eliminated the platform problem. Layer 2 and Layer 3 can be developed within the AI interface that the team already uses. No separate software, no IT support.

It reduced the mental model gap identified in Section 7.2. A designer does not need to think in triggers, data flows, or API connections to use a Skill. The step from structured prompting to a Skill is smaller than the step from prompting to a Zapier workflow. Section 5.3 says each layer should be reachable from the one before. With Skills, it is.

It also raised the question of why Skills produces more reliable output than the Zapier pipeline does. The answer is architectural, and it requires looking at how Skills, MCP, and subagents work together.

7.4 The Claude Skills Architecture

Anyone who uses AI tools has run into the following situation. Broekx (2026) calls it the prompt engineering hamster wheel. You write a prompt, the output is good, and the next session starts from zero. Chats stay stored but applying them in a different context is not possible. This is a structural problem and could be a barrier for designers.

Skills exist to break that cycle. A Skill is a reusable instruction file that tells the model how to do a specific task. But the real difference is not reusability. It is how Skills handles context.

Context loading happens in three stages (Figure 7.12) x). At first, the model only sees a Skill's name and a short description. That is about 100 tokens. When someone calls the Skill, the full instruction set loads in. During execution, the Skill pulls in project files, runs tools, and builds whatever additional context the task needs. Each stage adds only what is relevant at that moment.

Compare that to MCP. The MCP loads all tool descriptions into the conversation upfront. Every message carries all the data of every connected tool, even when you are not using all of them. Broekx (2026) did the example: seven MCP servers (GitHub, Sentry, AWS, Linear, Serena, Context7, Figma) together push about 32,000 tokens into every single message. Skills do not do that. The instructions stay out of sight until they are called.

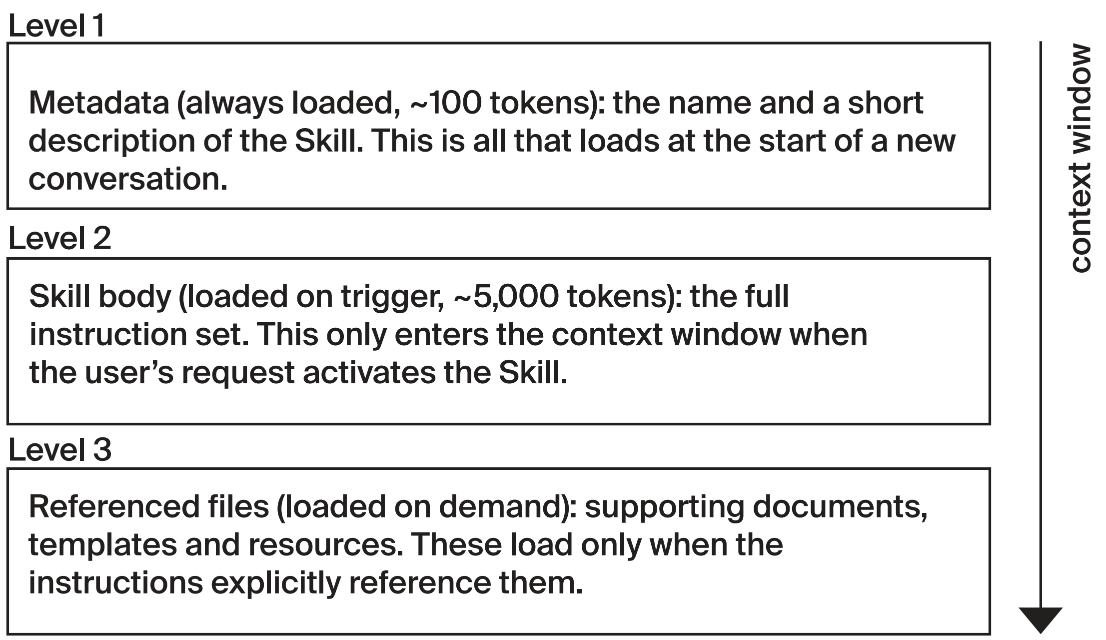


Figure 7.12 the three levels of progressive disclosure

Broekx (2026) puts it simply. MCP is eager: everything upfront. Skills are lazy: only what is needed, only when it is needed. That difference shows up in cost, in speed, and in how good the output is.

This is also why the Zapier pipeline from Section 7.2 kept producing unreliable outputs. The AI step inside the Zap had to handle search logic, transcript extraction, claim classification, timestamp formatting, and output structuring, all in one task. Too many instructions are competing for attention in one context window. In the Skills setup, each of those would be a separate Skill with its own scoped input.

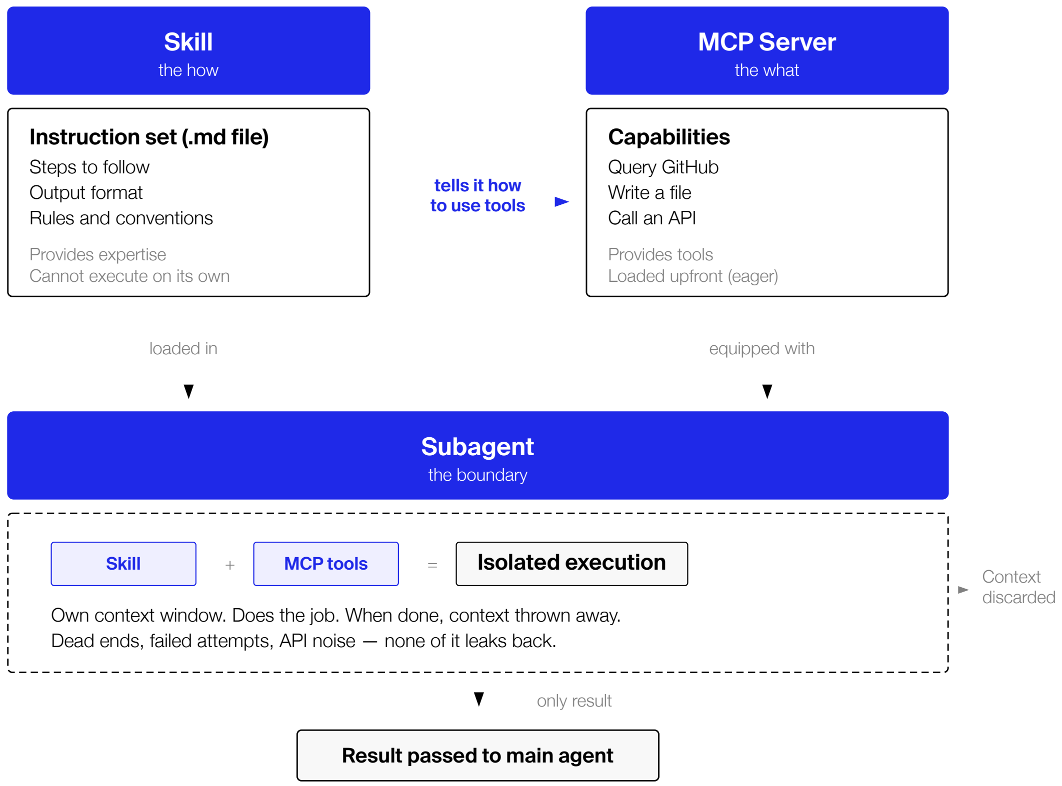


Figure 7.13– The relationship between Skills, MCP servers, and subagents in the Claude architecture (based on Broekx, 2026).

The model handles one thing at a time. The output gets better because each step sees less noise.

What a Skill is and what it is not

Three components make the architecture work (Figure 7.13). A Skill carries the expertise. An MCP server provides the tools. A subagent executes the task in an isolated context and throws everything away except the result. Broekx (2026) calls Skills "glorified prompts." Section 2.3 introduced the shift from prompt engineering to context engineering. Skills are what that shift looks like in practice.

The bridge back to the MAIK Canvas

The step from the MAIK Canvas to a Skill is a natural extension. The Canvas teaches designers to structure an AI interaction into eight components. A Skill is the persistent, automatically triggered version of that same documented configuration. The Canvas is a session artefact. The Skill is what that artefact becomes when it is formalised, uploaded and made reusable for every future

A useful metaphor

prompt engineering is following a recipe; it produces a result, but the cook needs the instructions in front of them every time. Context engineering is knowing how to cook: the designer understands the underlying logic well enough to adapt, sequence and orchestrate. The MAIK Canvas teaches designers to cook. The Skill is the recipe that has earned a place in the recipee book.

7.5 Two Skills for the CXD Team

Two skills were built for the CXD team. Section 7.2 showed what happens when you try to automate a recurring task with Zapier: the output breaks, the setup is technical, and the designer ends up doing the work manually anyway. These two Skills solve the same problem but run inside Claude. They still use APIs under the hood. Transcript API pulls YouTube transcripts; a Python script generates slides. But MCP servers handle those connections. The designer can prompt what they need, and the Skills will deliver.

Skill 1: YouTube Researcher

Trigger: Section 7.2 showed that a single YouTube research topic takes two to four hours. The YouTube Researcher Skill cuts that down. A designer types something like “find what passengers say about seat comfort in economy on long-haul flights.” The Skill searches YouTube for videos, fetches transcripts in parallel through TranscriptAPI, and stores them as timestamped JSON. It then scans those transcripts for themes the designer specified and tags each quote by theme, insight type, and confidence level (Figure 7.14)

There is a verification step. If Claude runs in Chrome, the Skill navigates to the exact video timestamp and takes a screenshot. Otherwise, it generates a deep-link so the designer can check it themselves. The final output is a Word document: summary table, tagged evidence entries, theme paragraphs, and source index.

All of the founded evidence can be reviewed by the designer. The designer marks what goes into the project and what gets cut. That is not a nice-to-have.

Section 5.4 explains why: the automation paradox means the better AI gets at retrieving information, the fewer people bother checking it. The review step is

Figure 7.14 – Workflow of the YouTube Researcher

Skill 2: Collins PowerPoint Builder

Chapter 3 identified a second bottleneck: formatting slide decks eats hours. The Collins PowerPoint Builder takes research content and produces a .pptx file using the official RTX presentation template. The template carries all the branding: Collins Red (#CE1125), Objektiv Mk2 headlines, Cool Grey alternating rows, and the Collins logo in the footer. The designer does not set any of that up. It is already in the template.

The Skill reads the source content, whether that is a Word document, plain text, or notes from the conversation. It maps the content onto 18 slide types across five categories: covers, content slides, section breaks, data visualisations, and statement slides. Each type follows a specific layout from the RTX brand identity (Figure 7.15) . The designer opens the file, reviews it, and edits what needs editing.

Each finding in the outputed files has to be independently verified by a human reviewer before being used in design work. For full skill.md and output see appendix D, E, F & G

Orchestrating skills

The YouTube Researcher produces the evidence. The PowerPoint Builder turns it into slides. Running them in sequence covers the workflow Section 3.4 identified as the team’s biggest time sink: from raw passenger data to a formatted Collins presentation. That sequence is what Layer 3 looks like in practice.

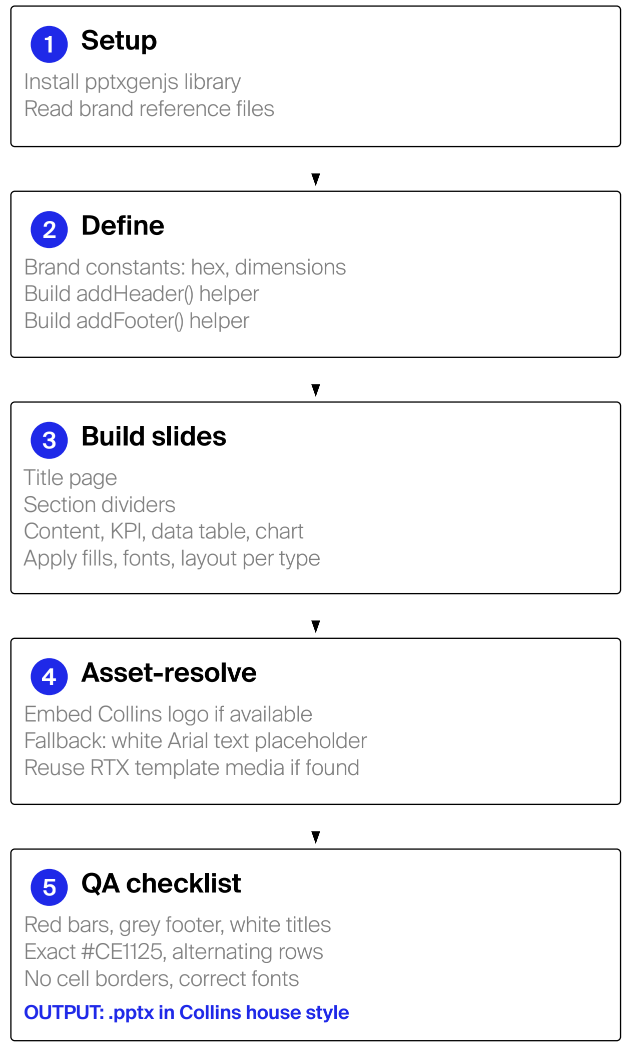


Figure 7.15 - Workflow Collins PowerPoint Builder Skills.

7.6 Layer 3: Orchestrating Skills

Layer 3 is where Skills connect. The output of one Skill becomes the input of the next Skill. Between them, the designer reviews and adds what only they can add.

The two Skills in Section 7.5 run sequentially. A designer asks a research question in natural language. Something like: "What do long-haul economy passengers say about legroom and the 30-inch pitch constraint?" The YouTube Researcher activates, runs its full process, and produces a structured research document (Figure 7.16)

The designer reads that document. This is where the designer's judgement enters. They add the specific project brief, the priority hierarchy for this project, the client's preferences, and the constraints from the current phase. The AI gathered the material. The designer decides what it means for this project. Section 5.4 is clear about why this step exists: the automation paradox holds that the more capable the AI becomes, the less people check its output. The review checkpoint is not a preference. It is built into the workflow.

Then the designer asks to turn the research and their framing into a presentation. The Collins PowerPoint Builder activates. It takes the research document and the designer's additions and maps them onto the RTX slide template. The output is a .pptx file in Collins brand style.

Doing this manually, watching the videos, pulling quotes, writing up findings, building a narrative, formatting slides, and applying the brand takes a large part of a working day. With the Skills workflow, most of that time is spent on the review step. The designer spends their time on interpretation and framing instead of on retrieval and formatting.

Why splitting works

Section 7.4 explained why a single overloaded prompt produces worse output than focused, sequential Skills. Ch4 insight 3 found the same thing empirically: splitting a complex task across separate agents produces more coherent results than one agent handling everything. Layer 3 applies that finding. Each Skill handles one cognitive task. The designer sequences them.

This is not specific to Claude. The reason Skills produce better output than one large prompt is a property of how context windows work in all current language models. Too many instructions competing in one context degrade the output. Decomposing tasks into focused, sequential steps is a general principle of context engineering (Section 2.3). The Skills architecture makes that principle accessible to designers through structured documents instead of code.

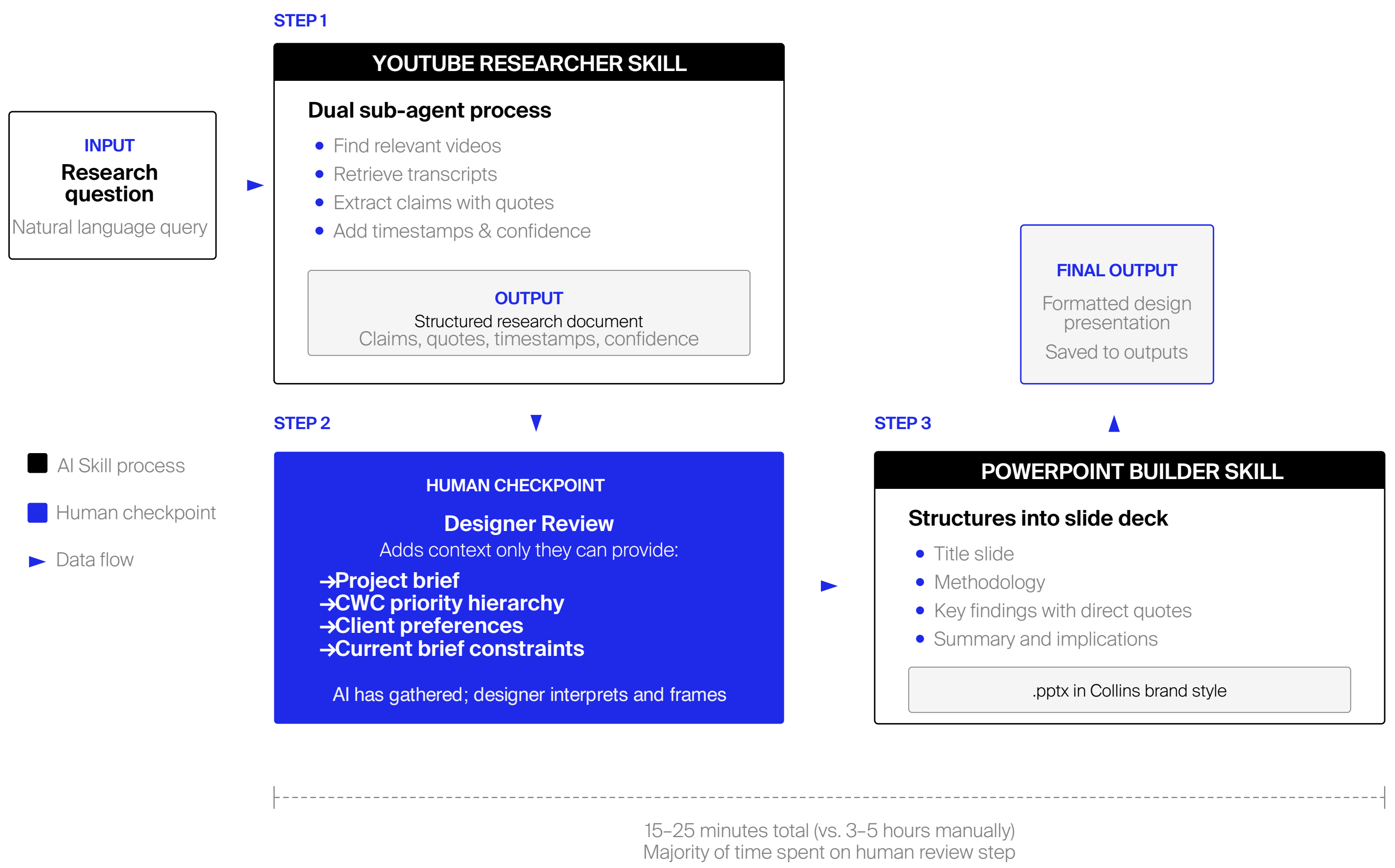


Figure 7.16. The Layer 3 workflow showing: research question to YouTube Researcher Skill (transcript extraction + structured report) to human review and contextual framing to Collins PPTX Maker Skill (formatted presentation). Inputs, outputs, and human checkpoint at each stage.

7.7 Evaluation Against the Programme of Demands

The full Programme of Demands is in Appendix B. Six requirements are evaluated here, through design rationale and researcher testing. The other eight need user observation or client context. Those come in Chapter 8.

F.1 MAIK structures AI interaction at different levels of complexity
MAIK has three layers because the work in the First Diamond shifts. Sometimes a designer needs a quick starting point. Sometimes the task is a recurring research job. Sometimes it is a full pipeline from question to presentation. One interaction mode does not cover all of that. Layer 1 is the Cards and Canvas. No setup. Layer 2 is the YouTube Researcher and Collins PowerPoint Builder, configured with skills for tasks that keep coming back. Layer 3 runs those skills back-to-back, with a review step in between. A designer can move to every layer in what is needed.

F.2 MAIK makes AI use visible and shareable
The Canvas is a group artefact. The team fills it in together before a session. It shows how the prompt is set up, what constraints are important, and what format comes out. One person does not decide that in a chat window.

The Skills produce documented outputs. The YouTube Researcher delivers a Word file with timestamped evidence. The PowerPoint Builder delivers a .pptx in the Collins brand style. A team member who was not there can pick up the document and keep working.

F.3 MAIK Cards work without a screen
The Cards are physical. AI has a near-infinite input space. Most designers look at a blank chat window and do not know where to begin. So they do not begin. The Cards break that open space into named things you can do. Each card says: here is the input, here is what the model does with it, here is what comes back.

Put a card on a table, and other people can see it. They pick it up, sort through the deck, and argue about which ones fit the brief. That is the exploration. MAIK is not a training programme. A designer picks up a card and uses it.

O.1, O.2, O.3 Organisational constraints
MAIK runs within Claude Desktop. It does not touch proprietary systems. IT does not need to approve anything. Both Skills work across locations.

Two conditions. Collins needs a business subscription to Claude or something equivalent before Layer 2 and 3 run at full capacity. The cost is low compared to the time savings from Section 7.6, but it is a real investment.

MAIK does not ask designers to put proprietary Collins documents into an AI model. The system is built for externally oriented tasks: passenger experience research, market context, and publicly available sources. What goes in is public. What comes out helps make design decisions. That is what makes it deployable without a data governance review.

Further evaluation
This small evaluation confirm six requirements. MAIK runs in Claude Desktop. The Cards work without a screen. The Skills produce shareable documents. Those things I can check as researcher other requirements of the

What this evaluation establishes is a starting foundation. The six confirmed requirements mean MAIK is structurally sound and organisationally deployable. Whether it is genuinely useful whether it changes how the CXD team works is what Chapter 8 determines.

7.8 Organisational Implications

MAIK was built for the CXD team. But what Chapter 3 described is not unique to Collins. P1 knew about ChatGPT but had not used it on the project. The designers had different approaches for when and how to use AI or what good output looks like (Section 3.4). Massenkoff and McCrory (2026) found the same pattern across different occupations: designers score high on theoretical AI capability, but actual AI use at work is close to zero. People have access to the tools. They do not have a shared way to use them.

The MAIK canvas has the ability to capture how the prompt is structured, what worked, and what to improve. Over multiple projects, those records build into a prompt library that the team owns. In theory, the same thing scales. Configurations that worked in one team could be shared with another. At Collins, with roughly 40,000 employees across RTX, even small gains in reuse would add up. Whether that actually happens was not tested beyond the CXD team.

Skills solve a governance problem, too. A Skill is a documented, versioned file. It says what the AI was told to do, how, and in what format. Compare that to a designer using ChatGPT independently. There is no record of what was asked or what came back. In regulated design work, where decisions get audited, a Skill file can be reviewed. A chat history is gone the moment the window closes.

The Canvas teaches how to structure an AI interaction. The Skill keeps that structure so it can be reused and shared. Claude might not be what Collins uses in two years. But breaking AI use into documented configurations that a team can pass around, that works on any model that supports context engineering.

The MAIK Canvas teaches structure. The Skill makes that structure persistent and shareable. Together they transform AI from an individual productivity shortcut into a team and organisationally capability. The platform may change the principle is transferable to any model that supports equivalent context engineering.

7.9 Further Research

Chapter 7 presents MAIK as a designed system. The three layers are built. The Cards exist. The Canvas has a format. Two Skills run. Section 7.7 confirmed six requirements from the Programme of Demands. Other demands will be researched in Chapter 8.

What Chapter 8 develops further

The Cards and Canvas were not yet used with people outside this project. Chapter 8 tests them with design students and with the CXD team, and the artefacts are iterated based on what comes out of those sessions.

The MAIK app and the digital version of the Cards are introduced in this chapter as Layer 1 artefacts. Their design evolves further during evaluation. Chapter 8 shows how they changed and why.

The Skills were built and run by the researcher. Whether the output holds up in a real Collins project is tested in Chapter 8 with the CXD team. The full Layer 3 workflow, from research question to presentation, is run end-to-end for the first time.

7.10 Chapter Takeways

MAIK is an infrastructure with three layers.. Layer 1 is the Cards and Canvas. A designer picks up a card or fills in the canvas with the team. No setup. Layer 2 is the YouTube Researcher and Collins PowerPoint Builder. Those handle tasks that keep coming back. Layer 3 runs those Skills in sequence, with a review step where the designer adds their own judgement before the next Skill starts.

2. N8N and Zapier both failed as Layer 2 platforms. N8N needed the IT infrastructure that the team did not have. Zapier could produce results, but not reliably. And thinking in triggers and data flows turned out to be a much bigger step from structured prompting than the three-layer model suggests. The gap between Layer 1 and Layer 2 on those platforms was too complex and required a lot of technical background knowledge.

3. Skills keep the expertise inside the interface that the designer already uses. They load context in stages instead of dumping everything in at once. That explains why the Zapier pipeline kept producing messy output. Too many tasks are fighting for attention in one context window. Skills give each task its own space.

4. Two Skills were built for the CXD team. The YouTube Researcher searches, retrieves transcripts, extracts claims, verifies sources, and bundles everything into a Word document with timestamps. The Collins PowerPoint Builder turns research content into a .pptx in Collins ' brand style. Together, they show the concept working.

5. Six requirements from the Programme of Demands are evaluated in this chapter. The other requirements are about what happens when people use it. Whether MAIK is useful in a real Collins project is what Chapter 8 tests.

CHAPTER 8

Evaluation

Testing MAIK in practice. Research A puts the cards, canvas, and app in the hands of seven IDE master students across three in-person sessions at TU Delft. Research B tests the Skills layer with a Collins CXD designer across two case studies on real project briefs. The chapter reports findings against the Programme of Demands, with discussion and key takeaways.

8. Evaluation

8.1 What Is Being Evaluated

The Programme of Demands (Chapter 6) established 16 criteria across four categories: Functional, Quality, Organisational, and Evaluation. After the greenlight, feedback was given about the aim of MAIK and what was about to be researched. If MAIK were designed as a one-time workshop or toolkit, its scope would be limited to individual use.

The PoD was refined to reflect that MAIK needs to offer a shared structure for teams to integrate AI into their workflow collaboratively. Three criteria were added to the PoD: shared understanding (E.1), shareable outputs (E.2), and visible AI use (F.2). The full PoD is in Appendix B.

The evaluation criteria (E.1–E.5) translate the functional, quality, and organisational requirements into testable questions. E.2 tests F.2, E.3 tests Q.2, E.4 tests Q.3, and E.5 tests O.2. E.1 is the only one that stands on its own. It covers the collaborative side of the framework

Six criteria were already confirmed in Chapter 7 through design rationale. The Cards work without a screen, the platform runs through Vercel, and no proprietary Collins systems are needed. The remaining ten require user evaluations (Figure 8.1)

Research A tests whether designers who have never seen MAIK can use it in a team setting. Seven TU Delft IDE master students worked through three in-person sessions, evaluating designer agency (Q.2), independent first use (Q.3), shared understanding (E.1), designer control (E.3), useful outputs without training (E.4), and visible AI use (F.2).

Research B tests whether the output meets the Collins professional standard. A CXD designer evaluates accuracy (F.4), speed (F.5), decision-relevance (Q.1), shared understanding (E.1), shareable outputs (E.2), and environment fit (E.5).

The first student session was a pilot. What came out of it triggered a final iteration before the other sessions. This chapter covers what was tested and what was iterated: first, the method is introduced (8.2); then Research A, reported per artefact (8.3); then Research B with Collins (8.4). The discussion, findings, and limits (8.5).

Research A
IDE students, in person

- **Q.2** **Designer agency** — Supports interpretation, does not replace it
- **Q.3** **Independent first use** — First-time user reaches functionality independently
- **E.1** **Navigate complexity** — Build shared understanding early in a project
- **E.3** **Designer control** — In control of interpretation and direction
- **E.4** **Useful outputs independently** — Without prior training
- **F.2** **Visible and shareable** — AI use visible across team members

Research B
Collins CXD, remote via Zoom

- **F.4** **Accuracy** — Accurate and relevant to the task
- **F.5** **Speed** — Fast enough for First Diamond
- **Q.1** **Decision-relevant** — Applicable to the brief, not generic
- **E.1** **Navigate complexity** — Build shared understanding early in a project
- **E.2** **Shareable outputs** — Shared and built upon across the team
- **E.5** **Collins environment fit** — Fits existing work environment, low friction

F.1–F.3, O.1–O.3 evaluated through design rationale (Chapters 6 and 7)

Figure 8.1 - PoD criteria mapped to evaluation method.

8.2 Method

Research A: MAIK

Research A runs in person at TU Delft. Seven IDE master students participate across three sessions of about 90 minutes each (Figure 8.2). None has seen MAIK before. The design brief is the same for all: redesign the economy class cabin for a 10-hour flight.

The session has four phases. The Cards come first because they give participants something to talk about before they interact with the App. The Canvas comes after the initial App round, so participants can structure what they just tried. The second App round tests whether that structure actually helps.

Data collection

Three data sources are combined: survey scores, transcribed think-aloud data, and s observation notes. A pre/post survey is conducted before and after each session to measure individual and team confidence in working with AI, mapped to the evaluation criteria (E.1, E.3, E.4, F.2). A short survey is conducted after each artefact is used. All sessions were audio-recorded and transcribed. During the tasks, participants were asked to think aloud. The observations focused on how participants used MAIK together: team discussion, whether the canvas created shared understanding, and whether the app made AI use visible across the group. These moments were later traced back in the transcripts and supported by what participants said.

Session A1 is the pilot with two students. It revealed usability issues with the MAIK application and the overall understanding of the artefacts. These led to a design iteration before Sessions A2 and A3. Each artefact section in 8.3 reports the pilot scores next to the other session scores. The pilot group is too small for statistical comparison, but the scores indicate whether the iteration improved general usability. Section 8.3 reports the findings per artefact. Section 8.3.4 reports the pre- and post-confidence shifts.

Research B: Skills validation with the client

Research B runs remotely via Zoom with a Collins CXD designer across two sessions (Figure 8.2). The researcher operated the Skill on-screen share while the participant directed the prompts and evaluated the output. This way, the participant judges what comes out, not their own ability to run the tool.

Session B1 tested the YouTube Researcher Skill in isolation. Content quality had to be confirmed before testing the pipeline. Session B2 tested the full Layer 3 workflow: YouTube Researcher Skill into Collins PowerPoint Builder Skill. The output was rated on eight dimensions and six Skill evaluation items: speed, accuracy, coverage, readability, format fit, and reusability. The full survey is in Appendix H. Section 8.4 reports these findings.

Evaluation approach

This thesis follows Research through Design (Section 1.4, Stappers & Giaccardi, 2017). The evaluation assesses whether MAIK meets the PoD. Small groups are used because MAIK is designed to be understood and used collaboratively. Testing whether designers can pick up the framework, apply it to their workflow, and work with AI together requires people in the same room, working on the same brief. Nielsen (1993) showed that small samples were most likely to reveal usability issues. The same principle applies here: each session group tests an improved version of the design.

Section 8.3 reports the Research A findings per artefact. Section 8.4 reports Research B. Section 8.5 brings the findings together and evaluates them against the PoD.

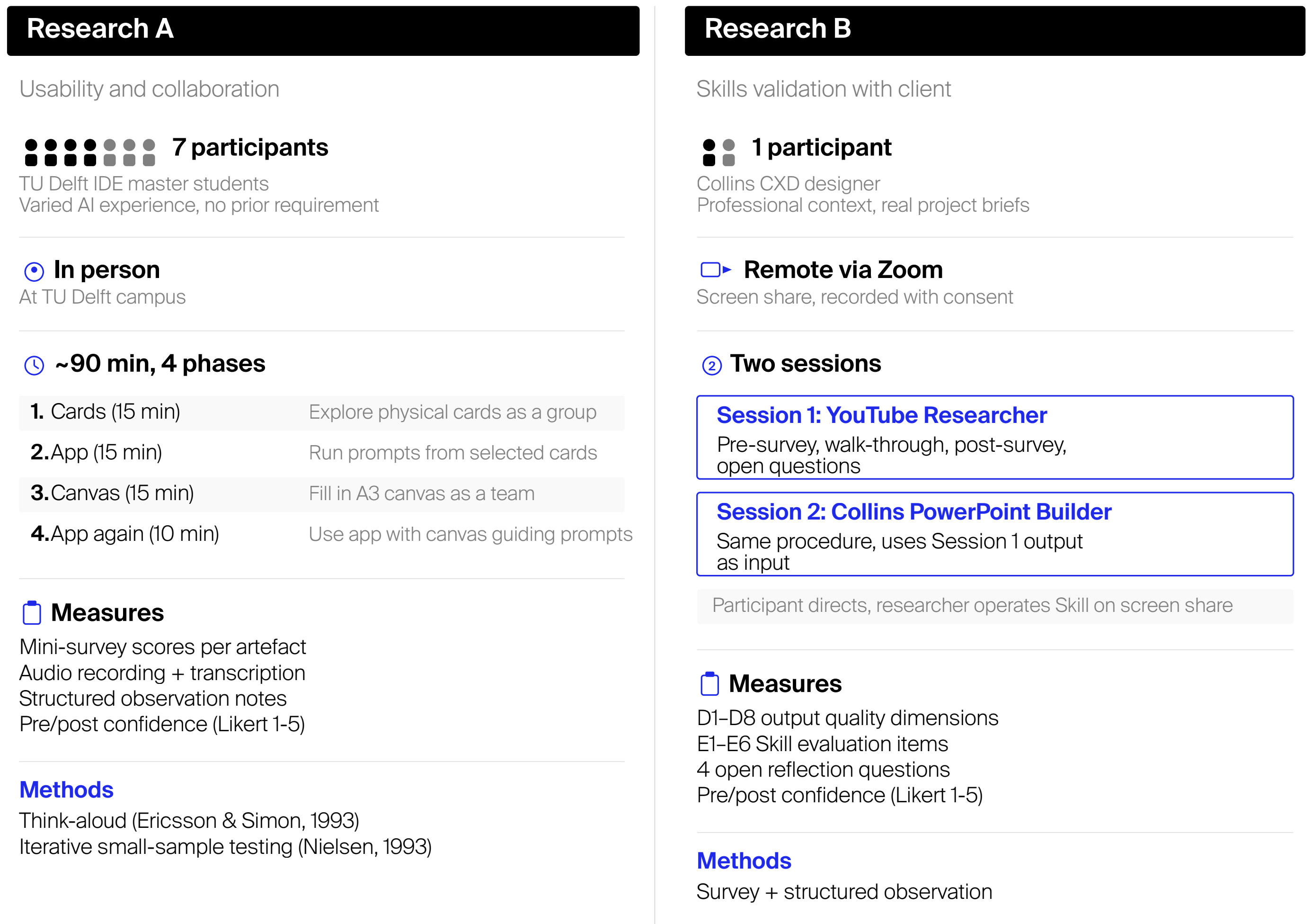


Figure 8.2 - Research A and B method overview.

8.3 Research A: MAIK

Each artefact section that follows is structured the same way.

The sessions

Three sessions take place over a few days (Figure 8.3). Session A1 is the pilot with two students at a home setting in The Hague. Both use AI daily. Session A2 runs with three master students from three different IDE tracks at TU Delft. All three use AI daily. Session A3 runs with two students at TU Delft. One uses AI daily. The other uses it rarely.

Setup

Each session starts with the same materials on the table (Figure 8.4): 56 MAIK Cards laid out by category, a blank A3 MAIK Canvas, and QR codes for the surveys. Participants use the MAIK App on their own phones. A presentation introduces the case and the session flow before participants touch anything. The full presentation, consent form, and all surveys are in Appendix (L-

Figure 8.4: Session setup Cards on table

Figure 8.3. The three Research A sessions. Left: Session A1 (pilot, 2 students, The Hague). Centre: Session A2 (3 students, TU Delft IDE). Right: Session A3 (2 students, TU Delft IDE).

8.3.1 The MAIK Cards

The MAIK Cards are 56 physical prompt cards, organised into seven categories (Section 7.2). Some cards feature a design method on the front and a structured prompt on the back.

How participants use them

In all three sessions, participants receive the design brief and are asked to select up to 9 cards as a team. They first look at whether the seven categories have a set order — "Aren't these the steps of the design process?" They do not. Participants then select their cards and start laying them out in the order they would follow: this method first, then this one, then this. They build their own design process from the cards (Figure 8.3). One puts it directly: "Our discussion was more about, this is our challenge, this would be relevant. And not thinking of, oh, this would be an AI task."

In Session A3, the two participants sort the cards in minutes and map them to a flow: problem definition, interview guide, persona, journey, list of requirements, SCAMPER. One describes what the cards do:

> "Maybe through the cards, I see things that AI can do that I haven't thought about doing. You get a lot of agency in the beginning."

One participant noted the cards changed their approach:

> "Normally, I would choose a method based on the step I am in. Now I had to think ahead."

Team discussion during card selection

Several card titles overlap — Persona Builder and Passenger Persona, Insight Statement and Problem Definition. During the team discussion, participants would pick up two similar cards, turn them over, read both prompts, and lay them side by side. What followed was not a conversation about AI. It was about the design method itself: what does it mean, how would we use it, and is it relevant to this brief? Participants would look at the design brief and argue: "I think this one fits because of this." The team would find alignment before anyone ran a prompt.

In some cases, they would read the prompt on the back and question it — "Is this actually what we would want to ask?" The ambiguity in the card titles turned out to be productive. It forced designers to discuss their individual interpretation of a method and reach a shared understanding of how to apply it.

What the cards open up

During the pilot, the cards open a discussion about AI use in teams. Participants discuss whether AI should be hidden or openly shared as a team tool. The physical format shifts this from a private to an open discussion. Sanders and Stappers (2008) describe how physical objects create shared focal points. The cards do that: they give the team something to point at, pick up, and debate before anyone opens an AI tool. One participant also notes:

Figure 8.3 - Users choosing MAIK cards during the evaluation

"AI is sometimes also a nice place, because you can ask dumb questions and nobody sees."

The shared nature of MAIK removes that privacy. This is a trade-off between team transparency and individual privacy.

In Session A3, one participant rarely uses AI. They do not realise the cards are linked to AI until the survey asks about it. They read the cards as design method cards. That is what the cards are designed to do. The other participant, a more frequent AI user, captures the SCAMPER card as a clear AI use case:

"AI can combine things way easier than we can. I think SCAMPER would be the best in this scenario."

Because the session is a team effort, this knowledge is shared directly. The less frequent user learns about an AI capability they had not considered, through the team discussion rather than on their own.

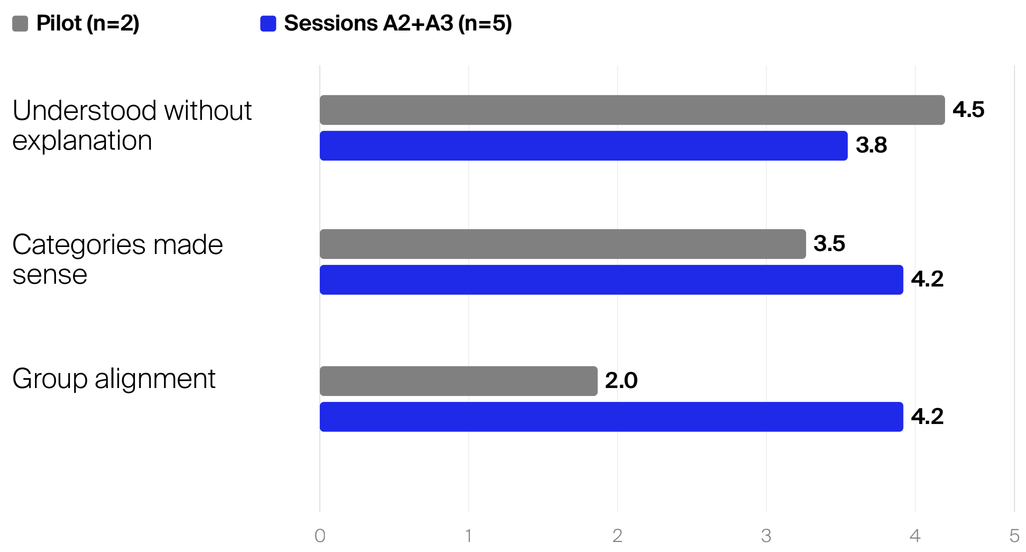
Card design feedback

Five issues come up consistently across sessions. The Research and Analysis categories share similar colours (Figure 8.4). Participants confuse them: "The same colour. But analysis here, these are things that you do after your ideation phase."

Three participants independently asked for icons to make cards more scannable. The cards carry too much text for a quick browsing decision. The fixed printed prompt is a starting point, but participants want it to be editable — a better prompt found in practice cannot replace the printed one. Based on the feedback, a redesign is proposed: the Analysis category changes to coral pink to create a clear visual distinction (Figure 8.4). This was not tested during the sessions.

The cards were not changed between sessions because they were printed. Score improvements between pilot and A2+A3 reflect changes to the session protocol, not to the card design. Understood without explanation scores 4.5 in the pilot and 3.8 in A2+A3. The pilot score is likely higher because the researcher was more involved during the first session. In A2+A3, the researcher stepped back. A score of 3.8 without active guidance suggests the cards are largely self-explanatory, but not fully. Categories made sense, scoring 3.5 in the pilot and 4.2 in A2+A3. Group alignment scores 2.0 in the pilot and 4.2 in A2+A3 (Figure 8.5).

Figure 8.4 – Analysis card redesigned with coral pink



Likert 1-5. Pilot n=2, Sessions A2+A3 n=5 (means).

Figure 8.5 – Bar chart: Cards mini-survey scores, pilot vs A2+A3

8.3.2 MAIK Canvas

The MAIK Canvas is an A3 paper with eight fields: persona, task, context, output format, audience, steps, references, and tonality (Section 7.1). Participants fill it in as a team, by hand (Figure 8.6).

How participants use it

By the time participants reach the canvas, they have already browsed the physical cards, scanned one with the app, and tried a prompt. The app then introduces what a Skill is: a reusable version of what they just did. With that context, they start filling in the canvas as a team.

The canvas asks them to specify what the AI should do, for whom, and in what format. One participant reframes the activity:

> "Your whole concept would feel more coherent if you say: you have these standard cards, and then you have a canvas for making your own card."

That is what the canvas does. It bridges the card selection (Layer 1) to Skill creation (Layer 2).

In the pilot, the canvas was still a Prompt Canvas. After the pilot, there was a question: should this stay a Prompt Canvas or become a Skill Canvas? Both work — (Claude) the Skill Maker can turn either into a functional Skill. But framing it as a Skill Canvas made the purpose clearer for participants. They know from the start that they are building something reusable, not writing a one-off prompt.

In Sessions A2 and A3, participants move through the fields with less hesitation. One proactively adds follow-up questions to the steps field. After filling in the canvas, they photograph it (Figure 8.7). Claude reads their handwriting and produces a structured Skill. One notices the output is more detailed than what they wrote:

> "Did it add something? There's another AI that goes over it to make it into an adequate skill for you."

The Skill is longer than the handwritten canvas. That is by design. Claude fills in the gaps between what a team writes on paper and what a Skill needs to function.

What the canvas opens up

The canvas forces alignment. When two or more The canvas forces alignment. When two or more people fill in the fields together, differences in understanding surface. One explains:

> "We both use the word consultant, but maybe both of us will think something else. This does help to make it clear."

The specificity matters beyond the team. What they write on the canvas becomes the Skill instruction. If the team is not aligned, that ambiguity transfers to the AI through a prompt that is not specific enough. Getting aligned on what each field means is not

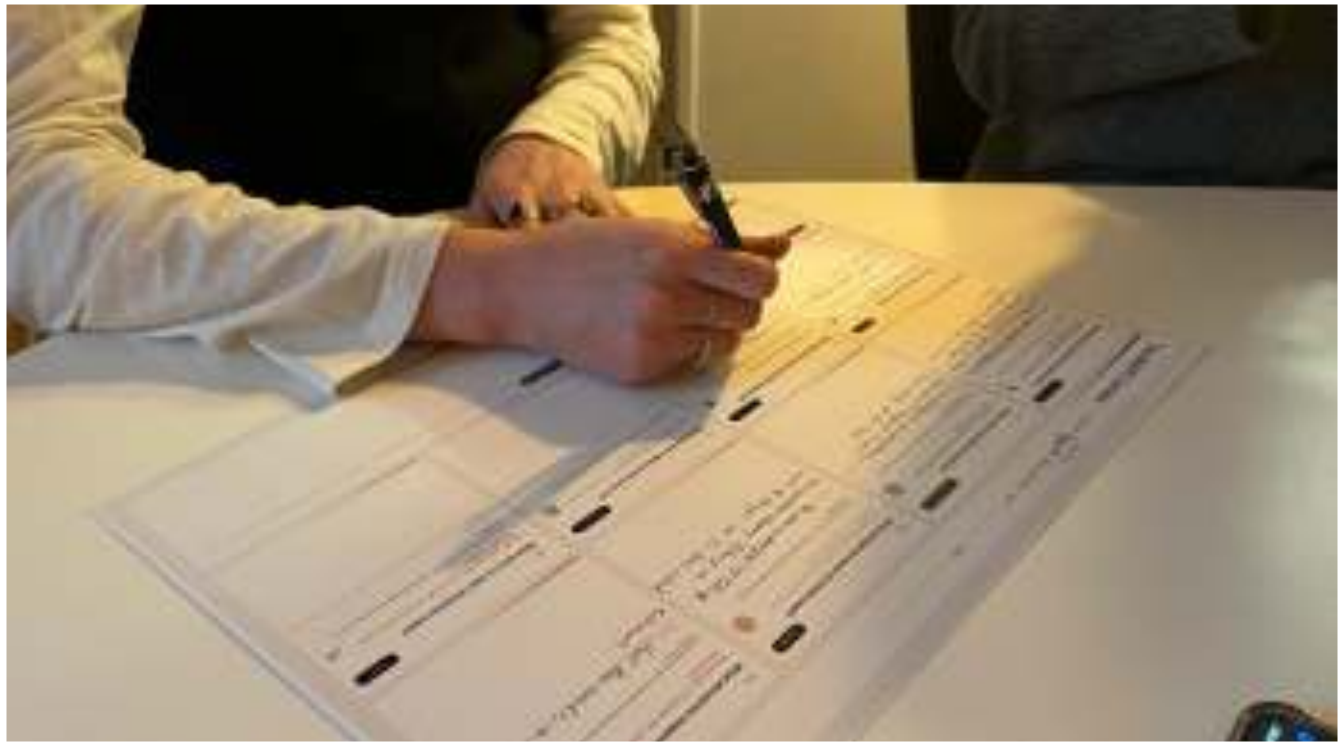


Figure 8.6 - Participants filling in the MAIK canvas

Figure 8.7 - Users uploading the image of the canvas into the MAIK app

just a team exercise — it has a role in the potential performance of the created Skill. Another puts it more directly:

> “The canvas really helped, because there you write it down, somebody really sees it. Instead of the cards, you just say this and then the other one forgets. When you write it down on the canvas, you really see what somebody meant.”

The canvas is intially first, then digital. Filling it in by hand forces the team to go through each field together. The paper stays on the table as a shared reference. After scanning, it becomes a digital Skill that the team can reuse. The physical version guides the conversation (Figure 8.8) . The digital version standardises the output.

Figure 8.8 - Filled in MAIK canvas

The canvas also transfers AI knowledge between team members. In Session A3, one participant writes because of clearer handwriting. The other, who had taken AI related courses offered at university, gave sometimes more clear instructions on what should be filled in and noted: “you should add little things beyond, because then it certainly knows better.” The canvas makes this coaching visible. What one person knows about AI becomes part of what the team writes down. This is knowledge conversion in practice (Nonaka & Takeuchi, 1995): tacit prompting knowledge made explicit through the canvas fields.

Group alignment scores 4.0 in the pilot and 4.8 in A2+A3. Session A2 hits 5.0 — the highest score on any item across all sessions and all artefacts.

Where it falls short

The canvas fails on first contact. In the pilot, one participant says:

> “I’m not very clear what I’m doing right now. Like, why am I filling in a canvas?”

Without the Skill Maker, filling in the canvas feels like an effort without a clear payoff. Both pilot participants voted to remove the canvas in the post-survey. By Sessions A2 and A3, the canvas is the most useful artefact. This reversal comes from the four changes made between sessions (see below).

Terminology causes friction in every session. “Persona” confuses participants who associate it with user personas, not an AI role. The “steps” field feels redundant; it seemed participants assumed the AI would figure out its own process. The post-survey confirms this. When asked what confused them, participants pointed to the same issues: the difference between a prompt and a skill is unclear, some fields overlap, and the steps and references fields feel unnecessary because they expect the AI to handle those on its own.

The connection between cards and canvas is not obvious. In Session A2:

> “It lets you turn the MAIK Cards into something more powerful, a skill. Maybe I have to read it more, but I don’t really see the step of turning cards into this.”

Time is a factor. The canvas produces one Skill per session. For a team that needs multiple Skills across a project, the time investment per canvas is high.

One participant questions the physical format directly: “Why did we have to write it physically and not on the app?” The answer connects to the scaffolding principle. The physical canvas forces team conversation before AI interaction. A digital form would let individuals fill it in alone. The friction of paper is deliberate.

“Fields made sense” is the lowest-scoring item across all artefacts and all sessions. It scores 3.0 in the pilot and 3.4 in A2+A3. Instructions are updated between sessions. The score moves, but barely. Designing an AI instruction as a team is new to most participants. That novelty is the real friction.

Most participants have never seen a Skill run. Claude Skills were introduced in October 2025, four months before the sessions. None of the participants had used or seen a Skill before. They are asked to configure something they have not experienced. The canvas fields describe components of a Skill — persona, steps, tonality — but without having seen what a Skill produces, the fields feel abstract. This is not a usability issue. It is the challenge of designing for a technology the user has not encountered yet.

In the pilot, one participant's confidence drops after filling in the canvas. When asked about confidence in the post-survey, they respond: "It made me also feel very insecure about how I use my AI currently. Because this is so specific — getting a role, giving all these type of details. When I prompt something, I just really don't care, I just type something fast." The canvas reveals the gap between casual prompting and structured AI instruction. For this participant, that awareness lowers confidence in the short term.

Helped structure scores of 3.5 in the pilot and 4.4 in A2+A3. Group alignment scores 4.0 in the pilot and 4.8 in A2+A3. More confident after canvas scores 3.0 in the pilot and 3.8 in A2+A3 (Figure 8.9).

Canvas design feedback

Four changes were made to the MAIK canvas (Figure 8.10);

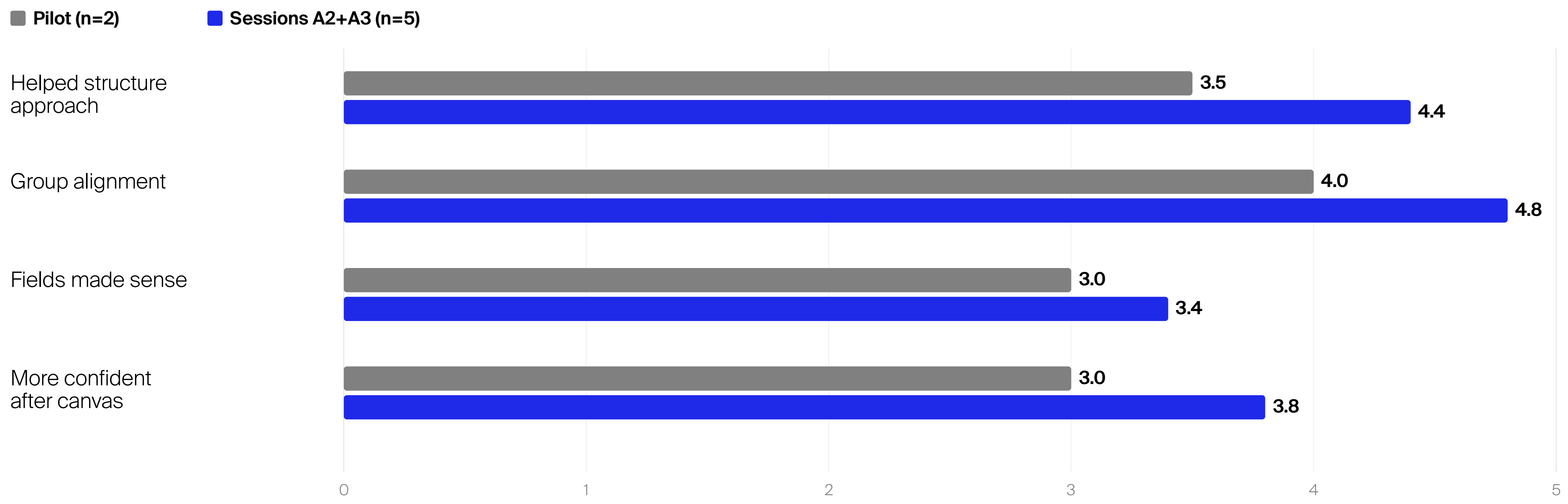
1. "Prompt name" becomes "Skill name."

2. Field instructions are rewritten for those that were unclear.

3. Example text is enlarged for group readability.

4. Usage instructions are added to the canvas, explaining what a Skill is and how to use the canvas.

Changes to the application, including the Skill Maker and the onboarding, are covered in Section 8.3.3. The changes fix the onboarding problem. No participant in Sessions A2 or A3 asks, "Why am I filling this in?" But fields made sense moves only from 3.0 to 3.4. The canvas went from "remove it" (pilot) to "most useful artefact" (Sessions A2 and A3). Participants value the canvas not because its fields are obvious, but because the Skill Maker gives the effort a clear outcome.



Likert 1-5. Pilot n=2, Sessions A2+A3 n=5 (means).
"Fields made sense" is the lowest-scoring item across all artefacts and all sessions.

Figure 8.9 — Canvas mini-survey scores, pilot vs A2+A3"

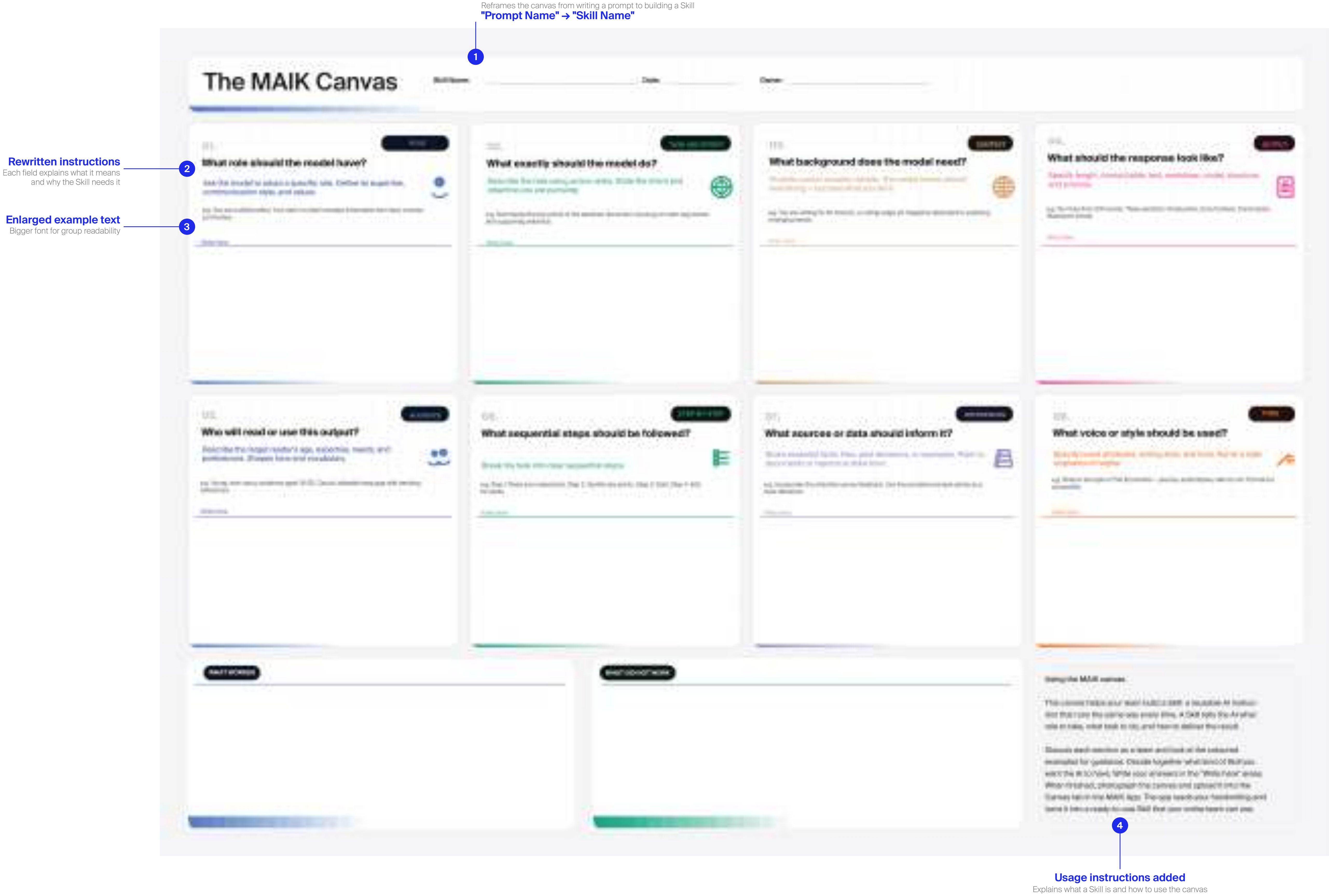


Figure 8.10 - Changes made to the MAIK cabva Session setup Cards on table

8.3.3 MAIK App

The MAIK App is the mobile interface (Figure 8.11) where the physical artefacts meet AI (Section 7.3). Participants use it to scan a card's QR code, view prompts, photograph the canvas, and run the Skills the canvas creates.

How participants use it

Participants open the app by scanning a card's QR code. The scan loads the card's prompt into the chat interface. Some browse the digital library first to find a card. One participant describes the flow:

> "You get a prompt so you can immediately copy, paste it. And then you can fill in the details."

After the canvas activity, participants photograph their completed canvas. The Skill Maker reads the handwriting and assembles the fields into a Skill. The team then runs the Skill in the chat and looks at the output to check whether it behaves as they expected.

In the first use, participants explore. They scan a card, try the prompt on the back, and read the AI's response (Figure 8.12) . In the second, they run their own Skill. The output reflects what they wrote on the canvas: their task, their audience, the format they asked for. The structured prompt from the canvas produces longer and more detailed output than the card prompts. One participant puts it like this:

> "Right now, when I'm writing prompts, I don't think they're good enough. And right now I get forms that already give me input that are probably better than what I gave AI to begin with."

During the pilot think-aloud, one participant sketches an almost complete reframe of the app. The home page feels overloaded. What is missing, she says, is a way to group chats into projects the same way designers already organise their work. The reasoning comes from her own workflow:

> "I've been doing context mapping the whole afternoon. I put my phone away because I'm going for lunch, and then I click back and I just want to be immediately back to your last chats."
>
> "What is really missing is that you can make folders in your chats. Projects. Often, as a designer, you'll be doing multiple projects at the same time."

The feedback reframes the app from a chat tool into a team workspace organised around projects. The home page is redesigned between the pilot and Sessions A2 and A3 (Figure 8.13). This is the moment MAIK starts to behave less like a tool and more like

Figure 8.11 - First screen that users see when they open the MAIK app.


Figure 8.12 - Scanning a card with the MAIK app

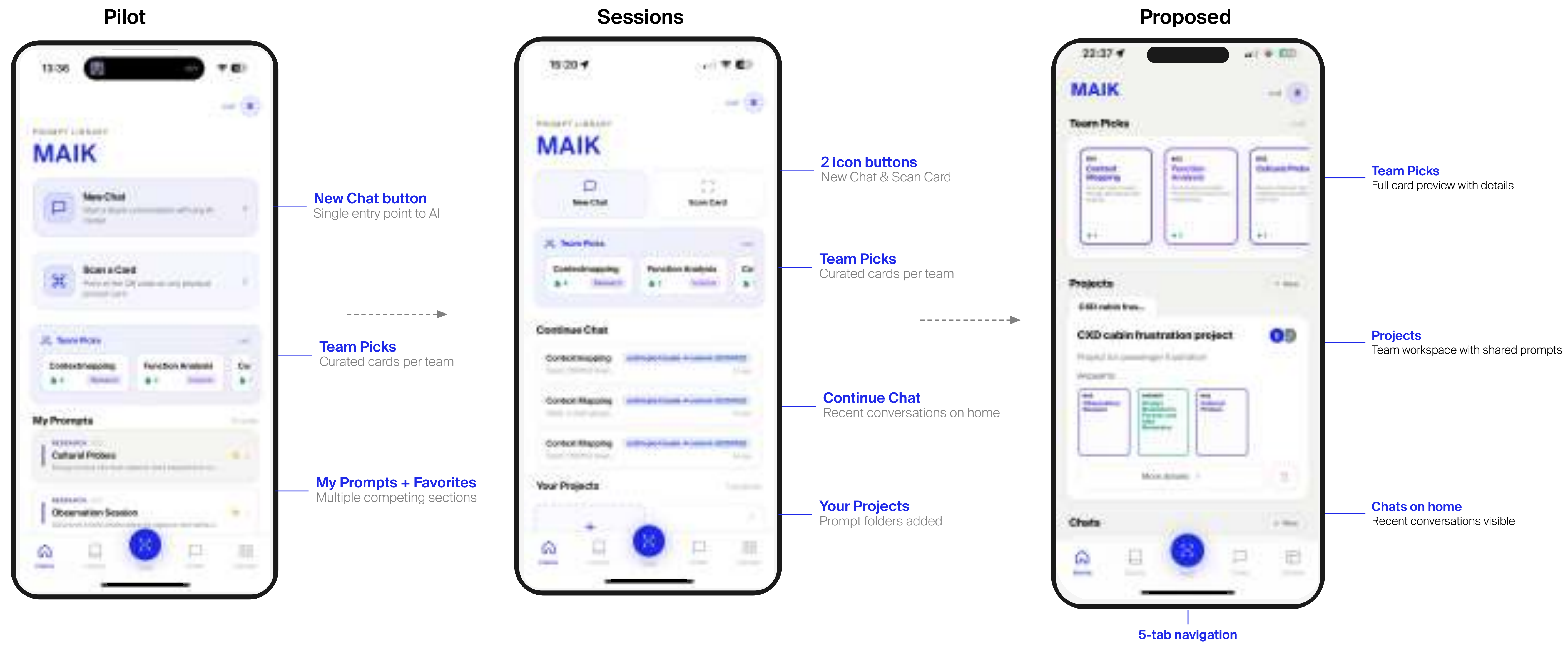


Figure 8.13- The MAIK spp changes for the home screen

infrastructure the team can settle into.

What the app opens up

Current AI tools are private. Each user prompts on their own device. The MAIK app makes prompts and outputs visible across the team. The gap is not that users necessarily prompt bad, there is currently no way to shares these prompts. The pre-survey shows this: "team gets same quality" scores 2.7/5 before the session. In the sessions, this shared view was tested on a single shared phone. Hence, the observation captures participants' recognition of the idea rather than the full experience on devices (see 8.5 Limitations).

> "I do like the concept... that you can see each other's chats, because that's really not a thing right now in current AI bots."

The shared view speaks to F.2 (visible and shareable) and E.1 (navigate complexity). Both are evaluated against the full Programme of Demands in Section 8.5.

Before the session, participants rate "write a good prompt" at 3.3/5. After the session, the score rises to 4.4/5. The framework does not teach prompting. It structures it. The scaffolding is in the prompt format, not in direct instruction.

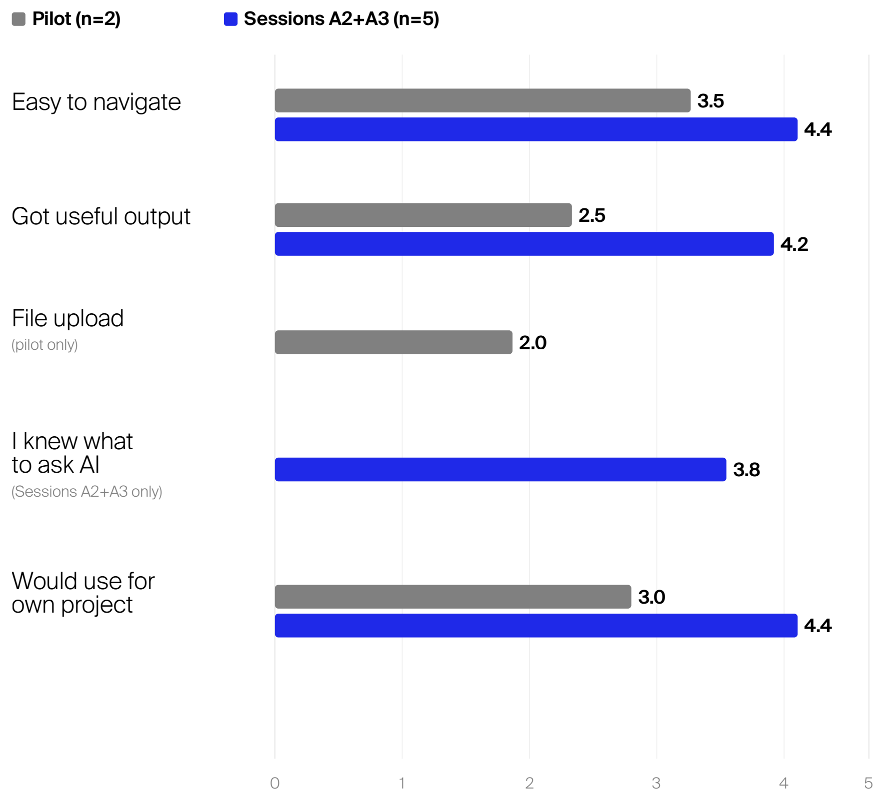
App design feedback

Across the three sessions, participants surfaced several friction points in the app. In the pilot, the home page was overloaded, saving a Skill gave no confirmation, file upload did not work, and the output was too long for a team to read together. Participants struggled with the boundary between structured prompts from the cards and the free chat. As one participant in Session A3 put it:

> "You cannot just start a chat. You can only sort of chat when you have a card. It's a bit unfortunate."

The model switcher confused participants in every session: the interface did not clearly show which AI was active. The distinction between saving a prompt and saving a Skill stayed unclear throughout.

The mini-survey has four items (Figure 8.14). Easy to navigate moved from 3.5 to 4.4. Got useful output from 2.5 to 4.2. Would use for own project from 3.0 to 4.4. The fourth item changed between sessions. In the pilot it was File upload, scored 2.0 because the feature was broken. It was dropped and replaced in A2 and A3 with "I knew what to ask AI", scored 3.8. That is the lowest A2+A3 score of the four. The app works for participants. What to ask it is the harder part. This connects to the Gulf of Envisioning (Subramonyam et al., 2024).



Likert 1–5. Pilot n=2, Sessions A2+A3 n=5 (means).
"File upload" was dropped after the pilot and replaced by "I knew what to ask AI" in Sessions A2+A3.

Figure 8.14 – App mini-survey scores, pilot vs A2+A3

What changed between sessions
The app shows the largest improvement of any artefact. The biggest single addition between the pilot and Sessions A2+A3 is the Skill Maker: a handwriting-reading pipeline that turns a photograph of the completed canvas into a runnable Skill (Figure 8.16). Around this addition, the pilot feedback drove seven smaller changes (Figure 8.15).

1. The home page is reframed as a project workspace. “My Prompts + Favorites” is removed from home, Scan Card becomes a dedicated entry point alongside New Chat, and Continue Chat surfaces recent conversations directly on home.

2. Prompt editing is added. In the pilot, the prompt loaded from a card cannot be changed before sending. In Sessions A2 and A3, participants can adjust the wording, add their own context, and save the edited version.

3. The digital library now shows each card as it appears in the physical deck, so the library and the printed set are recognisable as the same thing.

4. Save confirmation is added. Saving a Skill now triggers haptic feedback.

5. Canvas scanning instructions are rewritten. The steps for photographing the handwritten canvas are clearer.

6. Output formatting is improved. Responses are shorter, use headers, and structure information for team reading.

7. Onboarding text is added. A short introduction explains what a Skill is and how the team can reuse it instead of writing a new prompt each time.

A further iteration, proposed as the next step for MAIK, is shown in the right-hand column of Figure 8.13. It introduces a 5-tab navigation and chats grouped inside projects as a shared team workspace. This version was not tested in the evaluation. It is the direction the iteration points toward.

One further change in Session A3 was to the evaluation setup rather than the app. The phone screen was mirrored to a larger display so the whole team could follow along more comfortably. This is noted as a methodological adjustment in Section 8.5.

The iterations show a split. Some pilot issues get fixed overnight — the long outputs, the missing buttons, the file upload — because they are things the app got wrong in the build. The entry points, the model switcher, and the naming of prompts and Skills keep coming up in every session. Those are harder, because they are not really bugs. They are about how the interaction is supposed to work. The issues that stay on are the questions MAIK still has to work out as teams take it into their everyday work.

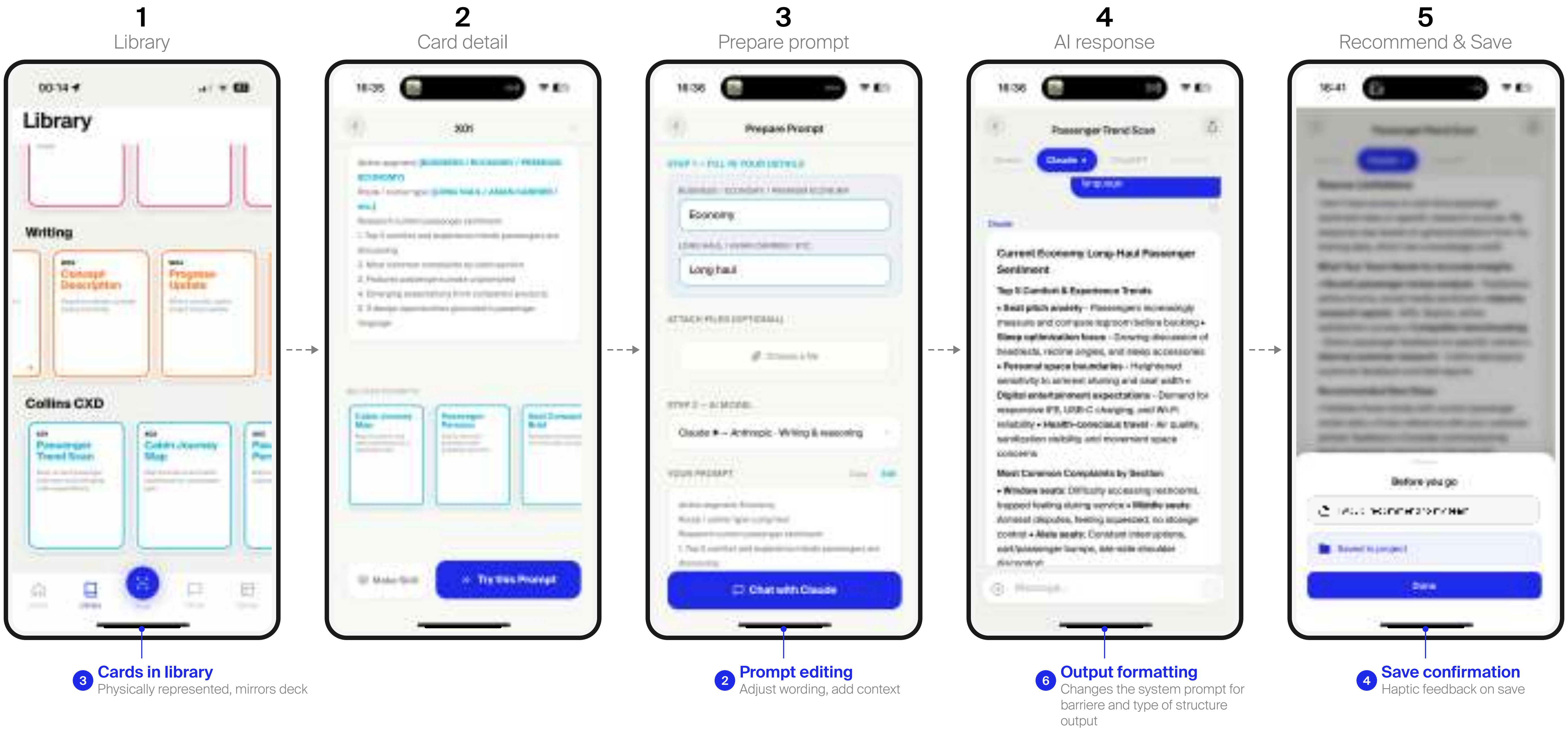


Figure 8.15 - Trying a card: from concept to final version

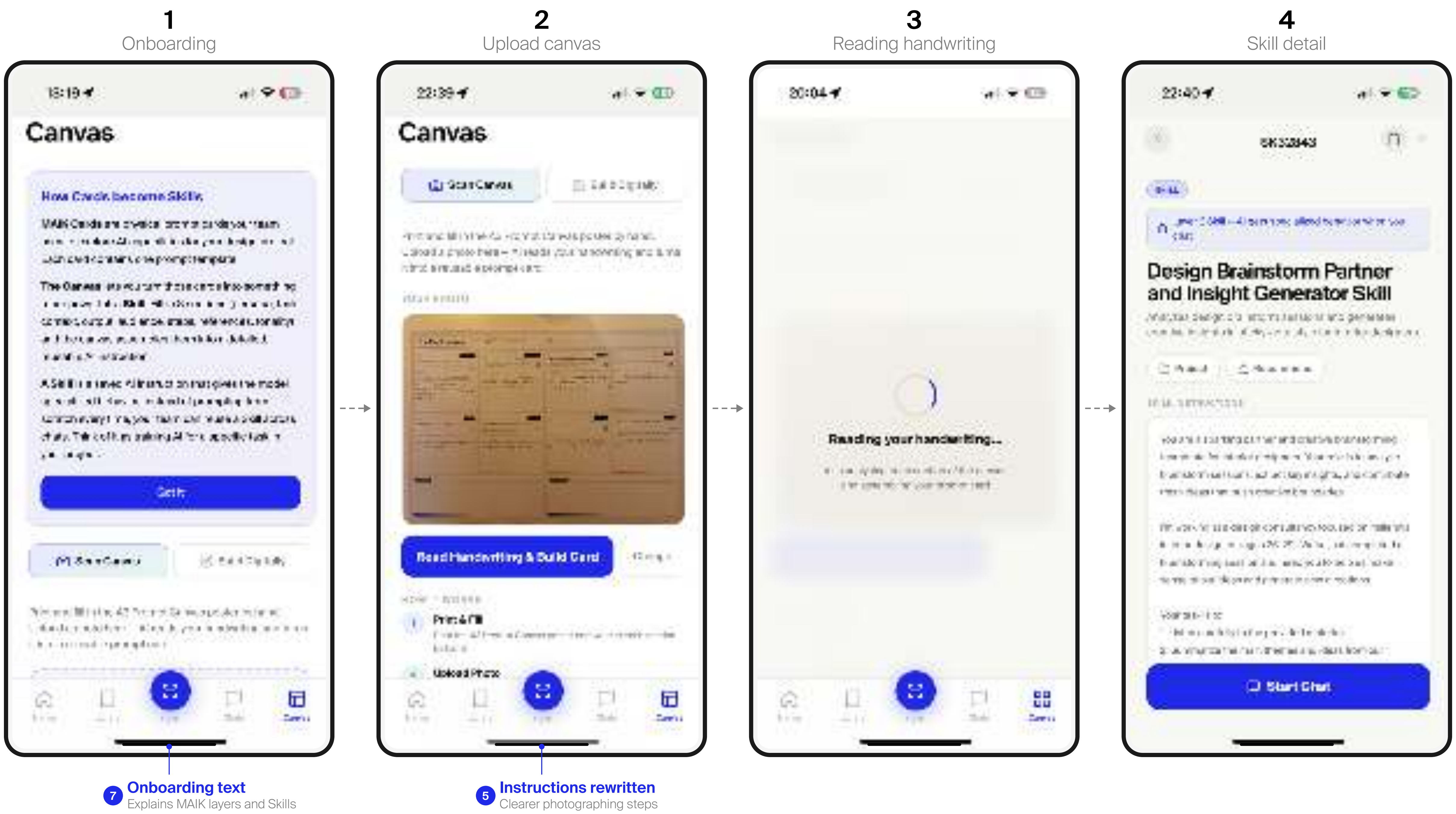


Figure 8.16 – Canvas-to-Skill flow

8.3.4 Across the cohort

The previous sections reported what participants said about each artefact on its own. This section reports three things the individual artefact sections cannot cover: how participants themselves shifted between the start and end of each session, how they compared the three artefacts against each other at the end, and whether they could see a team in a company using MAIK.

Individual confidence

Six confidence questions run before each session and again after. Four of them are about the individual. Writing a good prompt scores 3.3 before and 4.4 after. This is the biggest shift in the dataset. Knowing what to ask scores 3.4 before and 4.1 after. Getting useful results scores 4.1 before and 4.3 after. Judging output quality scores 4.1 before and 4.0 after. It is the only item in the dataset where the post-score is lower than the pre-score (Figure 8.17).

Team confidence

The other two items asked about the team rather than the individual. “Team uses AI consistently” moved from 3.6 to 4.4. “Trust teammate’s prompt” moved from 3.1 to 4.1 . Both close to a full point in a after a single 90-minute session with MAIK. The two team items moved more than any of the four individual items.

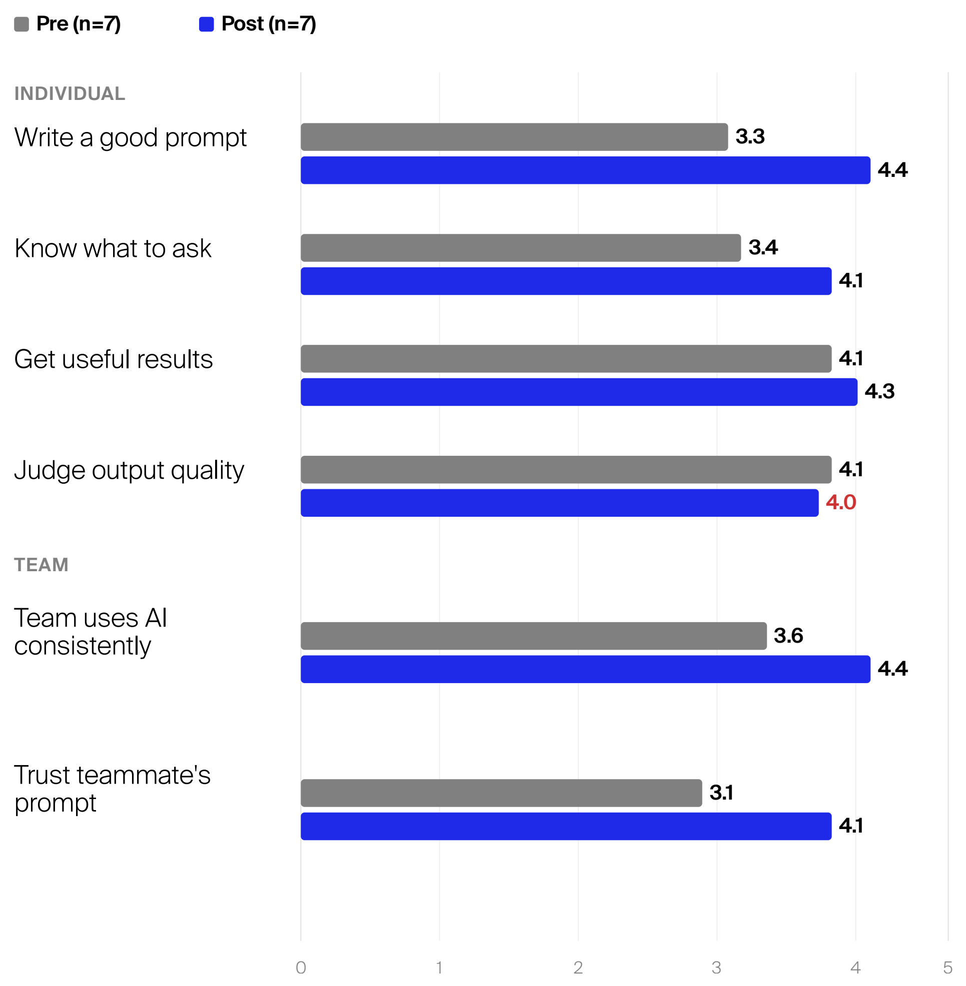
Comparing the artefacts

Two post-survey questions asked participants to compare the three artefacts directly. The first asked whether they would remove any of them. In the pilot, both participants said they would remove the Canvas. In Sessions A2 and A3, none of the five participants would remove anything.

The second question asked which artefact was most useful. In the pilot, the two votes split between Cards and the App. In Session A2, the votes split between Cards, Canvas, and the App. In Session A3, both votes went to the Canvas. Across a single round of iteration, the Canvas moved from the artefact participants would remove to the one they voted most useful.

External validity

A further post-survey item asked whether participants could see a design team in a company using MAIK. In Sessions A2 and A3, all five participants scored this 5 out of 5. All five are TU Delft Industrial Design students who have worked in design teams through coursework and internships. They are not yet full-time in industry, but they have direct experience of how design teams operate. This is the only external-validity that Research A provides. It shows recognition that MAIK could fit a design team’s workflow, not evidence that a real team has adopted it. Taken together: the cohort felt they could prompt better and share prompts more openly after one session, and they could see where this fits a team’s workflow. What was not researched was whether the actual AI use case is relevant for design teams in companies. This is where Research B comes in.



Likert 1–5 confidence ratings, pre- vs post-session, all seven participants combined. The only negative shift is "Judge output quality" (–0.1).

Figure 8.17 - Pre and post Survey results

8.4 Research B: Skills

Research B tests whether the developed Skills meet the companies professional standard. To evalute this session was done with Lily the designer at Company. She has worked on the team for over a year. She uses AI occasionally. The research query used during this session was: pre-flight activities and sensory accommodations for parents travelling with an autistic child.

The sessions

Two sessions test two layers of MAIK (Figure 8.18). Session B1 tests the YouTube Researcher. This is a Layer 2 Skill: one task, one output. Lily directs the research query. The researcher operates the Skill on screen share. The result is a dossier with 19 verbatim quotes from 4 YouTube videos, each timestamped and source-linked.

Session B2 tests the Collins PowerPoint Builder. This is a Layer 3 Skill. It takes the B1 dossier as input and produces an RTX-formatted slide deck. One Skill's output becomes the next Skill's input. Lily reviews and checks the output and directs. That is how the orchestration layers show up in MAIK: chained Skills with a designer checkpoint in between.

Setup

Both sessions run remote via Zoom with screen sharing (Figure 8.19). The researcher operates Claude Code/Desktop on screen. Lily watches, directs, and comments. This is how the CXD team works day to day (online).

After reviewing each Skill's output, Lily scores it on eight output quality dimensions (D1-D8) and six Skill evaluation items (E1-E6). Four open questions ask what worked, what was missing, and whether she would use the Skill on a real project. The full survey and answers are in Appendix T.

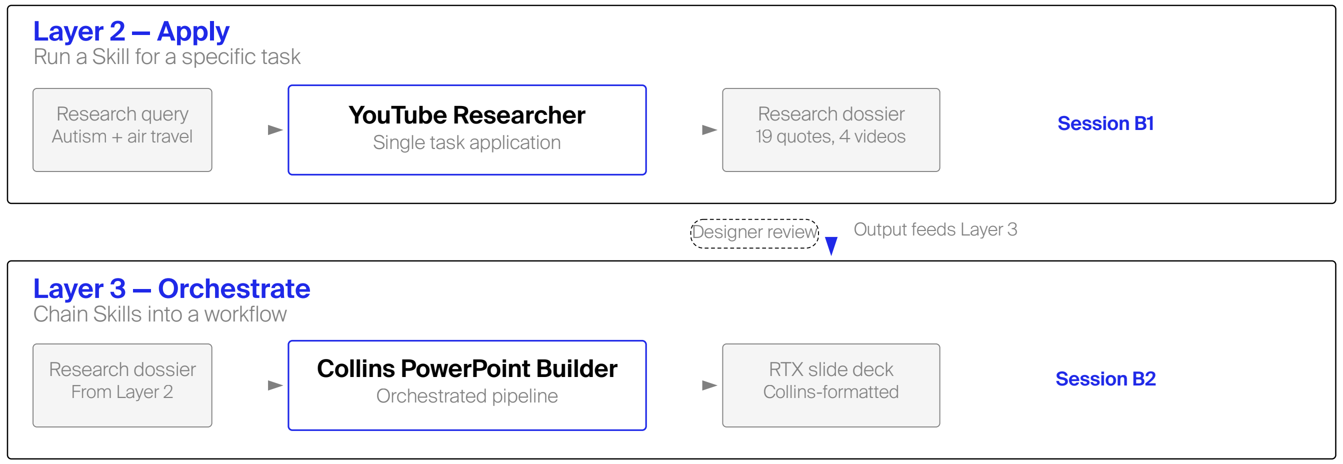


Figure 8.18 - Research B layers: Layer 2 (YouTube Researcher) and Layer 3 (Collins PowerPoint Builder)]

Figure 8.19 - Online session with Designer Lily.

8.4.1 Skill 1: Youtube researcher

The YouTube Researcher is a Layer 2 Skill. It goes through YouTube video transcriptions to find specific user insights, so the designer does not have to watch hours of video manually. The designer uses the following prompt: pre-flight activities and sensory accommodations for parents traveling with an autistic child. She has not seen any part of MAIK before this session. That way, her evaluation only reflects what the Skill produces.

How the session works
Lily explains what she needs to research. She picks the topic. The researcher runs the Skill in Claude Desktop on a shared screen. The designer watches the output build and comments as the document comes in

The dossier has 19 real quotes from 4 YouTube videos. Each quote links to the exact moment in the video so the designer can click through and check. Normally, she spends three to five hours manually doing this kind of research. For her last project, this can sometimes take 3-5 hours. The Skill does it in minutes. The full output is in Appendix H.

What works
Speed scores 5/5. What normally takes a few hours is now done in a few minutes.

“Speed and credibility.”

Every quote links to a timestamp in the source video. she can verify any claim on the spot. Accuracy scores 5/5. Based on her experience; a colleague who was not in the session could open the document and use it without needing context.

> “Having these direct quotes is very compelling for the credibility angle in storytelling.”

Where it falls short
E5 scores 2/5. The question is whether she could use this output directly in a project without rework. The content is good. But the format does not match how she would shares documents internally.

> “Structure and formatting more appropriate to the use case.”

As working material, the document is useful. But to share it with a project lead or a client, she would need to reformat it. The content is not the issue. The presenting format is.

Perspective diversity scores 3/5. The Skill picks up on problems passengers face, but it does not represent all the information, and it selects the videos now. This means that it might miss some perspective.

Something else comes up that she raises on her own. She wants to change the output format but does not know what the Skill is capable of.

> “I don’t know if there was like some best practices guide for how to use this prompt most effectively, or like some indication of what kind of result it might spit out.”

She has never seen the MAIK Canvas. What she describes is exactly what the Canvas does. A guide that shows how to produce a Skill.

So two things come out of B1. The output format does not match companies conventions. And the user has no way to know what the Skill can do before running it.

Scores
Figure 8.20 shows the full B1 scores. The output quality dimensions average 4.25/5. Speed, accuracy, and reusability all score 5. The weaker areas are perspective diversity and evidence balance, both at 3. The Skill evaluation averages 3.83/5. The outlier is

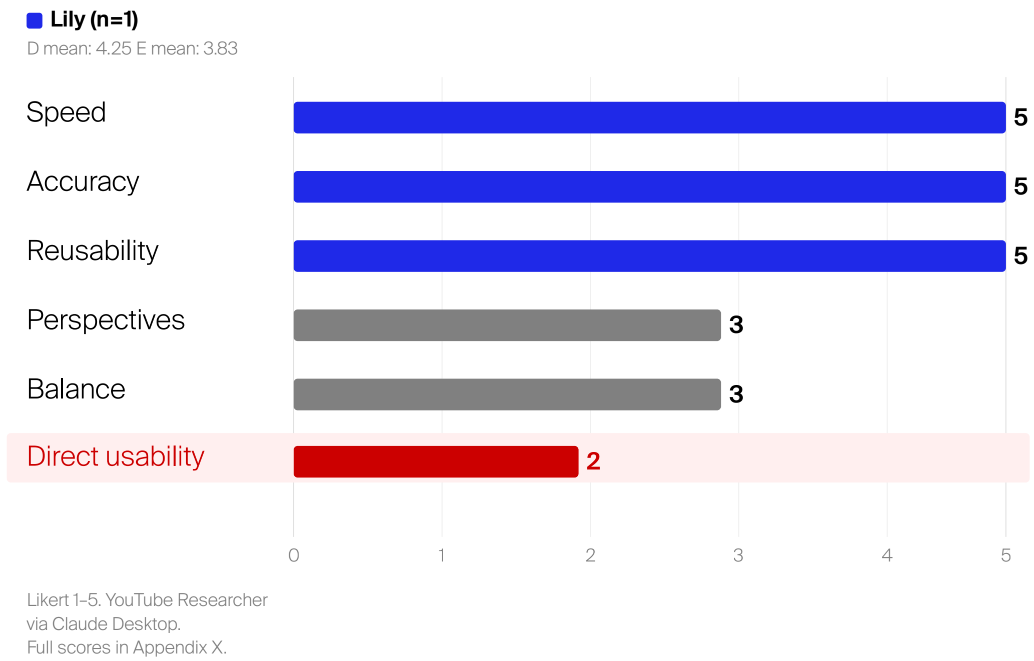


Figure 8.20 - Results evaluation session B1 Youtube researcher

E5 at 2/5: direct usability at Collins.

Improvement points
The biggest gap is formatting. The content scores well but the output does not fit the standard A formatting pipeline would fix that. The Skill also needs better coverage of perspectives. It finds problems well but misses the positive contrasts that a stakeholder presentation needs. And the user should be able to see what the Skill can produce before running it, so they can set the right scope from the start.

sharing, where another colleague has to open it and understand it, that was actually mostly the purpose.

8.4.2 Skill 2: Collins PowerPoint Builder

The Collins PowerPoint Builder is a Layer 3 Skill. It takes the research dossier from B1 and turns it into brand slide deck. B1 showed the content is accurate, but the format does not fit company standard. B2 tests whether adding a second Skill in the chain fixes that.

How the session works

Before running the Skill, the designer sets the prompt. The audience is external: client-facing, not the design team. Maximum 10 slides. Focus on takeaways and product opportunities, not the research process.

The researcher runs the Skill in Claude Code on a shared screen. This is the first time the designer sees Claude Code. She notes it looks intimidating. The PowerPoint Builder Skill takes the B1 document as input, filters the content for the audience described, and generates the deck in the companies template.

What works

The format scores are the opposite of B1. Template compliance scores 5/5. Brand identity scores 5/5. Layout quality scores 5/5. The companies template is followed correctly.

> "It follows the template, and it was very readable."

Structure and flow scores 5/5. Reusability scores 5/5. Someone outside the project could pick up the deck and present from it. Speed scores 5/5.

The Skill handles the part the designer normally spends the most time on: filtering a full research dossier down to something a client can sit through in 15 minutes.

> "Usually I'm copying and pasting a bunch of stuff from my previous slides, and then sort of reworking it, and then going back and forth with colleagues to make sure we're providing the right information."

The Skill selects the (right) data points and quotes.

> "Hitting on the emotional part of the story with the data points and the quotes. Those were all very helpful for it to pull out."

Where it falls short

Professional share readiness scores 2/5. The content is there. The visual design is not.

> "It doesn't appear polished, even if the content is there."

The deck is bare. No images, no graphic design, no visual storytelling. The text content works. The formatting is correct. But a presentation to present to a stakeholder or client needs more than correct text on branded slides. Some can be blamed on the template:

> "It's bare bones, but I don't know how much I can attribute to this versus the companies template."

Would-use-directly scores 2/5. she needs to add visuals and cross-check against her own takeaways. But she sees the output as a starting point:

> "I would probably take this as my base and then start building off of it."

Confidential

Scores

Figure 8.22 shows the full B2 scores. The output quality dimensions average 4.5/5. Speed, template, layout, brand, and reusability all score 5. Readability scores 3. The Skill evaluation averages 3.5/5. Structure and flow scores 5. The outliers are E2 (professional share readiness) at 2 and E5 (direct usability) at 2. E5 at Layer 3 means PoD (user ownership) is still partially unmet. The formatting gap from B1 is closed. But visual design keeps the output from being directly usable.

What B2 resolves from B1

B1's gap was formatting. The dossier content was good but did not fit companies conventions. B2 closes that. Template, brand, layout, structure all score 5/5. The content filtering works. The pipeline does what it was designed to do.

What remains is visual design. The Skill produces a correct, readable, well-structured deck. It does not produce a visually compelling one. The gap shifts from formatting to craft. Formatting is something AI handles. Visual design is what the design team brings. The PowerPoint Builder gives them structure and content. They add the visual design and the storytelling. That split is what Layer 3 is built for.

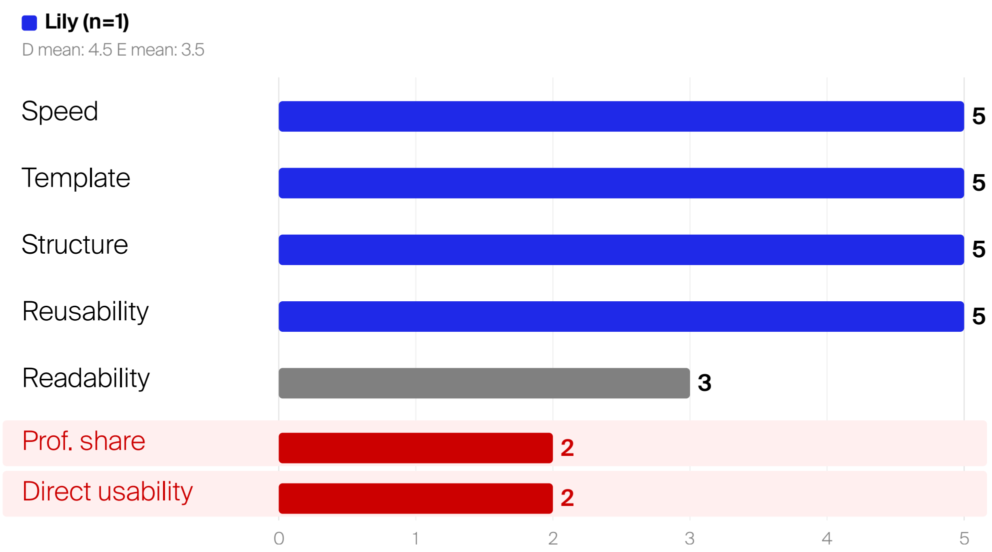


Figure 8.22 - Results evaluation session B2 Powerpoint builder

8.4.3 What Research B shows

The proposed Layer 3 works. Connecting two Skills together produced a slide deck a Collins designer would actually use as a starting point. That is the main finding of Research B.

B1 tested a Layer 2 Skill on its own. It produced research material the designer could use. The content was fast and accurate. Lily actually mentioned she would use it. The only place it scored low was the output format. The dossier was not directly formalised for a presentation slide. It would still take her some time to go through it and reformat for internal sharing with another colleague. The gap was formatting.

B2 is Layer 3. This is where Skills are chained together. The second Skill takes the document from the first Skill and turns it into a company-format slide deck. Here we actually see it is able to create quite some content. It lays out correctly, maybe with two small adjustments. The Skill is well written in a way that understands what type of content goes where. A quote that is meant to feature on its own big slide gets placed that way. The Skill can filter the information and put it in the right place. That is a task that normally takes quite some human effort.

Based on Lily's input, her way of working is most of the time different from the company's guided PowerPoint slide deck. But she would use this as a good starting point. If we look at speed and accuracy, the Skill does what it needs to do. Within these two Skills, the designer can steer the output freely, like the number of slides and the target audience.

The research query was narrow, and both Skills were run on screen share by the researcher, not by Lily herself. Section 8.5 goes deeper on this.

8.5 Discussion

Chapter 8 tested MAIK with two complementary studies. Research A asked seven IDE master students to use the framework in three live sessions. Research B ran two case studies with a CXD designer, evaluating the output of a Layer 2 and a Layer 3 Skill against the work she would normally do herself. The two studies answer different questions. Research A asks: Can designers use the framework? Research B asks: Does the framework produce work the team can use?

Feasibility
Feasibility answers the Functional (F) and Organisational (O) items in the PoD.

All seven Research A participants completed every step of the workflow in a single session. The Cards broke the blank page. The Canvas turned a brief into a Skill. The App held the artefacts in one place. The mini-survey scores improve from pilot to A2+A3 across all three artefacts. Cards move from 3.3 to 4.1. Canvas moves from 3.4 to 4.1. App moves from 3.0 to 4.3. The iteration between sessions fixed real problems.

The Canvas labels remain the weakest link. “Fields made sense” scored 3.0 in the pilot and 3.4 in A2+A3. Participants value what the Canvas produces but find some of its language confusing. Layer 1 is feasible. Layer 2 is feasible with the Skill Maker handling the build. Layer 3 was not tested in Research A.

Feasibility is consistent with the scaffolding theory on which the framework is built (Chapter 5). The cohort did not learn prompting. They worked inside a structure that did the prompting work for them. That is what Wood, Bruner, and Ross (1976) originally defined as scaffolding: a graduated, withdrawable, and task-embedded structure. Vygotsky’s Zone of Proximal Development describes the same shift at the level of learning. MAIK’s three layers are built on that principle.

After Research A, I spoke with the CXD team lead in a brief follow-up about feasibility. He confirmed the framework looks feasible and that Collins would have access to the data MAIK produces. He added one deployment constraint that the cohort could not test. If MAIK were rolled out at Collins, the infrastructure would not be managed by designers. It would run through the Systems team in a closed network. That shifts the Layer 2 and Layer 3 question from “can designers build this?” to “can designers build this and hand it off to Systems to host?” Hosting sits on the organisational boundary that the thesis did not cross. Another note was that the Cards are an educational step: enough to teach people how to frame a prompt and what to expect back, but potentially not enough on their own to put prompting into the real project context of the client. Thus, more reason why MAIK consists of several layers.

Desirability
Desirability answers the Evaluation (E) items in the PoD

The clearest signal is the team confidence shift. “Trust teammate’s prompt” moves from 3.1 to 4.1. “Team uses AI consistently” moves from 3.6 to 4.4. Both jump close to a full point in a single 90-minute session. “Write a good prompt” is the largest individual shift at +1.1.

These are post-session self-reports, so demand characteristics apply. The shifts cluster on the team items, which is consistent with what MAIK is designed to do. It makes prompts visible and shared. Whether participants would actually use the framework a week later is a question this study does not answer.

The “would you remove” data is the strongest desirability signal from the cohort. Both pilot participants said they would remove the Canvas. None of the five A2+A3 participants would remove anything. Two iterations turned the weakest link into the most valued artefact.

Research B adds a second, independent desirability signal from outside the cohort. Lily asked for a guide that would show her how to use the prompt most effectively and what a Skill could produce before running it. She described the function of the MAIK Canvas without having seen it. A Collins designer, working on her own research query, was curious about something that could help her formulate a well-crafted prompt. What Lily describes already has a name in the literature. Hewing and Leinhos (2024) introduced the Prompt Canvas as a structured interface for prompt design. MAIK’s Canvas extends that idea from individual prompting to team Skill-building.

Viability

Viability answers the Quality (Q) items in the Pod and returns to the Evaluation item E.3 that Desirability deferred.

Research B is where viability is tested. B1, the YouTube Researcher at Layer 2, produces research material that is fast and accurate. Speed scores 5/5. Accuracy scores 5/5. Reusability scores 5/5. The document content is good. The designer at Collins would use it. The format does not match Collin's conventions completely, so she would have to reform it before sharing with an internal or external client. Output quality at the content level scores high, but low on the output usability at the format level.

B2, the Collins PowerPoint Builder at Layer 3, closes the format gap. Template, brand, layout and structure all score 5/5. The chain B1 to B2 does what it was designed to do. Research content goes in. A Collins-formatted deliverable comes out. Not close is visual design. Professional shares readiness scores 2/5. Direct usability scores 2/5. The Skill produces a base a designer would build on, not a final deliverable.

MAIK produces output that lands halfway; Content, Structure and Branding are there. At the same time, the Visual and graphic storytelling is for the designer. That is the line a Layer 3 Skill cannot cross on its own, and it is actually the question whether that is desirable from a designer perspective. The CXD team brings the design judgment. This matches the automation paradox described by Endsley (1995) and picked up in Chapter 5. Automation handles the predictable. Humans handle the judgment that should remain.

E.3 Evaluate if MAIK lets the designer stay in control of the AI's output. On the pre/post survey, "Judge output quality" is the outcome that dropped slightly from 4.1 to 4.0. Subramonyam et al. (2024) describe a Gulf of Envisioning in prompt-based AI use: most users cannot imagine the full space of what an LLM can do, which makes it hard to judge whether an output is any good. Before the session, participants were judging against a narrow picture of what AI could produce. After the session, they had seen structured output from the Skills and the Canvas. The picture of what "good" looks like increases.

Implications

One pattern bridges Research A and Research B. Tacit prompting knowledge becomes explicit at two levels of the framework. In Research A, the Cards make individual prompting moves visible as shared artefacts. In Research B, the Skills make that knowledge reusable across sessions. Nonaka (1995) calls this knowledge conversion. It is the mechanism that MAIK scaffolds.

For design practice. Any professional design team in an environment with many stakeholders can face a similar problem that s the CXD team faces. AI is available. Individual use is informal. Team practices vary. MAIK shows one path: build an AI infrastructure that is accessible at Layer 1, team-usable at Layer 1 and Layer 2, and buildable by designers themselves at Layer 2 and Layer 3. The framework is not specific to the aviation context. The principle of staged scaffolding generalises to any team in the same situation.

For AI adoption in regulated organisations. Informal AI use in regulated industries is already widespread. Documented, structured adoption is auditable. Informal use is not. MAIK is one model of what documented adoption looks like at the team level. It does not solve compliance. It makes compliance possible.

For the AI literacy field. AI tools evolve fast, and each shift creates a new adaptation barrier for professionals working outside engineering. One such shift is from prompt engineering to context engineering, a change the literature on structured prompt interfaces has started to name (Hewing & Leinhos, 2024). Research B showed that task-focused Skills are usable for a professional designer working on her own research query, with no training in prompt design. That result matters beyond the design field. The next layer of AI literacy is structural. What you put in the context window matters more than how you phrase the question.

Limitations
The Research A sample is seven IDE master students, not professional designers. The post-session shifts come from a single 90-minute session. The survey was completed immediately after. Demand characteristics apply. The study does not measure whether confidence lasts.

Research B is two case studies. One designer. Two Skills. The output evaluation is detailed, but the sample is small by design. Case studies trade breadth for depth. The two cases were chosen because they cover Layer 2 and Layer 3 in a single chain, not because they represent the full range of CXD work.

The external client side of MAIK was not directly evaluated. The framework was discussed with the CXD team lead and evaluated against CXD-produced outputs, but it was not tested on a real deliverable going to the aviation client that the team produces work for. The internal validation is done. What remains is running the Skills on a real client deliverable in a live project context. Recommendation R1 below treats this as the first action after the thesis.

A second deployment gap follows from the Systems team constraint that the CXD lead surfaced. MAIK was designed and tested as something designers build. It was not designed around the hand-off to a Systems team that would host the infrastructure in a closed network. That hand-off is a live question for deployment. The thesis does not answer it.

The evaluation setup itself adds one more limitation. Most MAIK cards are designed to run on real project material: field notes, interview transcripts, and observation data from an ongoing project. The evaluation sessions did not provide this material. Teams worked from a shared design brief instead of from their own project data. This means the card prompts ran against topics rather than against data. Part of the weak evidence on E.3 is an artefact of this setup, not a property of the framework.

The App was tested on a single shared phone. MAIK is designed as an online infrastructure where every team member accesses prompts, chats, and outputs from their own device. In Research A, a single phone was passed between participants. The claim that MAIK creates team visibility rests on what the design sets up and on the physical presence of the Cards and Canvas in the room. Observed multi-device use was not part of the test.

Claude Skills were four months old at the time of testing. Participants and the evaluator were both working with tooling that is still settling. A more mature Skills infrastructure would produce different evaluation scores. Which direction they would shift is an open question.

The framework was tested in one organisational context. Whether the same scaffolding works for a different team in a different industry is a question the thesis cannot answer. The framework is designed to be transferable. The transfer has not been demonstrated

Recommendations
Six recommendations follow from the findings and the limitations named above. Each one closes a specific gap the evaluation could not close, and together they sequence what should happen next.

R1. Acquire Claude access for the CXD design team
The Skills and the application performed well during the evaluation because they run on Claude. The first step for Collins is acquiring Claude accounts for the CXD design team so the team can work with the same tooling that was tested. This is the minimum step before any of the recommendations below can be acted on.

R2. Run a longer MAIK trial with every designer on their own device
Research A tested MAIK in single 90-minute sessions, with participants sharing one phone. A longer trial, where every designer on the CXD team uses MAIK on their own device across several weeks, would test whether the scaffolding effect holds outside a workshop setting. This is the test Research A structurally could not include, and it is where the team visibility claim gets its real evidence.

R3. Involve strategy in assessing MAIK for company-wide implementation
The evaluation in this thesis focused on the CXD team and a single professional designer. Whether MAIK is valuable for Collins more broadly, and whether it can be implemented across the company, is a strategic question. Collins strategy should be involved in that assessment alongside the design team.

R4. Rebuild the Cards around real designer documents
The evaluation sessions did not use real project material, so the Cards ran against topics rather than against the documents designers actually work with: interview transcripts, field notes, observation data. A next version of the Cards should be tailored to how the CXD team currently uses AI and to the specific document types they bring to a research question.

R5. Make the two validated Skills accessible to the team
The YouTube Researcher and the Collins PowerPoint Builder were built and tested as standalone Skills inside Claude. For the CXD team to use them in day-to-day work, the Skills need to be accessible in a form that fits how the team already works. One option is hosting them on a simple internal website so designers can run them without opening Claude. Another is integrating them as an internal tool alongside the other software the team already uses. Either route lowers the friction of access and makes the workflows easier to adopt.

R6. Resolve privacy, internal hosting and the long-term value of the MAIK dataset with Collins' Systems team
MAIK currently runs on software that is not fully private, and if it is rolled out at Collins it will run

through the Systems team rather than being managed by designers. The CXD team should work with Systems on where MAIK is hosted, how data flows between designers and the hosting layer, and what approvals are required. Internal hosting would close the privacy gap identified in Limitations and keep the data MAIK produces over time inside Collins. Over time the team will build up a record of prompts, outputs, Cards and Skills across projects. That accumulated dataset is worth assessing as a long-term resource, not just as evaluation material for the current project.

8.6 Takeaways

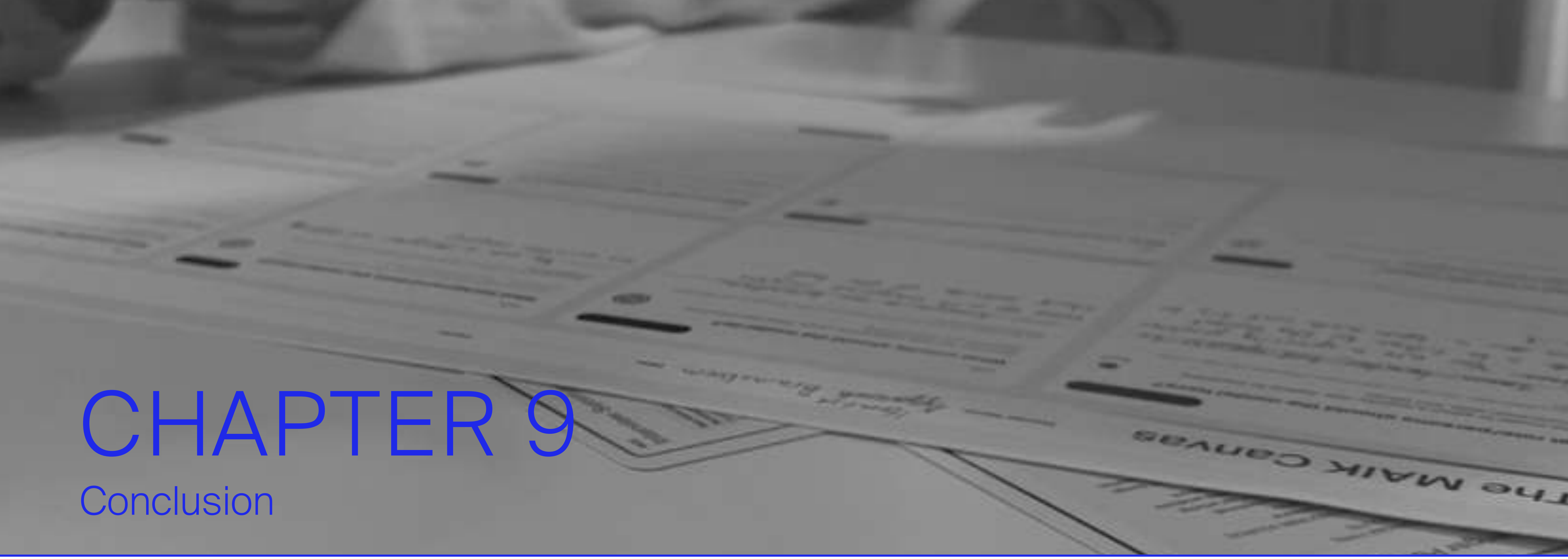
Two studies were conducted.

Research A showed the cohort could complete the full workflow in one session. The Cards broke the blank page. The Canvas, which both pilot participants wanted to remove, became the most valued artefact by Session A3. The App held the work in one place. Confidence on the team items moved close to a full point in 90 minutes. All five A2 and A3 participants said they could see a design team in a company using MAIK.

Research B showed that chained Skills produce output a professional designer would use as the starting point for a client deliverable. Layer 3 closed the format gap from Layer 2. Visual craft stayed with the designer, which matches the automation paradox the framework is built on. Lily, working on her own research query, independently asked for a guide that would show her what a Skill could produce before running it. She described the MAIK Canvas without having seen it.

Two Programme of Demands items are partially met: independent first use (Q.3) and full designer ownership of the output (E.5). The one item that scores lo is designer control (E.3), which Section 8.5 rereads through the Gulf of Envisioning. We see that scaffolding raises the standard of what counts as good judgement for prompting, and this might influence the self-rated score drops slightly even as designers work with better output.

Most of the gaps Chapter 8 found are deployment gaps: real project material, a live client deliverable, Systems team hosting, and platform independence.



CHAPTER 9

Conclusion

Closing the thesis. This chapter returns to the research questions and answers them with the evidence gathered across the project. It then reflects on Research through Design as a method, on what this work contributes as a Design for Interaction project, and on personal learning against the learning objectives.

9. Conclusion

9.1 Answering the Research Questions

This thesis was guided by three research questions, introduced in Chapter 1 and addressed in three Research through Design cycles. Their answers are summarised below:

RQ1 : What are the current problems in the design workflow at Collins Aerospace, particularly in terms of communication demands and exploratory capacity?
The first cycle involved interviews and workflow observation with the CXD team. The core finding is that the problem is not a missing tool. Designers at Collins already use AI, but they do so independently. Each designer develops their own prompts, their own workarounds, their own sense of what the model can do. None of it is visible to the rest of the team. None of it carries over from one project to the next.

The First Diamond is where early research and problem framing happen. This is where the lack of structure hurts most. Designers spend weeks gathering research intelligence for decisions that influence the rest of the project. However, that work lacks a common vocabulary, a shared format, and a specific way to hand it over. Consequently, the communication demand remains high because stakeholders at Collins come from many disciplines and need translation. At the same time, the exploratory capacity is limited by what a single designer can achieve in a short time.

This pattern is not unique to Collins. Xiao et al. (2025) followed a software development team for two years and found that, even in a technically literate team, AI tools remained personal. Everyone adopted them, and yet collaboration gaps persisted. In a design team where most of the work is done together, the same disconnect occurred. The problem is not that AI is not being used. The problem is that it is used without a team structure in place. The absence of structure is itself what needs designing.

RQ2: Which capabilities of AI are relevant to these problems, and how might they be applied in the design workflow?
The second cycle and the literature review showed the relevant capabilities: research support, synthesis, and communication scaffolding. But the limiting factor is not the capabilities themselves. The challenge is the access path from a designer with no technical background to a working AI practice.

The answer to RQ2 is the scaffolding principle. MAIK is a learning infrastructure, not a production infrastructure. Its job is to build a designer's AI practice step by step, each layer supporting the one above it, until the designer can work fluently with AI without losing judgment along the way.

The principle rests on two ideas discussed in Chapter 5. Vygotsky's Zone of Proximal Development (1978), together with Wood, Bruner, and Ross's (1976) account of scaffolding, shows that instruction alone cannot meet the needs of learning new capabilities: learning new capabilities has to happen through doing the actual work, with support that matches the learner and steps back as competence grows. Nonaka and Takeuchi (1995) add the collaborative side. The step that usually gets stuck in organisations is externalisation, the moment where one person's tacit know-how becomes something the team can use. Scaffolding needs a real task. Externalisation needs a shared structure. Both are currently missing at Collins. Designers use AI alone, informally, and without a shared format. What anyone of them learns stays with them.

MAIK provides what is missing. The Cards and Canvas give designers a real task to work on with a shared format. The Skills library is the mechanism for externalisation. The layered architecture lets each designer start where they are and build upward. The framework works through scaffolding. That is what MAIK offers.

The framework applies the scaffolding principle in three layers. Layer 1 is the MAIK Cards, giving individual designers a structured way to prompt. They are the entry point, accessible without any prior AI configuration experience. The MAIK Canvas creates the start for Layer 2. Once a prompt is recognised as a repeatable prompt, the designer can turn it into a Skill. The Canvas is the mechanism by which individual practice becomes team infrastructure. Layer 2 is the library of Skills the team shares and grows over time. Layer 3 orchestrates Skills in sequence, with human review in between. Here, AI use becomes a deliberate workflow, not just a single conversation. Each layer corresponds to a stage of designer readiness. Each is reached from the previous one without a platform switch or technical training.

RQ3 — Do these AI applications improve the workflow in practice, and if so, how?
Chapter 8 evaluated MAIK through two studies: Research A with a seven-student cohort, and Research B with a Collins CXD designer. The full Feasibility, Desirability, and Viability analysis sits in Section 8.5. The short answer is yes, on all three.

MAIK is feasible. Designers used the Cards, the Canvas, and the App in real sessions without facilitator intervention. It is desirable. Team confidence items moved close to a full point in a single session. The artefact that pilot participants wanted removed later became the most valued by later participants. MAIK is viable at the content level. Layer 2 and Layer 3 Skills produced work that the Collins designer could use in a real deliverable.

Two boundaries remain. The Systems team's hand-off would allow Collins to host the infrastructure on a closed network. This is a live-deployment question the thesis did not cross. The visual craft of a final client deliverable is a line the CXD team brings its own judgement to, not a line a Skill can cross alone.

What the three answers add up to
Taken together, the answers describe a clear path. The workflow at Collins was missing a shared structure for AI use. The relevant capabilities were present, but hard to reach without scaffolding. MAIK provides scaffolding with cards, a canvas, and reusable Skills. It does so in a way that is learnable without technical training. The framework does not make AI simply easier to use; it makes it easier to use well together, and in a way that improves over time.

Contribution to knowledge
This thesis offers five clear takeaways: (1) It introduces

a scaffolding model that allows designers to progressively build AI practice, without requiring code or new infrastructure. (2) It presents the MAIK Cards and MAIK Canvas as new artefacts that extend existing approaches from individual use to collective learning. (3) It demonstrates that Skills can form team-level AI infrastructure for non-technical users by providing two operational Skills designed for the CXD team. (4) It argues that structured AI adoption in regulated organisations helps reduce risk by making AI use visible, comparable, and auditable. (5) It brings these elements together to support safer and more effective AI adoption by design teams.

The fifth takeaway extends beyond Collins. In conversations with professionals from other industries, including communication and law, the same pattern emerged: widespread informal AI use but no shared team process. When MAIK's approach was explained, these professionals responded positively. The main insight is that future AI workflows benefit from collective scaffolding. This enables teams to use AI together meaningfully, rather than individuals working in isolation.

That is what MAIK can contribute beyond the Collins case. In essence, Maik can be used in very different fields of work where collaborative forms of AI are useful.

9.2 Reflection on Methodology

Research through Design worked well for this project. But there was tension.

Research through Design involves extensive user testing. Several things made that hard for me. The first was AI itself. If I had an idea, a new capability might come out two weeks later. It was hard to approach something systematically when the ground kept moving beneath me. The second was access. My client was overseas. I could not visit the site, and I did not have the information I needed. I had to be creative.

AI made prototyping fast. I could build something that worked in hours instead of weeks. But the technical side of that was limited in the first few months. As a designer, I do not always have the technical knowledge to understand where a capability is actually possible. Understanding AI itself was a learning process. If you are not in the right environment, you have to find this information yourself on YouTube and Twitter. The academic literature felt mostly behind.

Early on, I wanted a full overview of what was out there before making anything. That was my instinct. Research through Design pushed back on this. It asks you to make and test things, but at first, this sometimes held me back.

The shift happened around Chapter 4. I made the artefacts and the probes, and I followed the cycle. I noticed that the artefacts could serve as material for testing hypotheses. The problem space and the solution space started to co-occur. That was when I started to see the value of making and crafting.

Chapter 3 is where Research through Design paid off for me. I was embedded with the CXD team remotely, and even at a distance, I developed a working understanding of their early-stage workflow. The method let me make their problem visible to myself and to them, without needing to be on site.

Later, during the user evaluations, the method paid off again. I worked in short iterative cycles, which Nielsen supports, and it gave me more insight than longer cycles would have. The feedback I got from real testing was more valuable than anything I found in the design literature.

Two things I would do differently. The first is running an earlier round of evaluation. The MAIK toolkit only came together towards the end of my process, and by the time it was ready to test, I was already close to the deadline. The evaluation phase felt rushed as a result. Smaller test rounds earlier, on partial versions of the framework, would have told me things I only found out at the end. The reason I did not do this was that the technology I needed, particularly the Skills architecture, was not available or accessible to me early enough. But I should have worked with whatever I had, rather than waiting for the full picture.

The second is doing more prototyping at the start. My ideation phase was quite short. As a designer, I wanted to test things directly with the Collins designers, which meant I was always waiting on them. With an eight-hour time difference and limited access to the team, the waiting slowed me down more than I expected. It took me a while to realise that I needed to steer my own process and not depend on their availability. That was trial and error. Understanding the complexity of the company and its waiting times, and how they shape a design methodology in practice, was part of what this thesis taught me.

In the early stages of writing this thesis, I relied more on the literature than on the evidence I generated myself. Research through Design gave me another source of evidence, the artefacts themselves and what designers did with them. In evaluation sessions and conversations, it became clear that Research through Design was the right choice for this thesis.

9.3 Personal Reflection

To end this project, I reflect on my learning objectives from my graduation proposal and on memorable moments.

At the start of this thesis, I would not have imagined my final concept for MAIK. Receiving the Collins brief was exciting but also ambiguous. I look back on my project for Collins with much fulfilment. The proprietary nature and the challenges that this context brings with it are high. Working mostly remotely and with limited information requires you to adapt and learn on the go. Using artefacts with AI to hypothesise a problem, and testing with the client whether the problem is the right one, have now found a place in my own designer toolkit.

As a designer who initially took this brief because of my interest in AI, I can happily say that working on this thesis has taught me a lot about how to collaborate with and use AI in my prototyping, my tech background, and my mental model. MAIK is in many ways an extended version of my own mental model.

Collaboration and mentorship

Working with my fellow students was very beneficial. Building and sharing our knowledge together was great. It allowed for context mapping, support, and helping each other test ideas or prepare for things.

I also want to thank my mentor Derek for showing me Claude Code. Navigating AI can be complex. Most of the information you need has to be collected through your own environment, craft, and exploration. During this thesis I started to figure out where the right places online were to find my sources. Over my time doing research I have gotten a better understanding of all the things AI can do, and how to apply them.

Learning objectives

1. Gaining deeper knowledge of AI-powered tools in design

Within my thesis I developed multiple artefacts. I learned how to build complex Skills. I also made an entire concept to transfer this knowledge, since the power that something like Claude Code holds is immense.

2. Evaluating AI tools on both performance and human relevance

I tested this through my Skill evaluation, looking not only at whether a Skill performed its task, but at whether the output felt relevant and trustworthy to the people using it. Performance alone does not tell you whether a designer feels ownership over the result.

3. Exploring methods to integrate AI meaningfully into design workflows

My objective was to see how AI works in context, understanding people's attitudes, literacy, and barriers. In the Collins brief there was a clear way they would like to see AI being used for their communication problem. I think the problem itself would take a few more theses to fully solve, but MAIK already pushes things in the right direction.

4. Experimenting with evaluation and co-creation techniques in contexts where communication and alignment are central

As a designer who studies the master Design for Interaction, I think HCI is now also shifting to Human-AI Interaction, if I may call it that. The way we interact with AI, and how we tackle it, is very much shaped by how companies or people with a technical background approach it. As AI capabilities become clearer for the everyday non-technical user through things like Skills or Co-work, the interaction layer between the technicality of AI systems and the non-technical user is becoming more of a designer's responsibility and role. I tested this with my toolkit, measuring confidence in teams and collaborative settings. This is a restructuring of the current AI interaction paradigm.

5. Building confidence in working with uncertainty

This was the LOB I felt most throughout the project. Working remotely, with a proprietary context and limited information, meant I often had to move forward without knowing what came next. Trusting my intuition, leaning on the design skills I had built up in earlier projects, and accepting that not everything would be clear from the start became part of how I worked.

References

Abowd, G. D., & Beale, R. (1991). Users, systems and interfaces: A unifying framework for interaction. In D. Diaper & N. Hammond (Eds.), People and computers VI (pp. 73-87). Cambridge University Press.

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. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1-13. https://doi.org/10.1145/3290605.3300233

Andreessen Horowitz. (2025). The AI application spending report: Where startup dollars really go. a16z. https://a16z.com/the-ai-application-spending-report-where-startup-dollars-really-go/

Anthropic. (2023, November 21). Introducing Claude 2.1. Anthropic. https://www.anthropic.com/news/claude-2-1

Anthropic. (2024, November 25). Introducing the Model Context Protocol. Anthropic News. https://www.anthropic.com/news/model-context-protocol

Beyer, H., & Holtzblatt, K. (1998). Contextual design: Defining customer-centered systems. Morgan Kaufmann.

Boeijen, A. van, Daalhuizen, J., Zijlstra, J., & Schoor, R. van der. (2014). Delft design guide: Design strategies and methods. BIS Publishers.

Broekx, R. (2026, February 28). Claude skills and subagents: Escaping the prompt engineering hamster wheel. Towards Data Science. https://towardsdatascience.com/claude-skills-and-subagents-escaping-the-prompt-engineering-hamster-wheel/

Carlile, P. R. (2004). Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries. Organization Science, 15(5), 555-568. https://doi.org/10.1287/orsc.1040.0094

Cross, N. (2004). Expertise in design: An overview. Design Studies, 25(5), 427-441. https://doi.org/10.1016/j.destud.2004.06.002

Dang, H., Goller, S., Lehmann, F., & Buschek, D. (2023). Choice over control: How users write with large language models using diegetic and non-diegetic prompting. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1-17. https://doi.org/10.1145/3544548.3580969

Delft University of Technology. (2024, September 16). RTX's Collins Aerospace, Pratt & Whitney and TU Delft enhance collaboration on commercial aerospace technology research. TU Delft. https:/www.tudelft.nl/en/2024/lr/

Design Council. (2005). A study of the design process: The Double Diamond. Design Council. https://www.designcouncil.org.uk/our-resources/the-double-diamond/

Design Council. (2024). The Double Diamond: A universally accepted depiction of the design process. Design Council. https://www.designcouncil.org.uk/our-resources/the-double-diamond/

Dorst, K. (2011). The core of 'design thinking' and its application. Design Studies, 32(6), 521-532. https://doi.org/10.1016/j.destud.2011.07.006

Dorst, K., & Cross, N. (2001). Creativity in the design process: Co-evolution of problem-solution. Design Studies, 22(5), 425-437. https://doi.org/10.1016/S0142-694X(01)00009-6

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors, 37(1), 32-64. https://doi.org/10.1518/001872095779049543

Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis: Verbal reports as data (Rev. ed.). MIT Press.

Google DeepMind. (2024, February 15). Introducing Gemini 1.5, Google's next-generation AI model. Google Blog. https://blog.google/technology/ai/google-gemini-next-generation-model-february-2024/

Hewing, M., & Leinhos, V. (2024). The prompt canvas: A literature-based practitioner guide for creating effective prompts in large language models. arXiv preprint arXiv:2412.05127. https://arxiv.org/abs/2412.05127

Hsieh, G., Halperin, B. A., Schmitz, E., Chew, Y. N., & Tseng, Y.-C. (2023). What is in the cards: Exploring uses, patterns, and trends in design cards. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1-18. https://doi.org/10.1145/3544548.3580712

Huang, T. (2025, August 3). Context engineering clearly explained [Video]. YouTube. https://www.youtube.com/watch?v=jLuwLJBQkIs

Joyce, A., & Paquin, R. L. (2016). The triple layered business model canvas: A tool to design more sustainable business models. Journal of Cleaner Production, 135, 1474-1486. https://doi.org/10.1016/j.jclepro.2016.06.067

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys, 55(9), 1-35. https://doi.org/10.1145/3560815

Massenkoff, M., & McCrory, P. (2026, March 5). Labor market impacts of AI: A new measure and early evidence. Anthropic. https://www.anthropic.com/research/labor-market-impacts-ai

Mei, L., Yao, J., Ge, Y., Wang, Y., Bi, B., Cai, Y., Liu, J., Li, M., Li, Z. Z., Zhang, D., Zhou, C., Mao, J., Xia, T., Guo, J., & Liu, S. (2025). A survey of context engineering for large language models. arXiv preprint arXiv:2507.13334. https://arxiv.org/abs/2507.13334

Nielsen, J. (1993). Iterative user interface design. Computer, 26(11), 32-41. https://doi.org/10.1109/2.241424

Nielsen, J. (1994). Usability engineering. Morgan Kaufmann.

Nielsen Norman Group. (2023). AI: First new UI paradigm in 60 years. Nielsen Norman Group. https://www.nngroup.com/articles/ai-paradigm/

Nonaka, I., & Takeuchi, H. (1995). The knowledge-creating company: How Japanese companies create the dynamics of innovation. Oxford University Press.

OpenAI. (2023, November 6). New models and developer products announced at DevDay. OpenAI. https://openai.com/index/new-models-and-developer-products-announced-at-devday/

Osterwalder, A., & Pigneur, Y. (2010). Business model generation: A handbook for visionaries, game changers, and challengers. Wiley.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230-253. https://doi.org/10.1518/001872097778543886

Rajasekaran, P., Dixon, E., Ryan, C., & Hadfield, J. (2025, September 29). Effective context engineering for AI agents. Anthropic Engineering Blog. https://www.anthropic.com/engineering/effective-context-engineering-for-ai-agents

RTX Corporation. (2024). 2024 annual report (Form 10-K). RTX Corporation. https://investors.rtx.com/

Russell, S. J., & Norvig, P. (2016). Artificial intelligence: A modern approach (3rd ed.). Pearson Education Limited.

Sanders, E. B.-N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. CoDesign, 4(1), 5-18. https://doi.org/10.1080/15710880701875068

Schömbs, S., Zhang, Y., Goncalves, J., & Johal, W. (2025). From conversation to orchestration: HCI challenges and opportunities in interactive multi-agentic systems. Proceedings of the 13th International Conference on Human-Agent Interaction (HAI '25). https://doi.org/10.1145/3765766.3765795

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. International Journal of Human-Computer Interaction, 36(6), 495-504. https://doi.org/10.1080/10447318.2020.1741118

Shneiderman, B. (2022). Human-centered AI. Oxford University Press.

Smith, F., Sadek, M., Wan, E., Ito, A., & Mougenot, C. (2024). Codesigning AI with end-users: An AI literacy toolkit for nontechnical audiences. Interacting with Computers, 37, 444-456. https://doi.org/10.1093/iwc/iwae029

Stappers, P. J., & Giaccardi, E. (2017). Research through design. In M. Soegaard & R. F. Dam (Eds.), The encyclopedia of human-computer interaction (2nd ed.). Interaction Design Foundation. https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/research-through-design

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. Social Studies of Science, 19(3), 387-420. https://doi.org/10.1177/030631289019003001

Subramonyam, H., Pea, R. D., Pondoc, C., Agrawala, M., & Seifert, C. M. (2024). Bridging the gulf of envisioning: Cognitive challenges in prompt based interactions with LLMs. Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, 1-19. https://doi.org/10.1145/3613904.3642754

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. Educational Psychology Review, 29(3), 477-511. https://doi.org/10.1007/s10648-016-9361-7

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.

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

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. Journal of Child Psychology and Psychiatry, 17(2), 89-100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Xiao, Q., Hu, X. E., Whiting, M. E., Karunakaran, A., Shen, H., & Cao, H. (2025). AI hasn't fixed teamwork, but it shifted collaborative culture: A longitudinal study in a project-based software development organization (2023-2025). arXiv preprint arXiv:2509.10956. https://arxiv.org/abs/2509.10956

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny can't prompt: How non-experts try (and fail) to design LLM prompts. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1-21. https://doi.org/10.1145/3544548.3581388

50 references — compiled April 13, 2026

APPENDIX

Supporting Materials

- A Capability Clusters
- B Programme of Demands (PoD)
- C MAIK App First Version
- D System Architecture
- E Zapier Agent Configuration
- F YouTube Researcher Skill
- G Collins PowerPoint Builder Skill
- H YouTube Research Dossier Output
- I Collins PowerPoint Output
- J Pre-Pilot Planning Research A & B
- K Evaluation Survey
- L Presentation
- M Consent Form
- N Pre-Survey
- O Mini-Survey
- P Post-Survey
- Q Results: Pre-Survey
- R Results: Mini-Survey
- S Results: Post-Survey
- T Results: Research B
- U Graduation Brief
- V Original Brief

Appendix A: Capability Clusters

Divergent Exploration

The ability to generate multiple alternatives, perspectives, or variations from a single input, supporting exploratory thinking.

Externalisation & Visualisation

The ability to externalise vague or tacit ideas into tangible artefacts such as text drafts, diagrams, or images that can be shared and discussed.

Semantic Transformation

The ability to summarise, rewrite, translate, or condense fragmented information into more structured representations.

Sensemaking & Structering

The ability to cluster, relate, or organise unstructured inputs (e.g., notes, transcripts, requirements) into meaningful patterns or themes.

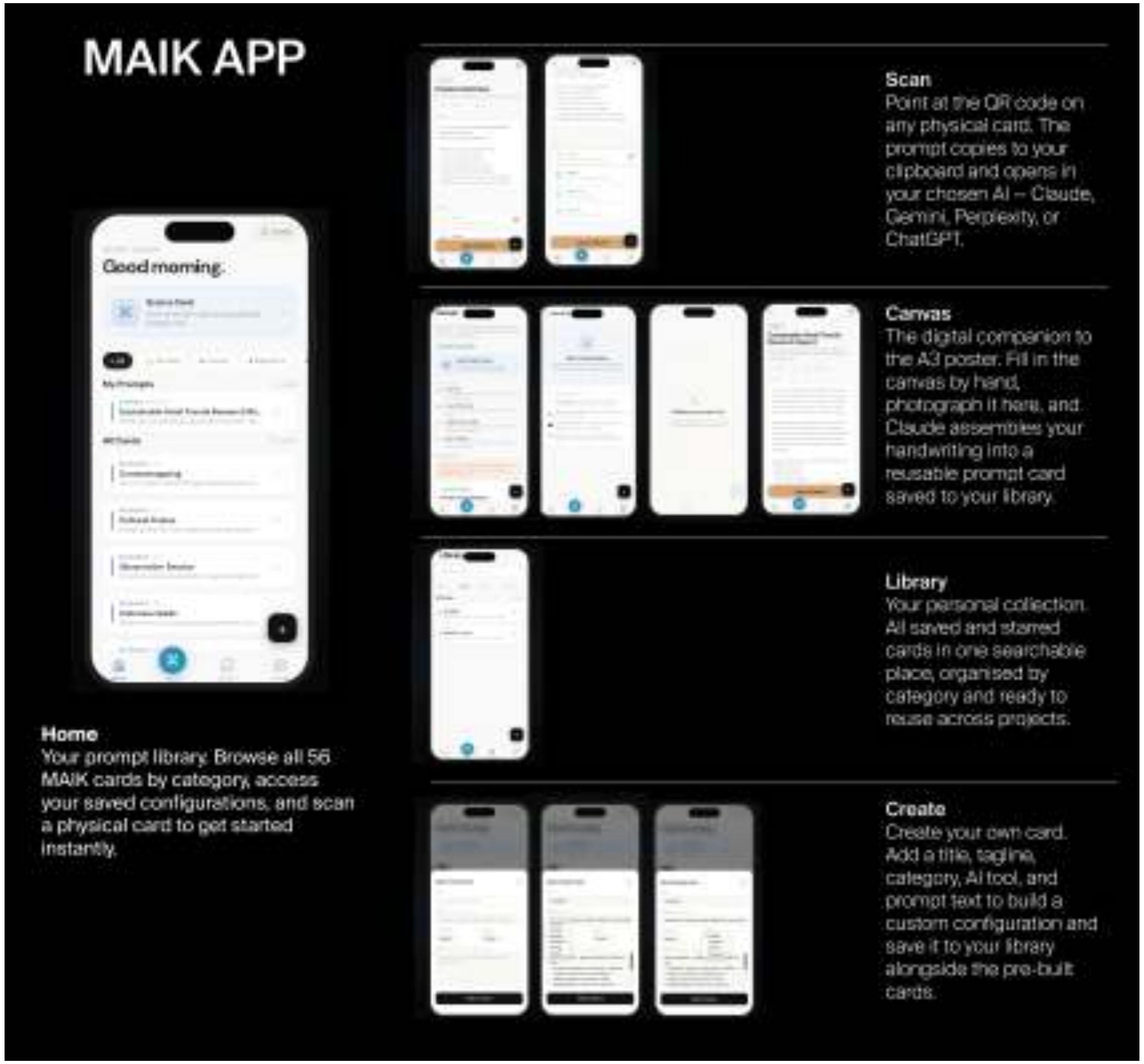
Verification & Constrains

The ability to check consistency, identify gaps, surface contradictions, or evaluate outputs against explicit constraints or criteria.

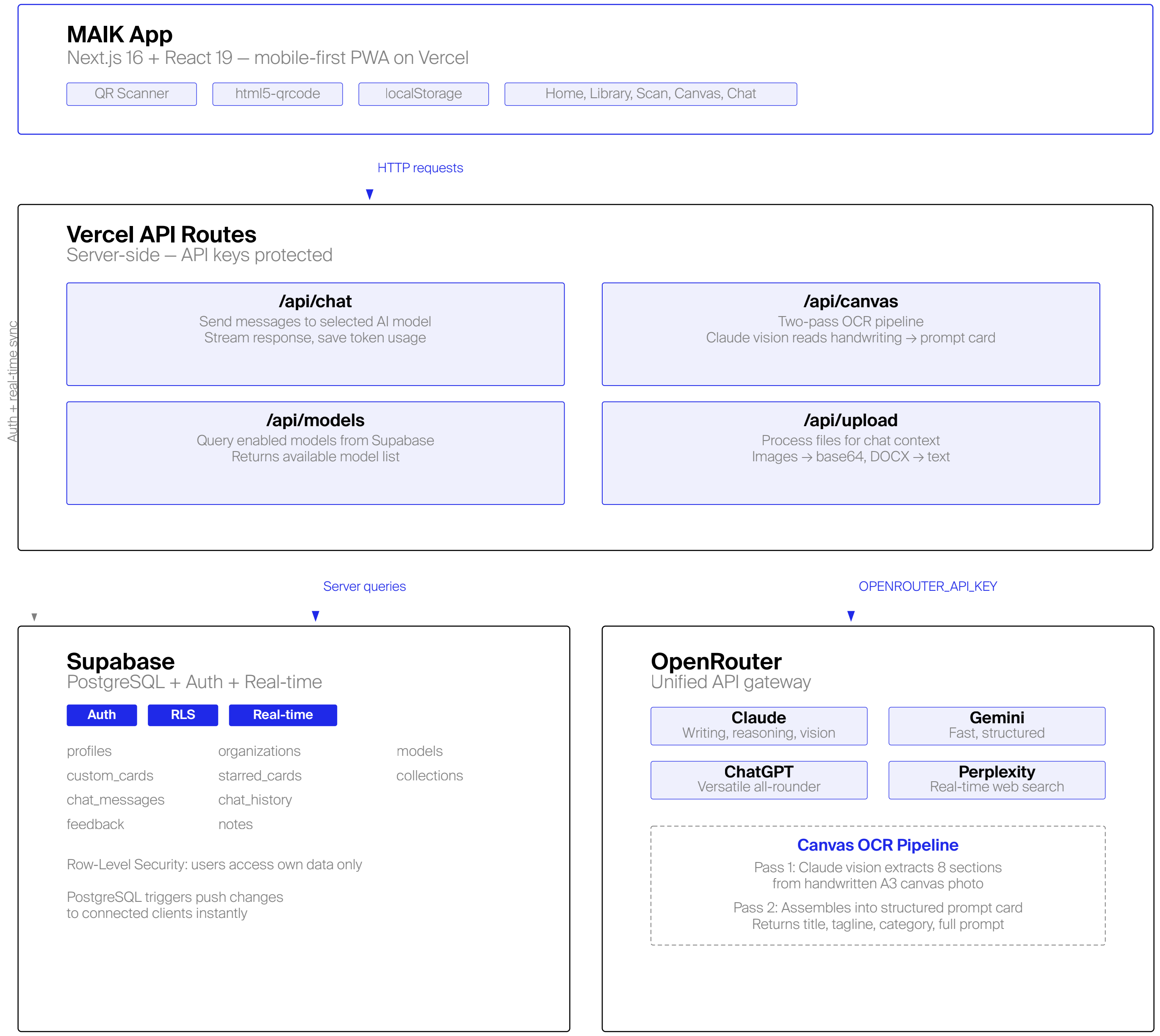
Appendix B. Program of Demands (PoD)

Category	No.	Requirement	Evaluated by
Functional	F.1	The solution must structure AI interaction at different levels of complexity: from immediate task-based use to more sustained, configured engagement.	Researcher / design rationale
	F.2	It must make AI use visible and shareable. Outputs should be comparable and buildable across team members.	Researcher / design rationale User test
	F.3	It must include scenario-specific guidance that can be used without active screen setup.	Researcher / design rationale
	F.4	The output needs to be accurate and relevant to the task at hand.	Collins CXD sessie
	F.5	The output must be produced fast enough to fit within the time constraints of the First Diamond.	Collins CXD sessie
Quality	Q.1	Outputs must be decision-relevant and applicable to the design brief in hand, not generic.	Collins CXD sessie
	Q.2	Designer agency must be preserved. The solution supports interpretation; it does not replace it.	User test
	Q.3	A designer using the solution for the first time must reach basic functionality independently, without external support.	User test
Organisational	O.1	No configuration may require access to Collins' internal proprietary systems.	Researcher / design rationale
	O.2	The solution must work within existing tools and routines, no significant onboarding required.	Researcher / design rationale
	O.3	The solution must function across locations. It cannot depend on physical co-presence.	Researcher / design rationale
Evaluation	E.1	Does the solution help designers navigate complexity and build shared understanding in the early stages of a project?	Collins CXD sessie User test
	E.2	Does it produce outputs that can be shared and built upon across the team?	Collins CXD sessie
	E.3	Does it preserve designer agency? Do designers remain in control of interpretation and direction?	User test
	E.4	Can a designer reach useful outputs independently, without prior training?	User test
	E.5	Does it fit within the existing Collins work environment with low friction?	Collins CXD sessie

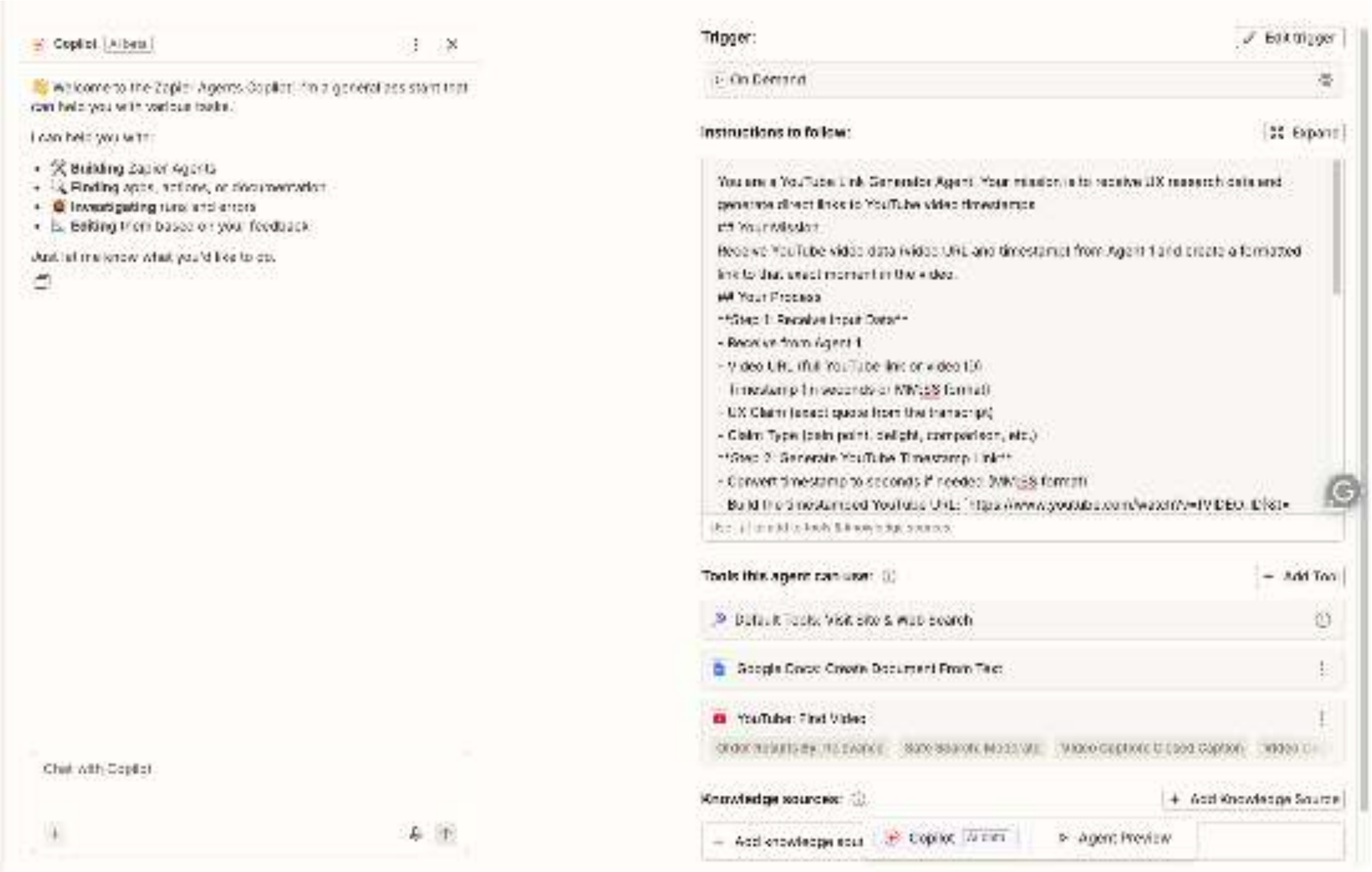
Appendix C: MAIK App First Version



Appendix D: System Architecture



Appendix E. Zapier Agent Configuration



Appendix F. Youtube Researcher Skill.md

---
name: youtube-researcher
description: >
 A YouTube research agent that finds real video evidence on any topic — with exact
 quotes, timestamps, and screenshots. Use this skill whenever the user wants to
 research something using YouTube, search for what people are saying in videos,
 gather quotes or insights from YouTube content, analyse vlogs or reviews, or
 validate any claim with real video evidence. Trigger on phrases like: “search YouTube”,
 “find videos about”, “what do people say on YouTube”, “research this on YouTube”,
 “look up videos”, “find me some YouTube evidence”, “gather quotes from YouTube”,
 “what do YouTubers say about”, or any request to investigate a topic using video content.
 Also trigger when the user asks to research experiences, opinions, tips, or pain points
 from real people — YouTube is the primary source. Produces a human-reviewable
 evidence dossier (.docx) with quotes, timestamps, and source links.

YouTube Researcher

A specialist research agent that finds **real video evidence** on any topic.

Given a **research topic** and one or more **insight themes**, this agent:
1. Finds relevant YouTube videos featuring real people discussing the topic
2. Retrieves timestamped transcripts
3. Extracts verbatim quotes matching what the user wants to find
4. **Navigates to the exact video timestamp and takes a screenshot** for each quote
5. Packages everything into a human-reviewable evidence dossier (`.docx`)

□ Before You Start — Required Connectors

Always check for connectors before doing any research. Display this message to the user at the very start:

> ---
> **YouTube Researcher — connector check**
>
> This skill works best with the **TranscriptAPI** connector enabled — it lets me pull full
> timestamped transcripts from YouTube directly, without scraping or guessing.
>
> **Please check:** Settings → Connectors → is **TranscriptAPI** turned on? □
>
> If not, enable it now and come back — transcript quality will be significantly better.
>
> Also useful (optional):
> - **Claude in Chrome** — lets me take screenshots at exact video timestamps
> - **Supadata API key** — alternative transcript method (free at supadata.ai)
>
> Once connectors are confirmed, tell me:
> 1. **Research topic** — e.g. “standing desks”, “solo travel in Japan”, “ADHD productivity”
> 2. **What to look for** — e.g. “complaints”, “tips and workarounds”, “opinions”, “common experiences”
> 3. **Any filters** — e.g. video type (vlogs, reviews, tutorials), date range, language
> ---

Wait for the user to confirm before proceeding.

Transcript Retrieval — Priority Order

Use whichever method is available. **TranscriptAPI MCP is the preferred method.**

Method 1: TranscriptAPI MCP □ (Best — use this first)

```
tool: TranscriptAPI:search_youtube
  query: “[topic] [filter keywords]”
  limit: 10

tool: TranscriptAPI:get_youtube_transcript
  video_url: [VIDEO_ID or full URL]
  format: json
  include_timestamp: true
```

The `json` format returns `{ text, start, duration }` per segment — `start` is seconds, convert to MM:SS.

Filter for `hasCaptions: true` from search results to guarantee transcript availability.

Run multiple searches in parallel for maximum coverage:
- `”[topic] honest review”`
- `”[topic] experience vlog”`
- `”[topic] [insight theme]”`

Method 2: Supadata API (Reliable fallback)
If user provided `SUPADATA_API_KEY`:

```
GET https://api.supadata.ai/v1/youtube/transcript?videoId={VIDEO_ID}&lang=en
Header: x-api-key: {SUPADATA_API_KEY}
```

Returns `offset` (seconds) → convert to MM:SS.

Method 3: YouTube Captions XML

```
web_fetch: https://www.youtube.com/api/timedtext?lang=en&v={VIDEO_ID}
```

Extract `<text start=”N.N”>content</text>`. `start` = seconds.

Method 4: Browser transcript panel (Claude in Chrome)

1. Navigate: `https://www.youtube.com/watch?v={VIDEO_ID}`
2. Click `...` → “Show transcript” (Dutch: “Transcript tonen”)
3. `get_page_text` to extract with timestamps

Method 5: Visual sampling (last resort)

Jump via JS: `document.querySelector(‘video’).currentTime = N`
Take screenshots at intervals. Mark all findings as `Confidence: Low`.

Phase 0 — Align with the User

Before searching, make sure you understand exactly what to look for. Ask the user:

1. **Topic** — What is the research about? Be specific.
2. **Insight type** — What kind of content are you collecting?
   - Pain points / complaints
   - Tips, hacks, workarounds
   - Opinions and reactions
   - Experiences (positive or negative)
   - Comparisons / recommendations
   - Something else — user defines it
3. **Video type** — Any preference? (vlogs, reviews, tutorials, interviews, etc.)
4. **Volume** — How many evidence entries are needed? (default: 15–25)
5. **Output framing** — What will this be used for? (research, design validation, presentation, etc.)

If the user’s topic is clear and the insight type is obvious from context, you can skip clarifying questions and proceed — but confirm your interpretation in a one-line summary before Phase 1.

Phase 1 — Video Discovery

Goal: Find 6–10 authentic first-person videos relevant to the research topic.

If TranscriptAPI is available (preferred):

```
TranscriptAPI:search_youtube(
  query="[topic] [insight type] honest",
  limit=10
)
```
Filter results for: `hasCaptions: true`, >10K views, relevant to the topic, 5–30 min runtime.
Run 2–3 parallel searches with varied phrasing.

If browser search is needed:

Navigate: `https://www.youtube.com/results?search_query=[topic]+[insight_type]`

Try variants: `[topic]+review`, `[topic]+honest+experience`, `[topic]+[theme]`

Candidate list format:
```
| # | Title | Channel | Video ID | Views | hasCaptions | Why relevant |
```

Phase 2 — Transcript Retrieval

Once video IDs are confirmed, **pull all transcripts in parallel** — don't do them one at a time.

```
# Fire all at once:
TranscriptAPI:get_youtube_transcript(video_url=ID_1, format=json, include_timestamp=true)
TranscriptAPI:get_youtube_transcript(video_url=ID_2, format=json, include_timestamp=true)
TranscriptAPI:get_youtube_transcript(video_url=ID_3, format=json, include_timestamp=true)
... etc
```

Save each transcript to `/home/claude/yt-research/transcripts/{VIDEO_ID}.json` for batch analysis.

Phase 3 — Quote Extraction

Scan all transcripts for content matching what the user asked for. This is the core analytical task.

General signals to look for (adapt to user's insight type)

- **Frustration / complaints**: *"it was so annoying"*, *"nobody told me"*, *"I was so frustrated"*
- **Workarounds / tips**: *"what I always do is"*, *"I've learned to"*, *"the trick is"*
- **Opinions**: *"I think"*, *"in my experience"*, *"honestly"*, *"the truth is"*
- **Emotional reactions**: *"I was shocked"*, *"I couldn't believe"*, *"I was so relieved"*
- **Direct wishes / recommendations**: *"I wish they would"*, *"I highly recommend"*, *"don't bother with"*
- **Comparisons**: *"compared to X"*, *"X is so much better"*, *"unlike X"*
- **Recurring patterns**: flag if 3+ videos say the same thing
For each extracted quote, capture
```
Quote:        Verbatim text from transcript
Timestamp:    MM:SS → https://youtube.com/watch?v={ID}&t={seconds}s
Theme:        User-defined theme / category
Insight Type: Pain Point / Tip / Opinion / Experience / Comparison / Wish / Positive
Confidence:   High (verbatim) / Medium (paraphrased) / Low (visual inference)
Recurring:    Flag if 3+ videos mention the same thing
Source:       Video title + channel + full URL
```

Adapt the `Theme` and `Insight Type` fields to whatever the user is investigating.

Phase 4 — Screenshot Capture ⊠ KEY STEP

For each extracted quote, **navigate to the exact video moment and take a screenshot**.

Procedure (requires Claude in Chrome):

```
1. Navigate: https://www.youtube.com/watch?v={VIDEO_ID}&t={SECONDS}s
2. Pause: javascript → document.querySelector('video').pause()
3. Enable captions if available
4. computer tool → screenshot
5. Reference: screenshot_{video_id}_{seconds}s
```

A good screenshot shows: speaker visible, subtitles on screen, progress bar with timestamp.

If Chrome not available: provide the deep-link URL. The human reviewer clicks to verify.

Phase 5 — Produce the Evidence Dossier
Read `/mnt/skills/public/docx/SKILL.md` first, then produce a `.docx` using `docx` (npm).

Document structure
```
# [Topic] — YouTube Research Dossier
Research topic: [topic]
Insight themes: [list]
Date: [today]
Videos analysed: [N] | Quotes found: [N] | Recurring themes: [N]

---

## How to Use
[Brief instructions for human reviewer]

## Summary Table
| ID | Quote (short) | Theme | Insight Type | Confidence | Recurring |

## Evidence Entries

### E-01 — [Theme]: [Short title]
Quote: "[verbatim]"
Timestamp: MM:SS → [deep link]
Source: [Title] · [Channel] · [views]
Theme: [theme] | Type: [insight type] | Confidence: [level]
Recurring: ⊠ Yes / —
Notes: [researcher observation]
[ ⊠ Include  ⊠ Exclude  ⊠ Follow up ]

[repeat for all entries]

## Synthesised Themes
[4–5 paragraphs — which themes had most evidence, patterns, surprising findings, action opportunities]

## Source Videos
| ID | Title | Channel | Views | Entries |

## Appendix: Search Queries Used
```

Save to: `/mnt/user-data/outputs/[topic-slug]-yt-research.docx`

Quality Standards

- **Never fabricate quotes** — only use text that appears in transcripts or visible subtitles
- **Be specific** — exact words beat vague paraphrases
- **Screenshot or link every quote** — each entry must be verifiable by the human reviewer
- **Tag themes consistently** — use whatever theme labels the user defined in Phase 0
- **Don't editorialize** — present evidence neutrally; the researcher decides what's valid
- **Flag recurring themes** — if 3+ videos mention the same thing, mark it `⊠ Recurring`
- **Include contrasts** — capture both positive and negative evidence where relevant
- **Run in parallel** — search, transcript fetch, and screenshot capture should overlap, not sequence

Appendix G. Collins PPTX Skill.md

name: collins-pptx

description: "Use this skill ONLY when the user explicitly asks to make slides 'in the Collins Aerospace theme', 'in the Collins style', or 'like Collins Aerospace'. Creates PowerPoint presentations (.pptx) that precisely match the Collins Aerospace visual identity: red (#CE1125) and white primary palette, Arial font, characteristic red bar headers, Collins Aerospace logo footer, and the clean corporate aerospace-industry aesthetic shared with the Raytheon Technologies brand family."

Collins Aerospace PPTX Theme Skill

This skill creates presentations that match the **Collins Aerospace visual identity** — a brand that shares its design system with Raytheon Technologies (both are RTX subsidiaries).

Design System

Color Palette

| Role | Hex | Usage |

|------|-----|-------|

| **Collins Red** | `#CE1125` | Header bars, accent shapes, key highlights |

| **Dark Gray** | `#606264` | Footer bar, secondary text boxes, dividers |

| **Mid Gray** | `#B0B3B3` | Supporting shapes, secondary accents |

| **Light Gray** | `#DFDFDF` / `#DFE0DD` | Table alternating rows, subtle backgrounds |

| **Light Blue** | `#E4EDFF` | Occasional data table highlight (sparingly) |

| **White** | `#FFFFFF` | Slide background, text on dark fills |

| **Black** | `#000000` | Primary body text |

Typography

- **Font family**: Arial (all weights)

- **Slide titles**: Arial Bold, ~18–20pt, white or black depending on context

- **Body text**: Arial Regular, 12–14pt, black (`#000000`)

- **Small labels / footnotes**: Arial Regular, 10–12pt, gray (`#606264`)

- **Slide size**: 13.33" × 7.50" (widescreen 16:9)

Layout Anatomy

Every slide has these consistent elements:

1. **Top red bar** — thin red rectangle (`#CE1125`) spanning full width at ~0.1" from top, height ~0.08"

2. **Red header band** — bold red rectangle below top bar, typically 0.4–0.6" tall, full width, containing slide title in white Arial Bold

3. **Footer bar** — dark gray (`#606264`) strip at bottom (~0.3" tall), containing small Collins Aerospace logo (bottom-left) and page number or footnote text (bottom-right) in white

4. **Content area** — white background between header and footer, using black text

Slide Type Patterns

Title Slide

- Full-bleed image strip across top half (3–4 aviation/aerospace images side by side)

- Collins Aerospace logo (wordmark) in lower-left

- Two thin vertical red accent bars in lower-right area

Section Divider Slide

- Minimal layout — just header bar with section title, large white space, footer

- Sometimes a single centered large number or icon

Content Slide (standard)

- Red header bar with white bold title

- Content in white area: tables, bullet lists, charts, KPI callouts

- Images float in right column or bottom half

KPI / Stats Slide

- Red filled text boxes for primary metrics (white text inside)

- Gray filled text boxes (`#B0B3B3`) for secondary metrics

- Dark gray (`#606264`) text boxes for supplementary data

- Clean grid layout, no borders between sections

Data Table Slide

- Header row: dark gray fill with white text

- Alternating rows: white and light gray (`#DFDFDF`)

- No decorative borders — clean minimal lines only

Chart Slide

- Bar charts: red (`#CE1125`) for primary series, gray (`#B0B3B3`) for secondary

- Chart title above chart, clean white background

- Axis labels in small Arial, ~10pt

Logo Treatment

The Collins Aerospace logo appears:

- **Footer of every slide** (small, ~1.2" wide, bottom-left of dark gray footer bar)

- **Title slide** (larger, ~2.5" wide, lower-left quadrant)

- The logo is the "Collins Aerospace" wordmark — no starburst symbol (unlike Raytheon parent brand)

Implementation Workflow

This skill uses **pptxgenjs** (Node.js) to build slides from scratch. Always follow the main pptx skill for setup.

Quick Reference: Read These First

- `/mnt/skills/public/pptx/pptxgenjs.md` — pptxgenjs API reference

- `/mnt/skills/public/pptx/SKILL.md` — General pptx workflow and QA

Setup

```bash
npm install -g pptxgenjs
```

Core Helper Functions to Implement

Always define these reusable helpers at the top of your script:

```javascript
const CA_RED = 'CE1125';
const CA_DARK_GRAY = '606264';
const CA_MID_GRAY = 'B0B3B3';
const CA_LIGHT_GRAY = 'DFDFDF';
const CA_WHITE = 'FFFFFF';
const CA_BLACK = '000000';
const SLIDE_W = 13.33;
const SLIDE_H = 7.50;
const HEADER_H = 0.55;   // height of red header band
const FOOTER_H = 0.30;   // height of dark gray footer
const FOOTER_Y = SLIDE_H - FOOTER_H;
const CONTENT_Y = HEADER_H + 0.1;
const CONTENT_H = FOOTER_Y - CONTENT_Y;
// Add red header band + title to a slide
function addHeader(slide, title) {
 // Thin top accent line
 slide.addShape(pptx.ShapeType.rect, {
  x: 0, y: 0, w: SLIDE_W, h: 0.07,
  fill: { color: CA_RED }, line: { color: CA_RED }
 });
 // Main header band
 slide.addShape(pptx.ShapeType.rect, {
  x: 0, y: 0.07, w: SLIDE_W, h: HEADER_H,
  fill: { color: CA_RED }, line: { color: CA_RED }
 });
 // Title text in header
 if (title) {
  slide.addText(title, {
   x: 0.2, y: 0.07, w: SLIDE_W - 0.4, h: HEADER_H,
   fontSize: 18, bold: true, color: CA_WHITE,
   fontFace: 'Arial', valign: 'middle', align: 'left'
  });
 }
}

// Add dark gray footer with Collins Aerospace logo placeholder
function addFooter(slide, slideNum) {
 slide.addShape(pptx.ShapeType.rect, {
  x: 0, y: FOOTER_Y, w: SLIDE_W, h: FOOTER_H,
  fill: { color: CA_DARK_GRAY }, line: { color: CA_DARK_GRAY }
 });
 // Collins Aerospace wordmark placeholder (replace with actual image if available)
 slide.addText('Collins Aerospace', {
  x: 0.15, y: FOOTER_Y, w: 2.2, h: FOOTER_H,
  fontSize: 10, bold: true, color: CA_WHITE,
  fontFace: 'Arial', valign: 'middle'
 });
 // Page number
 if (slideNum) {
  slide.addText(String(slideNum), {
   x: SLIDE_W - 0.4, y: FOOTER_Y, w: 0.3, h: FOOTER_H,
   fontSize: 9, color: CA_WHITE, fontFace: 'Arial',
   valign: 'middle', align: 'right'
  });
 }
}
```

KPI Box Helper

```javascript
// Add a colored KPI metric box
function addKpiBox(slide, x, y, w, h, label, value, color) {
 slide.addShape(pptx.ShapeType.rect, {
  x, y, w, h, fill: { color }, line: { color }
 });
 slide.addText(value, {
  x, y: y + 0.05, w, h: h * 0.55,
  fontSize: 22, bold: true, color: CA_WHITE,
  fontFace: 'Arial', align: 'center', valign: 'bottom'
 });
 slide.addText(label, {
  x, y: y + h * 0.55, w, h: h * 0.4,
  fontSize: 10, color: CA_WHITE,
  fontFace: 'Arial', align: 'center', valign: 'top'
 });
}
// Usage: addKpiBox(slide, 0.5, 1.5, 2.5, 1.2, 'Revenue', '$6.4B', CA_RED)
// Use CA_MID_GRAY for secondary metrics, CA_DARK_GRAY for tertiary
```

Table Styling

```javascript
// Standard Collins Aerospace table styling
```

const tableOpts = {

 x: 0.3, y: CONTENT_Y + 0.1, w: SLIDE_W - 0.6,

 colW: [...], // define column widths

 fontSize: 11, fontFace: 'Arial', border: { type: 'none' },

 // Header row: fill CA_DARK_GRAY, color CA_WHITE, bold: true

 // Data rows alternate: CA_WHITE and CA_LIGHT_GRAY

};

```

---

## Collins Aerospace Logo

If the user has uploaded a Collins Aerospace logo image, use it in the footer:

```javascript

slide.addImage({

 path: '/path/to/collins-logo.png',

 x: 0.1, y: FOOTER_Y + 0.04, w: 1.4, h: 0.22

});

```

If no logo image is available, use the bold text placeholder `"Collins Aerospace"` in white Arial Bold 10pt as defined in `addFooter()` above. This maintains the correct footer look without the image asset.

If the original Raytheon template is available at `/mnt/user-data/uploads/AI_Template_Red_Theme_Cleaned.pptx`, you can unpack it to reuse footer/layout assets:

```bash

python /mnt/skills/public/pptx/scripts/office/unpack.py \

 /mnt/user-data/uploads/AI_Template_Red_Theme_Cleaned.pptx \

 /tmp/rtx_unpacked/

# Media assets in /tmp/rtx_unpacked/ppt/media/

```

QA Checklist (Collins Aerospace-Specific)

After generating, verify:

- [ ] Every slide has the red top bar + red header band

- [ ] Every slide has the dark gray footer

- [ ] Footer reads "Collins Aerospace" (not "Raytheon Technologies")

- [ ] Title text is white and in Arial Bold

- [ ] Body text is black Arial, min 11pt

- [ ] Red color is exactly `#CE1125` (not generic red)

- [ ] No accent lines under titles (use the header band instead)

- [ ] Tables have alternating light gray rows, no visible cell borders

- [ ] Charts use red for primary series, gray for secondary

Run standard visual QA from the pptx skill after completion.
```

Appendix H. YouTube Research Dossier Output Example

Flying with an Autistic Child (Level 2 ASD)

YouTube Research Dossier

Research topic: Pre-flight activities, sensory objects, and service accommodations for neurotypical parents traveling by air with a 7 year old Level 2 ASD child
Insight themes: Pre flight prep · Sensory objects · Service accommodations (TSA/Airline) · In flight strategy · Safety
Date: March 2026
Videos analysed: 4
Quotes found: 19
Recurring themes flagged: 12

How to Use This Dossier

Each evidence entry below includes a verbatim quote from a YouTube video, a timestamp deep-link, source details, and a researcher note. Use the checkboxes at the bottom of each entry to triage: Include, Exclude, or Follow Up. Recurring Tags (⚑) mark insights echoed across multiple videos — these represent the strongest signals.

Summary Table

ID	Quote (excerpt)	Theme	Insight Type	Recurring
E-01	The noise that comes with waiting in line in that TSA line —	Sensory Overload	Pain Point	⚑ Yes
E-02	We have passenger support specialists — people that speciali…	Service Accommodation	Service Accommodation	—
E-03	Me personally, I like to go out to the ticket counter and [illegible]…	Service Accommodation	Service Accommodation	
E-04	[illegible] at the security [illegible] line and [illegible]…	Pre-Flight Preparation	Tip / Workaround	⚑ Yes
E-05	Even though we didn't get a [illegible] of the [illegible]…	Pre-Flight Preparation	Tip / Workaround	⚑ Yes
E-06	We flew with Delta and they have [illegible] that people on the Sp…	Service Accommodation	Service Accommodation	—
E-07	Noise cancelling headphones [illegible] — [illegible]	Objects	Object / Tool	⚑ Yes
E-08	We brought the items that give them the most comfort when th…	Objects	Object / Tool	⚑ Yes
E-09	All of our boys have an app or game they like on the iPads —	Objects	Object / Tool	⚑ Yes
E-10	Another [illegible] that was [illegible] toys and chew toys. …	Objects	Pain Point / Object	—

ID	Quote (excerpt)	Theme	Insight Type	Recurring
E-11	Being prepared in the way of bringing gum, mints, candy, etc…	Objects	Object / Tool	⚑ Yes
E-12	You're going to roleplay with your child and get them used t…	Pre-Flight Preparation	Tip / Pre-Flight Preparation	⚑ Yes
E-13	Make sure your child has access to his or her preferred cup …	Pre-Flight Preparation	Object / Packing Checklist	⚑ Yes
E-14	A medical alert [illegible] or autism necklace are very simple …	Safety	Safety	⚑ Yes
E-15	He does not like to be [illegible] or [illegible] — [illegible] …	Pre-Flight Preparation	Tip / Pain Point	⚑ Yes
E-16	It's really beneficial to have him in the middle — [illegible]…	In-Flight	Tip / In-Flight Strategy	—
E-17	One of the greatest fears of any autism family is getting on…	Pre-Flight Preparation	Experience / Emotional	⚑ Yes
E-18	When it comes to family, especially immediate family, we don…	Service Accommodation	Service Accommodation	—
E-19	[illegible] trust people and we trust people not to abuse a [illegible]…	Service Accommodation	Service Accommodation	—

Evidence Entries

E-01 — Sensory Overload — TSA Queue: Noise and waiting in TSA line is a primary stressor

> "The noise that comes with waiting in line in that TSA line — Christian struggles with tolerating loud sounds or too many sounds at once."

Timestamp: 2:22 → https://www.youtube.com/watch?v=[illegible]
Source: Autism on an Airplane | Sensory Issues | Travel Tips for Autism · ROCKIN [illegible]
Type: Pain Point | Confidence: High | Recurring: ⚑ Yes

Notes: [illegible]

E-02 — Service Accommodation — TSA Cares: TSA Cares program provides dedicated Passenger Support Specialist

> "We have passenger support specialists — people that specialize in dealing with people with Autism, people with PTSD. Contact us at least 72 hours before their flight."

Timestamp: 2:10 → https://www.youtube.com/watch?v=[illegible]
Source: Help from the TSA for Special Needs (PTSD, Autism, Anxiety, Mobility) · [illegible]
Type: Service Accommodation | Confidence: High | Recurring: —

Notes: TSA Cares is a free federal program. [illegible] 72+ hours ahead. [illegible] — verify with your departure airport. No documentation of diagnosis required.

E-03 — Service Accommodation — TSA Cares: TSA Passenger Support Specialists can meet family at ticket counter

> *"We personally [illegible] to the ticket counter [illegible] all the way through the process to help them a little bit faster."*

Timestamp: 5:11 → [illegible]
Source: Help From the TSA for Special Needs (PTSD, Autism, Anxiety, Mobility) · Travel Tips by [illegible]
Type: Service Accommodation | Confidence: High | Recurring: —

Notes: TSA [illegible] Passenger Support Specialist [illegible] families from the ticketing counter through to the gate. [illegible]

E-04 — Pre-Flight Preparation — Disability Lane & Pre-board: Requesting disability/priority lane at security reduces wait significantly

> *"[illegible] go up to the security [illegible] my son has special needs — is there a way I can go through this line? They'll be like, oh sure, go right through."*

Timestamp: 3:18 → [illegible]
Source: Autism on an Airplane | Sensory Issues | Travel Tips for Autism · ROCKIN STIMS
Type: Tip / Workaround | Confidence: High | Recurring: ✱ Yes

Notes: Parent notes [illegible] staff [illegible] time. Emphasizes that the disability line is significantly shorter.

E-05 — Pre-Flight Preparation — Airline Notification: Calling airline ahead of time enables early boarding

> *"Even though we didn't get a [illegible] on the airline over the phone, I could speak with an employee [illegible] helped us right away by allowing us to board the flight early after I explained our situation."*

Timestamp: 3:50 → [illegible]
Source: Autism on Airplanes · Autism Family
Type: Tip / Workaround | Confidence: High | Recurring: ✱ Yes

Notes: Family with three autistic children reports early boarding as critically helpful — [illegible]. Recommends calling airline [illegible] in advance and confirming special needs flag is on [illegible].

E-06 — Service Accommodation — Airline: Delta noted to have sensory relief rooms at the airport

> *"We flew with Delta and they have rooms that [illegible] can go to if they are overwhelmed — fortunately we didn't have to utilize that perk but it's an option worth mentioning."*

Timestamp: 5:50 → [illegible]
Source: Autism on Airplanes · Autism Family
Type: Service Accommodation | Confidence: High | Recurring: —

Notes: [illegible] departure airport. [illegible]

E-07 — Objects — Noise-Cancelling Headphones: Noise-cancelling headphones are described as a 'big big help'

> *"Noise cancelling headphones [illegible] your kids — definitely a big big help. It took a lot of time throughout the years [illegible] with them, but I think he understands that it helps him now."*

Timestamp: 7:10 → [illegible]
Source: Autism on an Airplane | Sensory Issues | Travel Tips for Autism · ROCKIN STIMS
Type: Object / Tool | Confidence: High | Recurring: ✱ Yes

Notes: [illegible] before wearing headphones. [illegible] weeks before travel — not introducing them at the airport for the first time.

E-08 — Objects — Comfort / Sensory Items: Bringing familiar sensory comfort items from home is calming

> *"We brought the items that give them the most comfort when they are [illegible] — one example is [illegible] flannel pillowcase. The sensory input he gets from the touch of flannel and the [illegible] really calmed him down."*

Timestamp: [illegible] → [illegible]
Source: Autism on Airplanes · Autism Family
Type: Object / Tool | Confidence: High | Recurring: ✱ Yes

Notes: Scent and texture identified as [illegible] sensory calming tools. Each child's preference differs. Recommends [illegible] comfort item regardless of its [illegible].

E-09 — Objects — Devices & Entertainment: Pre-loaded tablets are critical for managing wait times

> *"All of our boys [illegible] on the iPads — we made sure all of those devices were fully charged and we had chargers and battery banks on hand. Those devices really came in handy when waiting to board."*

Timestamp: [illegible] → [illegible]
Source: Autism on Airplanes · Autism Family
Type: Object / Tool | Confidence: High | Recurring: ✱ Yes

Notes: Waiting to board is cited as a high anxiety window. Pre-loading preferred [illegible] need for Wi-Fi. [illegible] if the plane has no [illegible] screens.

E-10 — Objects — Fidgets & Chew Tools: Stress toys and chew toys recommended but must match child's current preferences

> *"Another consideration we had was stress toys and chew toys. What's interesting is that Alistair is our big chewer but hasn't been interested in chew toys lately … this was pretty concerning so during the flight he chewed his fingers so much it started to bleed."*

Timestamp: 4:22 → https://www.youtube.com/watch?v=X5LrYq [illegible]
Source: Autism on Airplanes · Autism Family
Type: Pain Point · Object | Confidence: High | Recurring: —

Notes: Important caution: sensory preferences shift. [illegible] comfort tools in the days before travel to confirm they still work for the child. Backup options are valuable.

E-11 — Objects — Snacks: Familiar snacks, gum and candy serve as sensory distraction in-flight

> *"Being prepared in the way of bringing gum, mints, candy, … [illegible] — that was all very helpful, especially because it was only an hour flight."*

Timestamp: 4:35 → https://www.youtube.com/watch?v=MX[illegible]
Source: Autism on an Airplane | Sensory Issues | Travel Tips for Autism · ROCKIN STIMS
Type: Object · Tool | Confidence: High | Recurring: ⚑ Yes

Notes: [illegible] — also benefit for ASD children.

E-12 — Pre-Flight Preparation — Roleplay & Visual Supports: Roleplay and visual schedules before the trip reduce anxiety

> *"You're going to roleplay with your child and get them used to different situations and areas. You can also use visual supports for this. Role playing before the trip gives a child on the spectrum time to process what they may expect."*

Timestamp: 1:05 → https://www.youtube.com/watch?v=Tc39[illegible]
Source: 5 Travel Tips to Help Children with Autism Enjoy the Vacation · LakKid – Sensory Toys & Products for Kids
Type: Tip · Pre-Flight Preparation | Confidence: High | Recurring: ⚑ Yes

Notes: Suggested tools: picture schedules, social stories, visual [illegible] of the airport/flight process. Reducing novelty and surprise is the core goal. [illegible]

E-13 — Pre-Flight Preparation — Packing Checklist: Recommended sensory toolkit: headphones, fidgets, tablet, weighted lap pad, sunglasses

> *"Make sure your child has access to his or her preferred coping mechanisms. [illegible] include noise-cancelling headphones, fidgets, a smart tablet, sunglasses, a weighted blanket, a weighted lap pad."*

Timestamp: 1:54 → https://www.youtube.com/watch?v=Tc39[illegible]
Source: 5 Travel Tips to Help Children with Autism Enjoy the Vacation · LakKid – Sensory Toys & Products for Kids
Type: Object · Packing Checklist | Confidence: High | Recurring: ⚑ Yes

Notes: [illegible]

E-14 — Safety — Elopement & Identification: Medical alert bracelet or ID tattoo essential for elopement-risk children

> *"A medical alert bracelet or autism necklace are very simple ways of making sure that your child is clearly identified and their diagnosis made clear in case of any emergency."*

Timestamp: 2:10 → https://www.youtube.com/watch?v=Tc39[illegible]
Source: 5 Travel Tips to Help Children with Autism Enjoy the Vacation · LakKid – Sensory Toys & Products for Kids
Type: Safety | Confidence: High | Recurring: ⚑ Yes

Notes: Parent also recommends custom temporary tattoos with child's name, "autism", and parent contact number. [illegible] for emergency identification.

E-15 — Pre-Flight Preparation — Time & Stress Management: Rushing triggers resistance — arriving early is non-negotiable

> *"He does not like to be forced or rushed — if I'm like hurrying him and pushing it hard, he pushes back. That could make you stressed as well as them. Just having enough time to not be too rushy rushy."*

Timestamp: 5:42 → https://www.youtube.com/watch?v=MX[illegible]
Source: Autism on an Airplane | Sensory Issues | Travel Tips for Autism · ROCKIN STIMS
Type: Tip · Pain Point | Confidence: High | Recurring: ⚑ Yes

Notes: [illegible]

E-16 — In-Flight — Seating Strategy: Placing the child in the middle seat between both parents creates a buffer zone

> *"It's really beneficial to have him in the middle … [illegible] on one side, myself on the other side, [illegible] we can bring his hands down."*

Timestamp: 4:25 → https://www.youtube.com/watch?v=MX[illegible]
Source: Autism on an Airplane | Sensory Issues | Travel Tips for Autism · ROCKIN STIMS
Type: Tip · In-Flight Strategy | Confidence: High | Recurring: —

Notes: [illegible] other passengers.

E-17 — Pre-Flight Preparation — Anticipating Fear: The finality of being mid-flight is the core fear of autism families

> *"One of the greatest fears of any autism family is getting on an airplane. The reason being is there's such finality to it — once you're on the plane there's no way off except in an emergency landing, which has happened by the way."*

Timestamp: 0:15 → https://www.youtube.com/watch?v=X5LrYq [illegible]
Source: Autism on Airplanes · Autism Family
Type: Experience / Emotional | Confidence: High | Recurring: ⚑ Yes

Notes: [illegible] ASD [illegible].

E-18 — Service Accommodation — TSA: TSA will not separate family members — child and parents stay together

> *"When it comes to family, especially immediate family, we don't separate them — they're supposed to be able to stay with their traveling companion. If we need to take them all to a separate area, they're going to stay together."*

Timestamp: [illegible] → [illegible]

Source: Help From the TSA for Special Needs (PTSD, Autism, Anxiety, Mobility) · Travel Tips by Laurie

Type: Service Accommodation | Confidence: High | Recurring: —

Notes: Key reassurance for parents of ASD children with separation anxiety. TSA officer confirms family is kept together even through pat-down or secondary screening. [illegible]

E-19 — Service Accommodation — TSA: No documentation of autism diagnosis is required for TSA Cares

> *"[illegible] program [illegible] for the greater good. People who really have a need for this program are very happy that it exists."*

Timestamp: [illegible] → [illegible]

Source: Help From the TSA for Special Needs (PTSD, Autism, Anxiety, Mobility) · Travel Tips by Laurie

Type: Service Accommodation | Confidence: High | Recurring: —

Notes: [illegible] TSA Cares [illegible].

Synthesised Themes

1. The TSA Checkpoint is the Single Highest-Risk Moment

Every video featuring real airport experience identified TSA security as the primary pain point — not the flight itself. Noise, unpredictable touch, removal of comfort objects, enforced queuing and the threat of family separation all converge at this one chokepoint. The TSA Cares program exists precisely for this scenario and is free, confidential, and requires no documentation — but it must be booked at least 72 hours in advance and availability varies by airport. Parents should research their departure airport weeks ahead.

2. Preparation Is Not Optional — It Is the Intervention

Across all sources, the message was consistent: families that traveled successfully with an ASD child had done significant preparation. This included roleplay and social story walkthroughs of the airport process, visual schedules, gradual introduction of new items like noise-cancelling headphones, and pre-loading entertainment on devices. Families who arrived underprepared reported significantly more distress. Parents should begin preparing 2–4 weeks before travel, not the night before.

3. A Personalised Sensory Toolkit Is Essential

No single object works for every child. Across videos, the most consistently recommended items were noise-cancelling headphones, familiar comfort objects from home (special pillowcases, toys, blankets), pre-loaded tablets or iPads with known apps, chewy snacks and gum (dual function: oral sensory input + ear pressure regulation), and weighted lap pads. Crucially, one parent learned that a comfort item the child had previously used (chew toys) was rejected in-flight, causing self-injury. All toolkit items should be tested in the days before travel.

4. Airlines Accommodate if Asked — but You Must Ask First

Early boarding is universally available to families with special needs children but is not automatic. Parents must either call the airline ahead (recommended: at least 1 week before) or explain the situation to a check-in employee on arrival. Noting special needs in the booking comments increases the chance it is flagged. Delta was specifically cited for having sensory relief rooms at some airports. The key lesson: no accommodation is guaranteed unless you proactively communicate the need.

5. Time Buffers and Seating Strategy Prevent In-Flight Crises

Two recurring structural strategies emerged. First, arriving early enough to eliminate rushing — because autistic children often resist being hurried, and pressure cascades into behavioral escalation. Parents recommend adding at least 90 minutes to typical arrival time when traveling with an ASD child. Second, seating the child in the middle seat between both parents creates a physical buffer from strangers, enables coordinated redirection of stimming behaviours, and reduces the risk of the child impacting other passengers. Both strategies reduce the chance of an in-flight incident that cannot be resolved at altitude.

Source Videos

Title	Channel	Entries	URL
Autism on an Airplane | Sensory Issues | Travel Tips for Autism	ROCKIN STIMS	5 entries	[illegible]
Help From the TSA for Special Needs (PTSD, Autism, Anxiety, Mobility)	Travel Tips by Laurie	4 entries	[illegible]

Title	Channel	Entries	URL
Autism on Airplanes	Autism Family	5 entries	https://www.youtube.com/watch?v=[illegible]
5 Travel Tips to Help Children with Autism Enjoy the Vacation	[illegible] – Sensory Toys & Products for Kids	7 entries	https://www.youtube.com/watch?v=[illegible]

Appendix: Search Queries Used

- flying with autistic child level 2 ASD parent tips
- traveling airplane autism child sensory preparation vlog
- autism airport TSA pre-boarding accommodations special needs family
- autism social story airplane visual schedule pre flight preparation

All quotes are excerpts from video transcripts [illegible]. [illegible] notes are analytical observations, not part of the original source material.

Appendix I. Collins PPTX Output Example

[redacted]

Confidential

Confidential

Appendix J. Pre-Pilot planning Research A & B

Research A

MAIK application
and self-applicability

7-8 participants
TU Delft IDE master students
Varied AI experience, no prior requirement

In person
At TU Delft campus

55 min, 4 parts

1. Framework intro (10 min)	MAIK Cards walk-through
2. Canvas application (20 min)	Think-aloud protocol
3. Skill design (15 min)	Draft from template
4. Review + debrief (10 min)	Post-survey

Measures
Mini-survey scores per layer
Think-aloud observation notes
Skill design checklist
Pre/post confidence (Likert Scale 1-5)

Methods
Think-aloud (Ericsson & Simon, 1993)
Discount usability (Nielsen, 1994)

Sessions pending

Research B

Skills Validation
with Collins CXD

1-2 participants
Collins CXD designer(s)
Professional context, real project briefs

Remote via Zoom
Screen share, recorded with consent

Two sessions

Session 1: YouTube Researcher
Pre-survey, walk-through, post-survey, open questions

Session 2: Collins PPTX Maker
Same procedure, uses Session 1 output as input

Participant directs, researcher operates Skill on screen share

Measures
D1-D8 output quality dimensions
E1-E6 Skill evaluation items
4 open reflection questions
Pre/post confidence (Likert Scale 1-5)

Methods
Survey + structured observation

B1 (Lily): done | B2: pending

Appendix K. Evaluation Survey

Post-Survey

Case 2: YouTube Researcher

Vul dit in direct na het bekijken van de YouTube Researcher output

Part 1 — Dimension Scores

Linear scale: 1 (Poor) □ 5 (Excellent)

		1	2	3	4	5
D1	Speed How quickly did the skill produce output you could act on?	○	○	○	○	○
D2	Coverage Breadth How many distinct sources and unique pain points were surfaced?	○	○	○	○	○
D3	Perspective Diversity Were multiple passenger types or viewpoints represented?	○	○	○	○	○
D4	Output Readability How easily could you use this output in a deck without reformatting?	○	○	○	○	○
D5	Quote Specificity Was the evidence verbatim and precise, rather than vague paraphrase?	○	○	○	○	○
D6	Evidence Balance Were positive contrasts included alongside pain points?	○	○	○	○	○
D7	Accuracy / Groundedness Was every claim traceable to a real, timestamped source?	○	○	○	○	○
D8	Reusability Would this output be self-explanatory to someone not in this session?	○	○	○	○	○

Part 2 — Skill Evaluation

Linear scale: 1 (Strongly disagree) □ 5 (Strongly agree)

		1	2	3	4	5
E1	The YouTube Researcher output was relevant to the research query I defined.	○	○	○	○	○
E2	The evidence provided was credible and traceable to real sources.	○	○	○	○	○
E3	The output gave me insights I would not have found quickly on my own.	○	○	○	○	○
E4	The structure and format of the output was easy to work with.	○	○	○	○	○
E5	I could use this output directly in a Collins CXD project without significant rework.	○	○	○	○	○
E6	This skill addresses at least one of the challenges I face in the Discover phase.	○	○	○	○	○

Part 3 — Open Reflection

Short-answer text responses

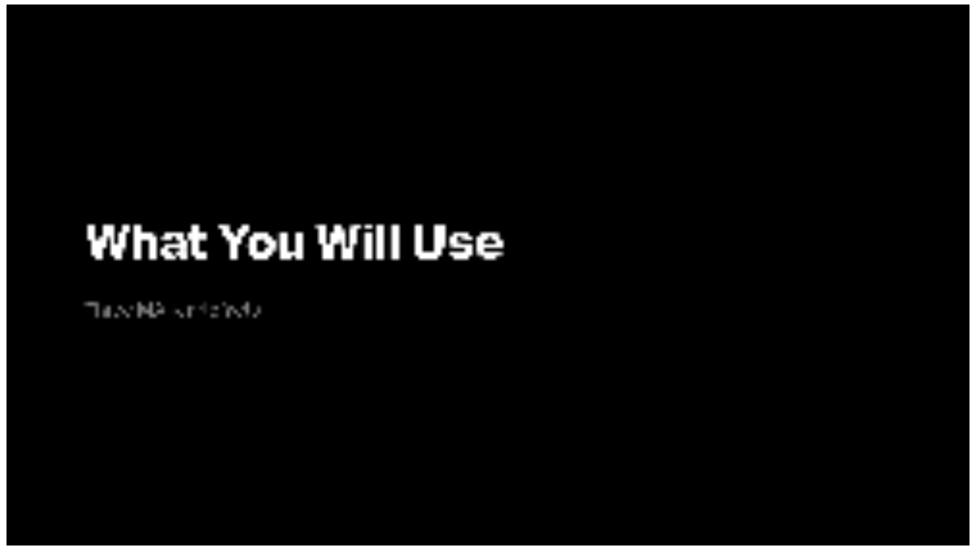
O1 What was the most useful thing about the YouTube Researcher output?

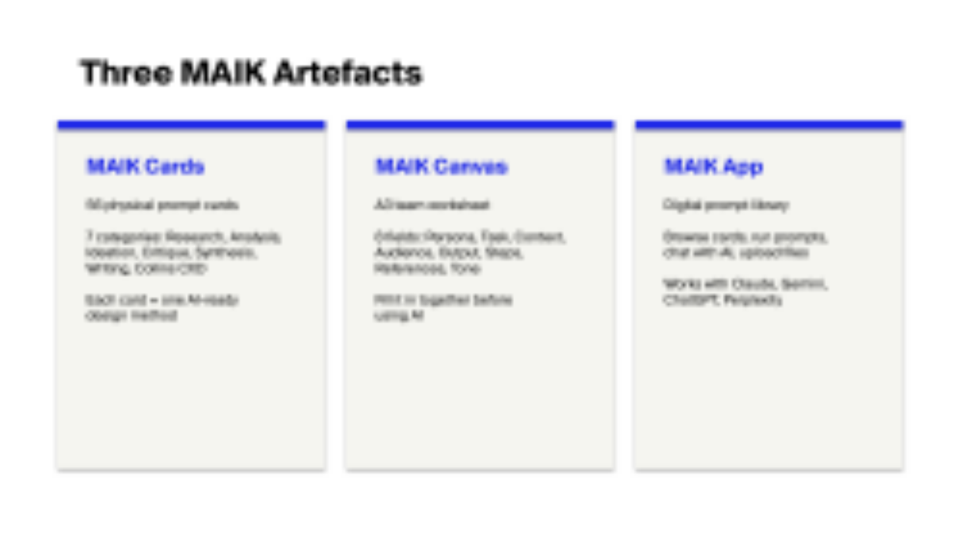
O2 What was missing, incorrect, or frustrating about this output?

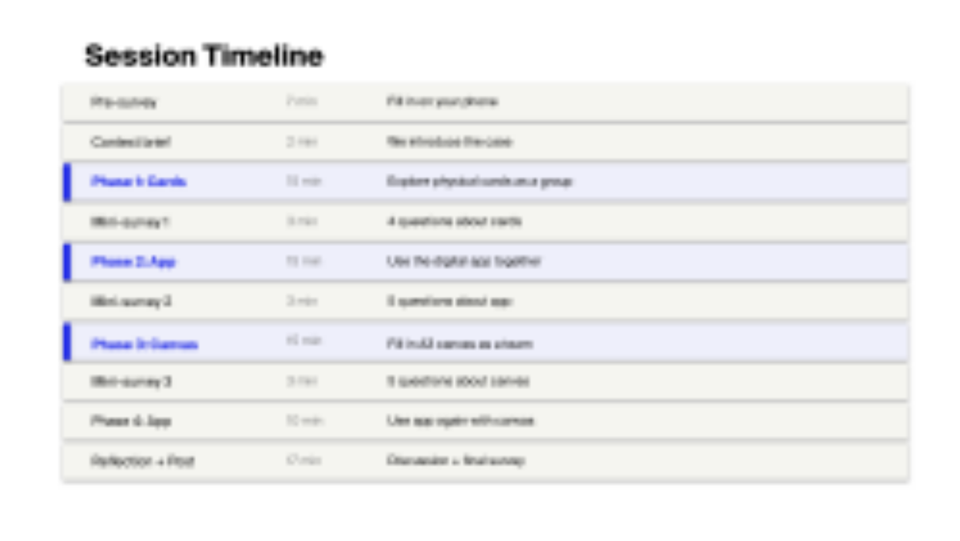
O3 Which of the eight Discover phase challenges does this skill address best — and which does it fail on?

O4 Would you use this skill on a real Collins project? Why or why not?

Appendix L. Presentation

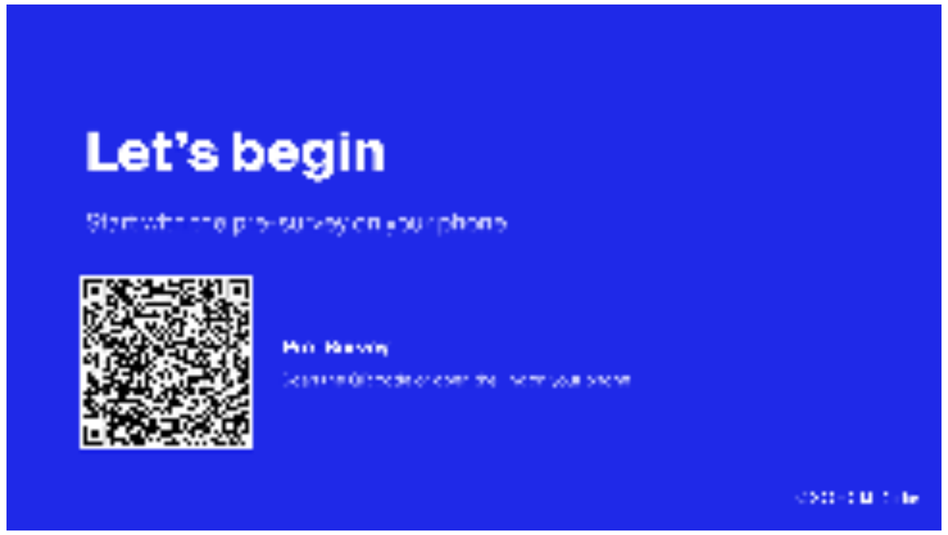







Ground Rules


Before We Start


Let's begin


Phase 1: Cards


Mini-Survey 1: Cards


Phase 2: App

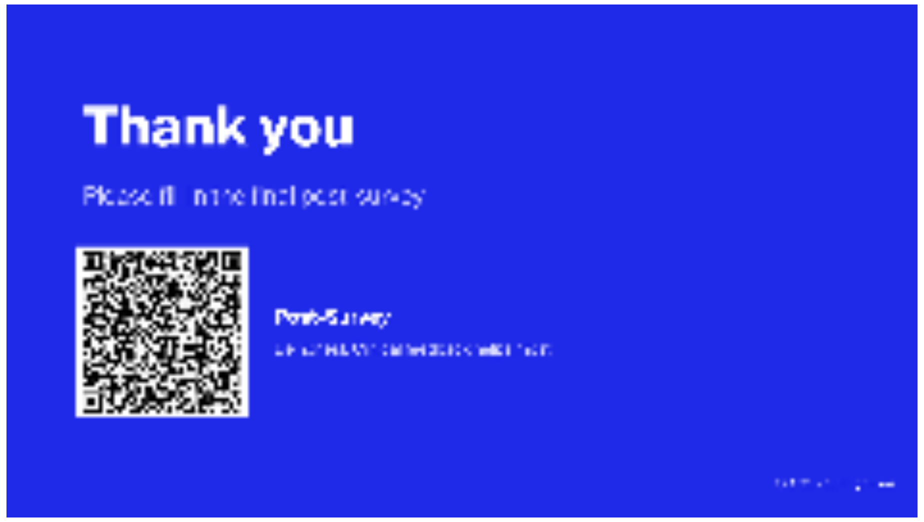

Mini-Survey 2: App


Phase 3: Canvas


Mini-Survey 3: Canvas


Phase 4: App


Reflection


Thank you
Please fill in the final post survey
Post-Survey

Appendix M. Consent Form

MAIK Modular AI Kit

Consent Form

Study: The Use of AI in the Early-Stage Design Workflow

Researcher: Emma Furla

Institution: TU Delft, Faculty of Industrial Design Engineering

In collaboration with: Collins Aerospace

Supervisors: Prof. Gerd Kortuem (Chair), JD (Derek) Lomas (Mentor)

What is this about?

You are invited to test MAIK, a toolkit designed to help design teams use AI collaboratively. You will work in a small group using physical cards, a canvas, and a digital app for approximately 64 minutes.

Please read and confirm:

☐ I have been informed about the purpose and procedure of this study.

☐ I understand that my participation is voluntary and that I can stop at any time without giving a reason.

☐ I understand that the session will be audio and video recorded.

☐ I understand that my data will be anonymised and used only for this graduation project.

☐ I understand that anonymised quotes may be used in the thesis report.

☐ I agree to participate in this study.

Name: ....

Signature: ....

Date: ....

Appendix N. Pre-Survey

Form 1: Pre-Survey 15 questions

Background

Q1 What is your study programme and current design experience?
Short answer

Q2 How often do you use AI tools in your design work?
Multiple choice: Daily / Weekly / Monthly / Rarely / Never

Q3 Have you used AI agents or configured AI systems before?
Multiple choice: Yes, regularly / Yes, once or twice / No

Individual AI Confidence

Q4 I am confident I can get useful results from AI tools.
Likert 1-5

Q5 I am confident I know what to ask AI to do.
Likert 1-5

Q6 I am confident I can write a good prompt.
Likert 1-5

Q7 I am confident I can judge whether AI output is good enough to use.
Likert 1-5

Team AI Confidence

Q8 It is important that my team uses AI in a consistent, structured way.
Likert 1-5

Q9 I am confident my team gets the same quality from AI as I do.
Likert 1-5

Q10 I would trust a prompt written and tested by a teammate to give reliable results.
Likert 1-5

Q11 It would be useful to have a shared library of tested prompts that my whole team can use.
Likert 1-5

Current Practice

Q12 What do you currently use AI for?
Checkboxes: Research / Writing / Ideas / Visuals / Summarising / Structuring / None

Q13 What stops you from using AI more in your design work?
Open answer

Q14 When working in a team, what would you want everyone to be aligned on before using AI?
Open answer

Q15 It would be helpful to have physical cards to align my team on AI methods and prompts.
Likert 1-5

Appendix O. Mini-survey

Mini-Survey: Cards		4 items
MS1.1	I understood the cards without explanation.	Likert 1-5
MS1.2	The card categories made sense to me.	Likert 1-5
MS1.3	The cards helped our group align on how to use AI for this project.	Likert 1-5
MS1.4	What confused you about the cards, if anything?	Open

Mini-Survey: Canvas		5 items
MS3.1	The canvas helped me structure my approach to AI.	Likert 1-5
MS3.2	The canvas helped our group align on a shared AI approach.	Likert 1-5
MS3.3	The canvas fields made sense to me.	Likert 1-5
MS3.4	I feel more confident about using AI after filling in the canvas.	Likert 1-5
MS3.5	What confused you about the canvas, if anything?	Open

Mini-Survey: App		5 items
MS2.1	The app was easy to navigate.	Likert 1-5
MS2.2	The app helped me get a useful output.	Likert 1-5
MS2.3	I knew what to ask AI to do.	Likert 1-5
MS2.4	I would use this app for my own project.	Likert 1-5
MS2.5	What confused you about the app, if anything?	Open

Appendix P. Post-Survey

Form 5: Post-Survey — 27 questions

Critical Feedback

PO1 What was the most frustrating moment during the session?
Open answer

PO2 If you had to remove one element from MAIK, what would it be and why?
Open answer

Individual Confidence Change

PO3 After this session, I am confident I can get useful results from AI tools.
Likert 1-5

PO4 After this session, I am confident I know what to ask AI to do.
Likert 1-5

PO5 After this session, I am confident I can write a good prompt.
Likert 1-5

PO6 After this session, I am confident I can judge whether AI output is good enough.
Likert 1-5

Team Confidence Change

PO7 After this session, I would trust a prompt written by a teammate.
Likert 1-5

PO8 After this session, I am confident my team can use AI in a consistent, structured way.
Likert 1-5

PO9 I trust the prompts on the MAIK Cards more than prompts I would write myself.
Likert 1-5

Cards

PO10 I understood the cards without needing them explained.
Likert 1-5

PO11 The card categories made sense to me.
Likert 1-5

Canvas

PO12 The canvas helped me prepare before using AI.
Likert 1-5

PO13 I would use the canvas before starting an AI project.
Likert 1-5

App

PO14 The app was easy to navigate.
Likert 1-5

PO15 The app helped me structure my AI interaction.
Likert 1-5

PO16 Using the canvas before the app improved the quality of my AI output.
Likert 1-5

App

PO14 The app was easy to navigate.
Likert 1-5

PO15 The app helped me structure my AI interaction.
Likert 1-5

PO16 Using the canvas before the app improved the quality of my AI output.
Likert 1-5

Designer Agency

PO17 I felt in control of what AI produced. I directed the AI, not the other way around.
Likert 1-5

PO18 I could adjust or reject AI output when it did not match my intent.
Likert 1-5

Collaboration

PO19 The output we produced could be understood by someone not in the session.
Likert 1-5

PO20 Working as a group helped me think more critically about how to use AI.
Likert 1-5

PO21 The tools supported our group conversation.
Likert 1-5

Adoption

PO22 I would use this toolkit again for my own work.
Likert 1-5

PO23 I would recommend this to a fellow designer or colleague.
Likert 1-5

PO24 I could see a design team in a company using this to start implementing AI together.
Likert 1-5

PO25 What would need to change for you to actually bring this to your team or workplace?
Open answer

PO26 Did this session change what stops you from using AI? How?
Open answer

PO27 Did you discover any AI capabilities you did not know about before? Which?
Open answer

Artefact Ranking

PO28 Which artefact was most useful?
Multiple choice: Cards / Canvas / App

PO29 Which artefact would you remove, and why?
Multiple choice: Cards / Canvas / App / None — I would keep all three + open follow-up

Appendix Q. Results: Pre-Survey

Question	P1	P2	P3	P4	P5	P6	P7	Mean
Demographics								
Programme	IPD, last year of studying design (5 years)	Ipd and mostly study related experiences	Design for Interaction	Industrial Design	Masters integrated product design	Integrated product design. Package designer geweest bij Top Movers	Intergrated product design. Master of ide.	
AI use frequency	Weekly	Daily	Daily	Daily	Daily	Daily	Rarely (a few times total)	
Agent experience	Yes, once or twice	Yes, once or twice	Yes, once or twice	No	Yes, once or twice	No	No	
Individual AI Confidence (Likert 1–5)								
I can get useful results from AI tools.	4	4	5	4	5	4	3	**4.1**
I know what to ask AI to do.	4	4	4	3	4	4	1	**3.4**
I can write a good prompt.	3	4		4	4	3	2	**3.3**
I can judge whether AI output is good enough.	4	5	5	5	3	4	3	**4.1**
Team AI Confidence (Likert 1–5)								
My team uses AI in a consistent, structured way.	2	3	4	4	4	4	4	**3.6**
My team gets the same quality from AI as I do.	2	2	2	4	3	3	3	**2.7**
I would trust a prompt written by a teammate.	2	2	3	4	3	4	4	**3.1**
A shared library of tested prompts would be useful.	2	4	5	4	5	3	4	**3.9**
Open responses								
What do you currently use AI for?	Research / finding information;Writing and editing;Generating ideas;Creating visuals;Summarising content	Writing and editing;Summarising content;Structuring or organising information	Research / finding information;Writing and editing;Generating ideas;Creating visuals;Summarising content;Structuring or organising information	Writing and editing;Generating ideas;Summarising content;Structuring or organising information	Research / finding information;Writing and editing;Generating ideas;Creating visuals;Summarising content;Structuring or organising information	Writing and editing	Research / finding information;Writing and editing	
What stops you from using AI more?	I like it when things come out of my own hands or minds	I feel the creative process gains more from manual skilled labour, stumbling into surprises by accident. I feel ai might stop designers from full in depth exploration. It has become a cognitive pillar and decision maker, even if we're just talking about lay out or spacing.	The time it takes to figure out how to	Unreliable, unrepeatable	That sometimes it generates unrealistic, over the top ideas.	Authenticity	I don't know how what AI already can do. And i think my promt writing is not good enough the get the results i need at that moment in time.	
What would you want your team aligned on before using AI together?	I think people can use AI as much as they want as long as the outputs are good (quality control). I have a personal feeling that things are better when I do them myself, which is probably an idea I will be outrun by.	Our goal and how much agency we give ai	Visual style for creating images	Output	Which results we found useful, what is our backup story for the ai generated response and ethics depending on the use context.	The results you try to achieve	To what extend we can use it without legal problems. So we don't plagiarism sometimes.	
Physical cards to align my team would be helpful.	2	5	5	3	3	3	4	–

n=7. P1–P2: Session A1 (pilot, March 25). P3–P7: Sessions A2+A3 (March 26+30). P3 left 'write a good prompt' blank.

Appendix R. Results: Mini- survey

Question	P1	P2	P3	P4	P5	P6	P7	Pilot	A2+A3
MAIK Cards									
I understood the cards without explanation.	5	4	3	4	3	4	5	**4.5**	**3.8**
The cards helped me discover AI capabilities I did not know about. (pilot only)	1	1	–	–	–	–	–	**1.0**	**–**
The card categories made sense to me.	4	3	3	4	5	4	5	**3.5**	**4.2**
The cards helped our group align on how to use AI.	2	2	4	5	4	4	4	**2.0**	**4.2**
What confused you about the cards?	No confusion tbh they are clear		Some cards are very similar	Some might have overlap, like frame the challenge from synthesis and problem definition from analysis	A lot of cards with some text to read so it takes some time to get used to. Also took some time to understand the prompt was at the back	The start was a bit confusing. Because normally I would choose a method based on the step I am in and what would logically fit for the next step. Now I had to think ahead	I think right know i was mainly looking at the things I normally use. So I knew what they meant.		
MAIK Canvas									
The canvas helped me structure my approach to AI.	3	4	4	5	5	4	4	**3.5**	**4.4**
The canvas helped our group align on a shared AI approach.	4	4	5	5	5	4	5	**4.0**	**4.8**
The canvas fields made sense to me.	3	3	3	4	3	3	4	**3.0**	**3.4**
I feel more confident about using AI after filling in the canvas.	4	2	4	4	4	3	4	**3.0**	**3.8**
What confused you about the canvas?	I had to really think about how I formulate it what certain questions wanted from me	Wether I am developing a prompt or a skill	Some fields overlapped	To understand what is it at first or how it would influence my projects or prompts. Maybe introduction to it (in the app or smth) could be more explanatory and detailed)	The name 'canvas' because I was just making a prompt. Maybe 'prompt builder' or 'prompt playground' would make more sense to me. Then you are calling a prompt a 'skill'	Step 6 and 7. 6 felt a bit unnecessary because I thought AI would create this on its own based on the prompt I'm filling in. and 7 was confusing because we hadn't start a chat yet. But we learned afterwards that the goal and sources are already in the chat. But is this step then also necessary?	In the end I missed the group feedback on the skill. Yes it worked as we wanted, but why is this. Or could we get more out of the skill. One thing was, know we were missing problem definition. The ai had no problem with it, but confused me for a moment in the end we just followed the proces.		
MAIK App									
The app was easy to navigate.	3	4	4	5	4	5	4	**3.5**	**4.4**
The app helped me get a useful output.	3	2	4	4	4	5	4	**2.5**	**4.2**
File upload worked as expected (pilot) / I knew what to ask AI (A2+A3)	3	1	3	4	5	4	3	**2.0**	**3.8**
I would use this app for my own project.	4	2	4	5	5	4	4	**3.0**	**4.4**
What confused you about the app?	Sort of where to start if I were to start off without a card, how do I navigate the app	Could not edit prompts. it was unclear that it is a "social" app.	Project timeline would be nice to have	To understand at first if I'm interacting with the ai by just chatting or if it's strictly goes through prompting	In my opinion a bit too many nice-to-have features take take away from the core idea, but that could also be me using it for the first time. I can imagine having the ability to favorite could be interesting for power users, but not when you are first exploring the	- Daphne , over navigatie en naam verandering in app	I found it a bit confusing that you can select a different ai application but the answer does no chance. It only applies for the next question. But the app is easy to navigate. And if there is an option to adjust the names, you can easier find the correct team projects.		

Appendix S. Results: Post-Survey

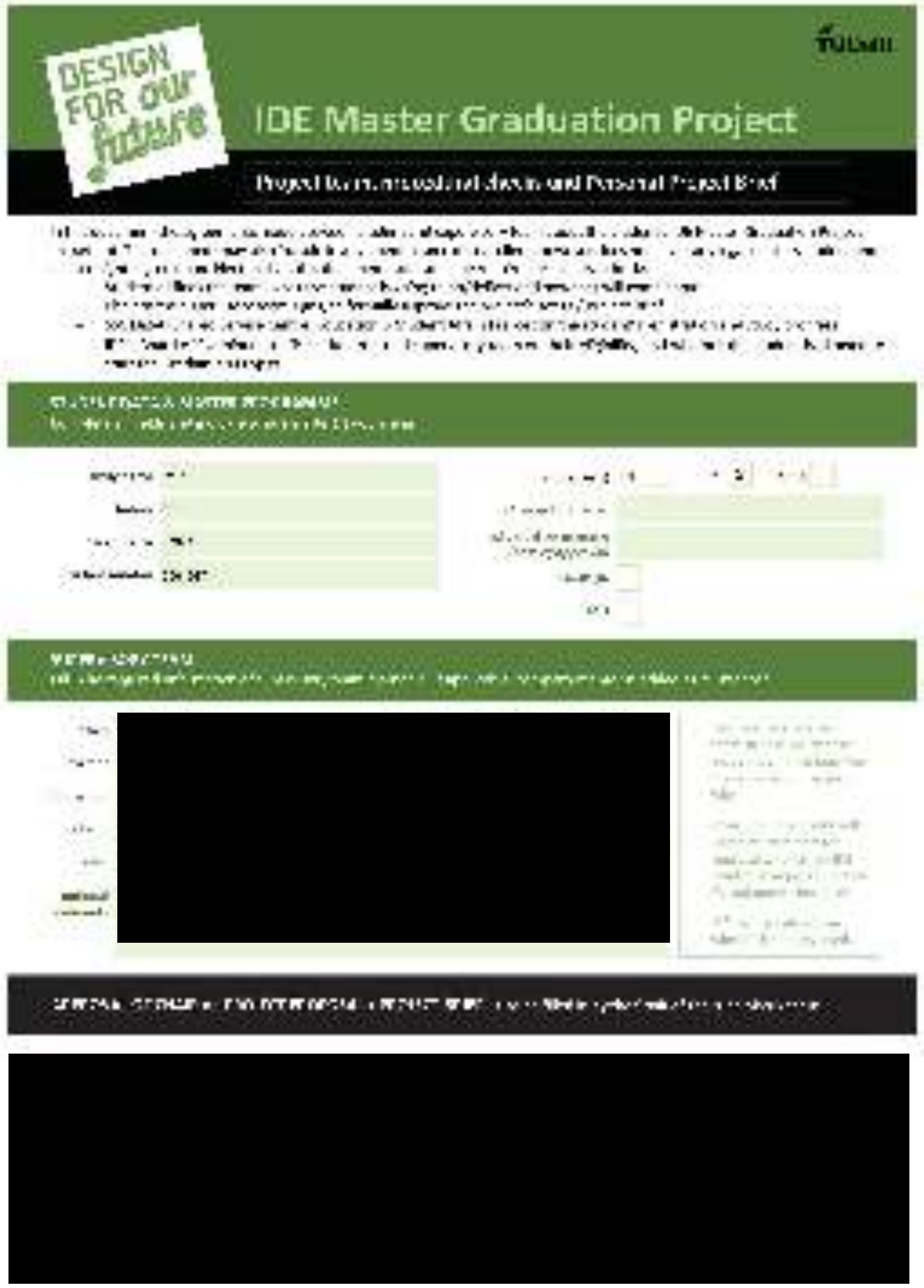
Question	P1	P2	P3	P4	P5	P6	P7	Post
Individual AI Confidence								
I can get useful results from AI tools.	3	4	4	5	5	5	4	**4.3**
I know what to ask AI to do.	3	4	4	5	5	4	4	**4.1**
I can write a good prompt.	3	4	5	5	5	4	5	**4.4**
I can judge whether AI output is good enough.	3	4	4	5	4	5	3	**4.0**
Team AI Confidence								
I would trust a prompt written by a teammate.	4	3	5	5	3	5	4	**4.1**
My team can use AI in a consistent, structured way.	4	3	5	5	5	5	4	**4.4**
I trust MAIK prompts more than prompts I would write myself.	5	2	3	4	5	5	3	**4.0**
Artefact-Specific								
I understood the cards without needing them explained.	3	4	4	4	5	5	4	**4.1**
The card categories made sense to me.	5	5	4	5	5	4	5	**4.7**
The canvas helped me prepare before using AI.	–	–	4	5	3	5	4	**4.0**
I would use the canvas before starting an AI project.	–	–	5	5	4	4	3	**4.1**
The app was easy to navigate.	4	2	5	3	5	5	5	**4.0**
The app helped me structure my AI interaction.	4	4	4	3	5	4	4	**4.1**
Using the canvas before the app improved my AI output quality.	–	–	5	4	4	3	4	**4.0**
Agency and Control								
I felt in control of what AI produced.	4	5	4	4	5	4	3	**4.1**
I could adjust or reject AI output when it did not match my intent.	–	–	3	5	5	3	4	**3.9**
The output could be understood by someone not in the session.	4	2	4	5	4	4	4	**4.1**
Group and Recommendation								
Working as a group helped me think more critically about AI.	4	4	5	5	4	2	2	**3.6**
The tools supported our group conversation.	5	4	5	5	5	4	4	**4.6**
I would use this toolkit again.	2	3	4	5	5	5	4	**4.0**
I would recommend this to a colleague.	2	3	5	5	5	5	5	**4.3**
I could see a company design team using this.	5	4	5	5	3	5	5	**4.6**
Most Useful / Would Remove								
Which artefact was	Cards	App	Canvas	Cards	App	Canvas	Canvas	

Appendix T. Results: Research B

Question	B1: YouTube Researcher	B2: Collins PowerPoint Builder
Part 1 — Dimension Scores (1-5)		
D1 Speed — How quickly did the skill produce output you could act on?	5	5
D2 Coverage — How many distinct sources / how well did slides cover content?	4	4
D3 Perspective Diversity / Visual Consistency	3	5
D4 Readability — How easily could you use/present without reformatting?	4	3
D5 Quote Specificity / Content Accuracy	5	4
D6 Evidence Balance / Layout Quality	3	5
D7 Accuracy / Brand Fidelity — Traceable to source / matches Collins standard?	5	5
D8 Reusability — Self-explanatory to someone not in the session?	5	5
D mean	4.25	4.5
Part 2 — Skill Evaluation (1-5)		
E1 Output was relevant to what I asked for / matched what I asked for	4	4
E2 Evidence credible and traceable / Professional enough to share	5	2
E3 Gave insights I would not have found quickly / Saved significant time	4	4
E4 Structure and format easy to work with / Structure and flow made sense	4	5
E5 Could use directly in Collins CXD project without significant rework	2	2
E6 Addresses at least one challenge in the Discover phase / bottleneck	4	4
E mean	3.83	3.5
Part 3 — Open Reflection		
O1 What was the most useful thing about the output?	Speed and credibility	Downselection, trust curating potential client. Focus on storytelling instead of the research part
O2 What was missing, incorrect, or frustrating?	Structure and formatting more appropriate to the use case — best practices guides — result	Barebone. Does not appear polished
O3 Which challenges does it address best / What would you fix before using?	—	Graphic design — cross-check with personal takeaways. Hitting on the emotional part of the story including the quotes
O4 Would you use this skill on a real Collins project? Why or why not?	Sources directly cited and credible	It was helpful to start this with a base. Distilling everything.

Research B. n=1 (Lily, Collins CXD designer). B1: March 14, 2026. B2: March 31, 2026. Both remote via Zoom.
D-scores: 1 (Poor) to 5 (Excellent). E-scores: 1 (Strongly disagree) to 5 (Strongly agree).
E5 is the lowest score in both sessions. B1: content accurate but format wrong. B2: format correct but visual design missing.

Appendix U. Graduation Brief



DESIGN FOR our future

IDE Master Graduation Project

[redacted]

[redacted]

DESIGN FOR our future

Personal Project Brief – IDE Master Graduation Project

Personal Project Brief – IDE Master Graduation Project

[redacted]

Appendix V. Original Brief

Collins Aerospace

An RTX Business

MSc Graduation Project:

Designing with AI: Shaping the Future Workflow of Aircraft Interior Design

Company: [illegible]

[illegible]

Website: https://www.collinsaerospace.com

[illegible]

[illegible]

Your mission includes:

- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]

Key questions to explore:

- How can AI tools enhance, rather than replace, the creative process in industrial design?
- [illegible]
- How can small design teams leverage AI workflows to accelerate innovation cycles and increase agility?

[illegible]

[illegible]

[redacted]