EDUCATION CO-PILOT

Prototype Evolution Report

From concept to functioning agent — design process, rationale and version history

24/7 Instructional Design · 24/7 Teach Platform

AI Education Agent Series · June 2026

Table of Contents

PART 1

Project Overview

Purpose, vision, and design approach

What This Document Is

This report documents the full design and development process for the Education Co-Pilot — an AI coaching agent for teachers and school administrators. It traces the journey from audience research through four prototype iterations to a deployment-ready Claude Project system prompt. It is intended as both a design record and a reference for future development.

The Core Design Problem

Most AI tools for education are either generic AI assistants (powerful but require significant prompting expertise) or narrow platforms (easy to use but inflexible). Neither builds the educator's long-term capability. The Education Co-Pilot is designed to occupy a third position: a conversational agent that helps educators solve real, immediate problems while simultaneously building their AI fluency. The key principle is teach by doing, not by lecturing.

> Design principle established at the start: The agent should help users understand where AI fits into their real work — not explain AI in general. Every interaction should deliver value first and embed learning in the process of delivering that value.

Two Audiences, One Agent

TEACHERS

- Classroom teachers: 1st year to veterans
- K-12, all subjects and grade levels
- Identity tied to craft and students
- Arrive with specific, immediate pains
- Need relief before they need education

ADMINISTRATORS

- Principals, APs, curriculum directors
- Instructional coaches, district leaders
- Operate at a systems level
- Arrive with strategic or compliance pressures
- Need institutional justification, not just task help

PART 2

Research Foundation

Audience personas — goals, constraints, concerns

Why Persona Research Came First

Before evaluating any conversation structure, the project required a detailed and honest portrait of each audience. Designing for 'educators' in general produces generic outputs. Designing for a burned-out elementary teacher with 28 essays to grade, or an AP who writes 14 observation reports per semester, produces something useful.

Teacher Persona Summary

Dimension	Key Findings
Goals	Survival (get through the week), craft (engage students), impact (reach hardest students)
Constraints	Time scarcity, classroom isolation, curriculum mandates, lack of PD support
Concerns	Am I being replaced? Is this cheating? Academic integrity with students. Equity of access.
Entry point	A specific pain this week — 'I have 28 essays to grade' — not a strategic goal
Activation energy	Very low required. If the first exchange delivers nothing useful, they don't return.

Administrator Persona Summary

Dimension	Key Findings
Goals	Accountability (outcomes, retention), personal (forward-thinking reputation, better decisions)
Constraints	Time scarcity, budget, district/state policy, uneven staff readiness, public accountability
Concerns	Data privacy, equity, teacher backlash, looking uninformed, initiative fatigue
Entry point	Strategic or systemic challenges — or a compliance directive from above
Institutional role	Often deciding for the whole school, not just themselves. This changes the stakes of every recommendation.

PART 3

Framework Evaluation

Testing the original 8-stage flow against each persona

The original conversation framework had eight stages: Welcome, Understand User, Identify Goals, Explore Workflow, Identify AI Opportunity, Recommendation, Responsible AI, Next Steps. Before building any prototype, each stage was evaluated against both personas.

Stage	vs. Teacher	vs. Administrator	Key gap
Welcome	Moderate	Strong	Three questions feels like intake; triggers teacher anxiety
Role & context	Moderate	Moderate	Missing grade/subject for teachers; no domain-narrowing for admins
Identify goals	Weak	Moderate	'AI fluency goal' framing doesn't match how either audience thinks
Explore workflow	Strong	Strong	Needs fallback for users without defined processes
AI opportunity	Good	Strong	Missing student-facing AI branch; missing 'deciding for school' branch
Recommendation	Moderate	Moderate	Language skews teacher; no admin-specific examples or framing
Responsible AI	Moderate	Strong	Missing academic integrity thread for teachers; leadership dimension for admins
Next steps	Moderate	Moderate	Too many options; not anchored to the conversation just had

> Critical insight from the evaluation: 'Identify your AI fluency goal' is an administrator-level question. Teachers don't think in terms of fluency goals — they think in terms of problems. This single observation drove the most important structural change in Prototype 1: replacing goal-identification with pain-point surfacing.

PART 4

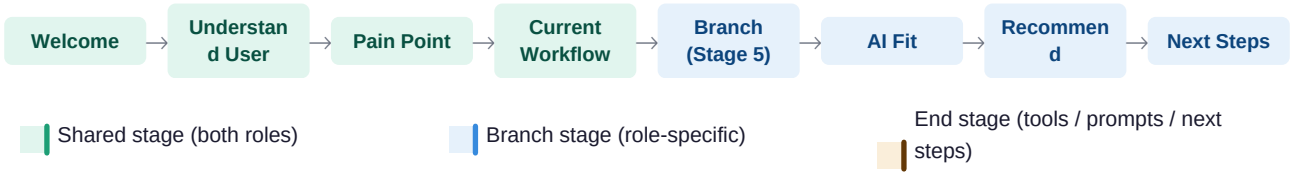
Prototype 1

Shared framework, first branch, flow bar navigation

What Problem This Prototype Solved

The original framework branched on role immediately — the first question was 'are you a teacher or administrator?' Prototype 1 delayed that branch to Stage 5 and replaced goal-identification with pain-point surfacing. The insight: the first answer the user gives about their heaviest task already signals their likely role implicitly. Forcing a declaration before any value is delivered reduces trust.

Prototype 1 Conversation Flow



Prototype 1 — Shared Framework + First Branch

WHAT IT DOES

- 8-stage flow: Welcome → Context → Pain Point → Workflow → Branch → AI Fit → Recommendation → Next Steps
- Detects role implicitly from first response (lesson planning = teacher, staff management = admin)
- Delays formal branching to Stage 5 — shared context established first
- Flow bar shows shared (green) vs branch (blue) stages visually
- Branch card at Stage 5 shows both teacher and administrator pathways with tabs
- Role-specific questions, recommendations, and next steps after the branch
- Choice buttons for routing at every stage
- 'Build it in chat' sendPrompt button at the end

WHAT IT DOESN'T DO / GAPS IDENTIFIED

- Choice buttons only — no free-text input for users with specific or unusual situations
- Pain point options at Stage 1 showed only administrator-facing choices (data, planning, staff) — teachers had no relevant entry points
- Prompting guidance was generic — same three patterns regardless of what the user selected
- No tool recommendations beyond Claude, ChatGPT, Gamma, and Perplexity
- No AI literacy instruction — the agent helped but did not teach
- Role detection relied entirely on first button click — free-text entry would break routing

CHANGES MADE FROM PREVIOUS VERSION

- FIRST version — no previous version to compare against

WHAT LED TO THE CHANGES

The framework evaluation revealed that the original 8-stage flow was too front-loaded with intake. The critical structural change was replacing 'identify your AI fluency goal' with a pain-point question. Branching at Stage 5 rather than Stage 1 was driven by the insight that shared context makes role-specific recommendations more credible.

PART 5

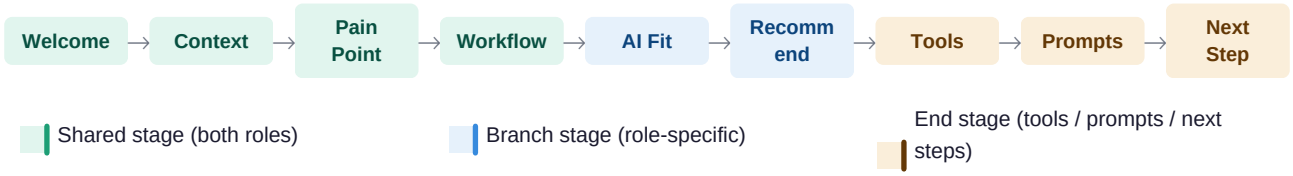
Prototype 2

Free text input, tool recommendations, prompting stage

What Problem This Prototype Solved

Three gaps in Prototype 1 drove the changes: users with complex or unusual situations could not express them through buttons alone; teacher pain point options were absent from Stage 1; and tool recommendations were too narrow to be genuinely useful. A dedicated prompting stage was added because a tool recommendation without prompting guidance is only half-useful.

Prototype 2 Conversation Flow



Prototype 2 — Free Text + Tool Library + Prompting Stage

WHAT IT DOES

- Free-text textarea added alongside choice buttons on every stage
- Lightweight intent detection routes free-text responses to correct role/pain path
- Eight pain-specific tool sets covering lesson, feedback, comms, staff, data, planning
- Tool logic organised by job type: conversation/drafting, visual design, research, workspace
- 'When to use two tools together' section for each task — the most important tool guidance
- Three prompt pattern templates: Role+Task, Constraint First, Iterative Build
- Five teacher pain point options added to Stage 1 (lesson, feedback, diff, comms, materials)
- Administrator pain points retained (staff, data, planning)

WHAT IT DOESN'T DO / GAPS IDENTIFIED

- AI literacy not present — the agent helped and recommended tools but never taught
- Even with teacher options added, the context stage still showed admin-skewed framing
- Prompt templates were generic — the same three patterns regardless of the user's pain point
- Role was still determined primarily from Stage 1 button selection — role declaration came after task selection, not before
- No expertise calibration — new users and experienced users received identical guidance depth

CHANGES MADE FROM PREVIOUS VERSION

- Added free-text input with intent detection at every stage
- Added dedicated tool recommendation stage with 8 pain-specific sets
- Added two-tool combination guidance for each task
- Added dedicated prompting stage with 3 patterns
- Added 5 teacher pain point options to Stage 1

WHAT LED TO THE CHANGES

User feedback on Prototype 1 identified three problems: buttons couldn't capture nuanced situations; teachers had no relevant entry points at Stage 1; and 'here are four tools' without prompting guidance left users no better equipped than before. The tool library was expanded after research identified that Fathom, MagicSchool, Diffit, Gradescope, and education-specific platforms were more useful for educators than general AI tools alone.

PART 6

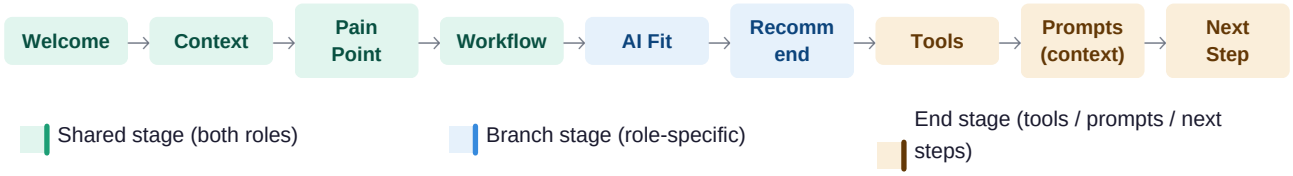
Prototype 3

AI literacy integration, fixed teacher entry, context-aware prompts

What Problem This Prototype Solved

Three specific failures in Prototype 2 were identified: teachers were not learning anything — they were getting outputs but not building capability; the context question still skewed toward administrators; and the prompting stage showed the same generic patterns regardless of what the user selected. All three were structural, not cosmetic.

Prototype 3 Conversation Flow



Prototype 3 — AI Literacy + Corrected Teacher Entry + Context-Aware Prompts

WHAT IT DOES

- AI literacy moments embedded as purple callout boxes at every stage — triggered by stage + pain point
- Literacy framed as 'here is why this works' or 'here is what to watch for' — never as a lesson
- Eight distinct prompt sets built — one per pain point — replacing the three generic patterns
- Each prompt set structured as a progression: First Prompt → Refine It → Push Further
- Every prompt includes a 'why this works' note that doubles as AI literacy instruction
- Teacher-specific context questions built for all five teacher pain points
- Literacy moments differ by expertise level implied — foundational for new users, strategic for experienced
- All eight pain point pathways now have complete, role-appropriate context questions

WHAT IT DOESN'T DO / GAPS IDENTIFIED

- Role declaration still happened implicitly — the first choice button determined role without asking
- Multi-select not possible — users could only select one pain area even if they had several
- Expertise level was inferred, not asked — calibration was approximate and could be wrong
- No explicit role question meant a teacher who selected an admin option by accident would be misrouted
- Context and pain point were determined in the same step — they couldn't inform each other

CHANGES MADE FROM PREVIOUS VERSION

- Added AI literacy callout boxes at every stage (8 total per path, 17 distinct literacy moments)
- Built 8 pain-specific prompt sequences replacing 3 generic patterns
- Added 'why this works' notes to every prompt
- Built teacher-specific context questions for all 5 teacher pain points
- Corrected the Stage 1 imbalance — equal teacher and administrator entry points

WHAT LED TO THE CHANGES

The core insight was that literacy delivered in context sticks; literacy delivered as a module before the task is forgotten. The prompt template failure was traced to the fact that 'how to prompt AI for differentiation' and 'how to prompt AI for a staff conversation' are completely different skills. Generic patterns were not transferable. The teacher context failure was traced to Stage 1 still having more administrator-recognisable options than teacher-recognisable ones.

PART 7

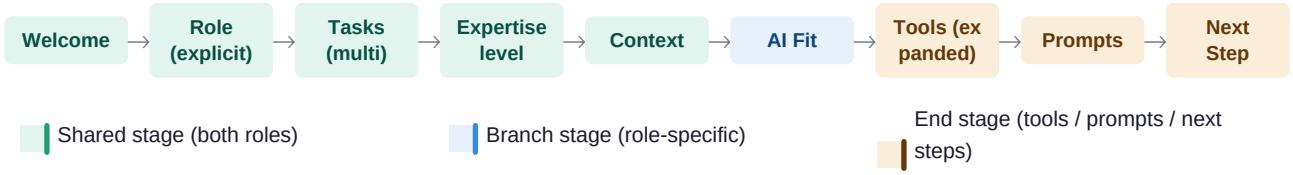
Prototype 4 (Current)

Role-first flow, multi-select, expertise calibration, expanded tools

What Problem This Prototype Solved

Three structural problems remained in Prototype 3. First, role was still determined by task selection — a teacher who selected an admin option would be misrouted with no way to correct. Second, users with multiple pain areas could only select one — a real educator's situation is rarely a single problem. Third, new and experienced users were receiving identical guidance depth, which over-explained to experienced users and under-supported new ones.

Prototype 4 Conversation Flow



Prototype 4 (Current) — Role-First + Multi-Select + Expertise Calibration + Full Tool Library

WHAT IT DOES

- Role declared explicitly at Stage 2 before any task selection — eliminates misrouting
- Task selection is now multi-select — users choose all relevant areas, not just one primary
- Primary task used for deep workflow focus; secondary tasks mapped to additional tool recommendations
- Dedicated expertise calibration stage: New / Some Experience / Confident User
- Tool recommendations, prompt framing, and AI literacy depth all adapt to expertise level
- New users receive one clear recommendation and explicit 'what to do with a first draft' guidance
- Confident users receive peer-level engagement, tool combination focus, and template-building emphasis
- Expanded tool library: 35+ tools across 8 categories including meeting capture (Fathom, Otter, Granola, Jamie, Zoom AI, Fireflies), ed-specific (MagicSchool, Diffit, Gradescope, Brisk, Khanmigo, SchoolAI, Curipod), research (NotebookLM, Consensus, DeepL)
- Meeting capture workflow explicitly recommended for administrators with post-meeting processing prompts
- 14 distinct task categories (8 teacher, 8 administrator) with unique tool sets

WHAT IT DOESN'T DO / GAPS IDENTIFIED

- The prototype currently shows the primary task workflow in depth — secondary tasks receive tool references but not full workflow support
- No persistent memory across conversations — each session starts fresh
- No ability to upload a document for AI to work with directly within the prototype
- The responsible AI guidance is embedded in prompts but not yet a dedicated conversation stage
- No branching for 'deciding for the whole school' vs 'personal use' in the administrator path

CHANGES MADE FROM PREVIOUS VERSION

- Restructured flow: Welcome → Role (explicit) → Tasks (multi-select) → Expertise → Context → AI Fit → Tools → Prompts → Next Step
- Added explicit role selection as Stage 2 before task selection
- Converted task selection from single-select to multi-select with chip confirmation
- Added dedicated expertise calibration stage with three levels
- Tool intro text adapts to expertise: one recommendation for new users, full set for experienced
- Prompt framing adapts: 'fill in the brackets exactly' for new users vs 'push further' for experienced
- Expanded tool library from 8 tools to 35+ tools across 8 job-type categories
- Added full meeting capture section with 6 tools, platform-specific guidance, and post-meeting workflow
- Added education-specific platforms section with 9 tools
- Added research and synthesis tools: NotebookLM, Consensus, DeepL

WHAT LED TO THE CHANGES

The role misrouting problem was identified when testing with an administrator who selected 'lesson planning' out of curiosity and was routed to a teacher pathway. The multi-select change came from recognising that real educators never have one problem — a teacher might need help with both feedback AND differentiation AND parent comms. The expertise gap was identified when testing showed the same explanation being offered to a 20-year veteran and a first-year teacher. The tool expansion was driven by research revealing that tools like Fathom, MagicSchool, Diffit, and Khanmigo are significantly more useful for their specific jobs than general AI assistants — and that administrators specifically were underserved without meeting capture recommendations.

PART 8

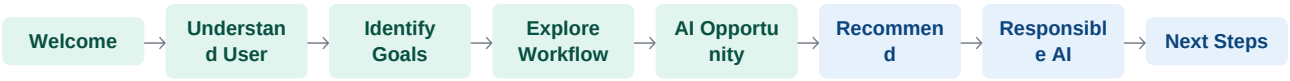
Flow Diagrams

Visual comparison of all four prototype conversation flows

How the Flow Evolved

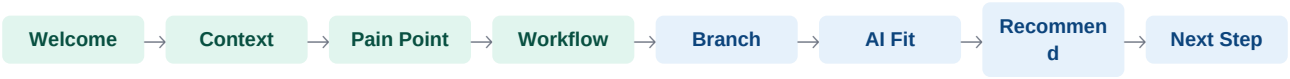
The most visible change across versions is where branching happens and what is added before and after it. The flow moved from a gate-first structure (branch immediately on role) to a value-first structure (establish context, then branch, then deliver). New stages were added sequentially as gaps were identified — never speculatively.

Original Framework (Pre-Prototype)



8 stages. Role declared at Stage 1. Fluency goal required at Stage 3. No tool recommendations. No prompting guidance.

Prototype 1 — Shared Framework + First Branch



Pain point replaces fluency goal. Branching delayed to Stage 5. Flow bar navigation added. Role implicit from first answer.

Prototype 2 — Free Text + Tools + Prompts

Free text added. Tools stage added (8 pain sets). Prompting stage added. Teacher pain points added to Stage 1.

Prototype 3 — AI Literacy + Context-Aware Prompts

AI literacy callouts at every stage. 8 pain-specific prompt sets. Teacher context questions fixed. Literacy by stage and pain.

Prototype 4 — Current Version

Role declared before task. Multi-select tasks. Expertise calibration stage. 35+ tools, meeting capture, ed-specific platforms.

PART 9

Design Principles

What stayed constant across all four versions

Eight design principles emerged from the research and were maintained through every prototype iteration. These are not aspirational values — they are structural constraints that shaped every decision. When a proposed change violated one of these principles, it was rejected or redesigned.

1 **Coach First, Teach Second**

The agent must understand the user's work before explaining AI concepts. Value delivery precedes education. If the first exchange does not produce something useful, learning cannot happen.

2 **Pain Point Over Fluency Goal**

'What's frustrating you?' consistently outperforms 'What do you want to learn?' for both audiences. Meeting users at their problem is the only reliable entry point.

3 **Delayed Branching**

Shared context before role-specific pathways. The first answer already signals role implicitly — forcing a declaration before value is delivered reduces trust and engagement.

4 **Value Before Volume**

Deliver something useful before asking more questions. The intake cannot feel longer than the value. Every additional question is a cost that must be justified by what it enables.

5 **AI Literacy in Context**

Embed learning in the workflow, not before it. Literacy delivered as a module is forgotten. Literacy delivered as 'here's why this worked' at the moment of doing is retained.

6 **Preserve Professional Identity**

AI is always the assistant; the educator is always the expert. Every recommendation must visibly preserve the educator's role. Any output that appears to replace professional judgment will be rejected.

7 **Specificity Over Generality**

Every recommendation must be anchored to something the user actually said. 'AI can help with this' is not useful. 'Here is what to try tonight, in this tool, with this prompt' is.

8 **The 80/20 Principle**

AI produces approximately 80% of what is needed. The educator's professional knowledge provides the 20% that makes it usable — student specifics, institutional context, final judgment. Name this explicitly.

PART 10

From Prototype to Agent

System prompt, knowledge base, and test personas

What Was Built After the Prototype

The four prototype iterations produced the design logic, the conversation architecture, and the tool and prompt libraries. These were then formalised into three deployable artefacts for the Claude Project implementation.

System Prompt

File: education_copilot_system_prompt.md

- Identity and purpose — teach by doing, not by lecturing
- Two audience profiles written in full (teacher and administrator)
- 8-stage conversation architecture with coaching language
- Complete tool recommendations with when/when-not-to guidance
- Responsible AI guidance integrated throughout (not appended)
- Three AI literacy teaching patterns: 'why this works', 'what to watch for', 'what this means broadly'
- Tone and formatting guidance — warm, specific, direct
- Sample exchanges for four user types to calibrate response style

Knowledge Base — Category 2: Tool Reference Library

File: tool_reference_library.md

- 35+ tools across 8 categories with full descriptions
- Job-type framework for matching tool to task
- Meeting capture section with 6 tools and platform guidance
- Education-specific platforms section with 9 tools
- Research and synthesis tools including NotebookLM
- 8 two-tool combination workflows with rationale

Test Personas — Administrator Roles

File: administrator_test_personas.md

- 5 detailed administrator personas across all role types
- Each persona includes: profile, technology expertise, pressures, fears, goals
- Each persona includes: an opening message, 3 follow-up prompts, success criteria
- Roles covered: Principal (Title I), AP (multilingual school), Curriculum Director, Instructional Coach, CAO
- Cross-persona testing notes for evaluating role recognition and expertise calibration

Final principle: Every conversation should end with the educator being more capable than when they started — not just with a better output, but with a better understanding of how to get there themselves. That is the difference between a tool that creates dependency and a coach that builds capacity.