

AI CAMP BERLIN · 24 APRIL 2026

# 17.8x your UX testing process

**AI Camp Berlin · 24 April 2026 · PoDojo with  
guest Ayse Delikaya, mobile.de**

 **Bring a laptop if you want to follow along.**

Scan to install before we start



Installation Guide → [Claude Desktop](#) + [podojo CLI](#)

# Coaching for your team. Tools for your agents.

So people lead. Agents assist.



**Hannah Carlotta Kelbel**

Innovation & Design Thinking ·  
research + discovery



**Stefan Haas**

Co-Founder · product flow



**Marion Boeing**

UX Design & Research · user  
perspective



**Jochen Ade**

Co-Founder, CTO · tools +  
building blocks

# UX testing and research is hard.

Why we're spending 45 minutes on this together.



## Setup

Design the test (Hypothesis, what to test, study guide) **The real pain:** recruiting the right customers and coordinating schedules.



## Running the test

Talking to many people, one conversation at a time. **Some of it can be automated.**



## Synthesis & review

Hours of video, spreadsheets, miro or post-its – before any decision gets made.

**That loop is what we help shrink. Two live cases + one live demo — starting now.**

**"We would have an idea in the morning and in the afternoon we're already running a pretty full scale experiment... and by the evening we can go through all of it together."**

**Grant Lee — Co-Founder & CEO, Gamma**

Lenny's Podcast · November 2025



# What you'll see

45 minutes – two live cases, one live demo.

<b>1</b>	<b>Case mobile.de — 6.4 h → 30 min • Ayse</b>
<b>2</b>	<b>What's an agent • Why Claude Code • Jochen</b>
<b>3</b>	<b>Live demo — plan → test → analyze • Stefan</b>
<b>4</b>	<b>Bolt: 100+ interviews a week, built by a researcher • Jochen</b>
<b>5</b>	<b>Q &amp; A • All</b>

*Questions on Post-its – answered at the end.*

# Agentic User Research

Notes from our collaboration with Bolt

**Jochen Ade**

podojo · AI Camp Berlin

Chapter 01

# The basics

Agents, Claude Code, and how they reach your stack

---

## What is an agent? What is **agentic**?

An agent is a model running in a loop, using tools to achieve a goal.



**Andrew Ng**  
Founder, DeepLearning.AI

### A SPECTRUM, NOT A BINARY

Rather than arguing whether something **is** an agent, Ng asks **how agentic** it is — a gradient from a single prompt to systems that plan and act on their own.

**Reflection** — the model critiques its own output

**Tool use** — calling external functions and APIs

**Planning** — breaking a task into steps

**Multi-agent** — specialised agents collaborating

# Why Claude Code?

## STATE OF THE ART

Coding is where **agents** excelled first.

## IT'S NOT JUST A CODING TOOL

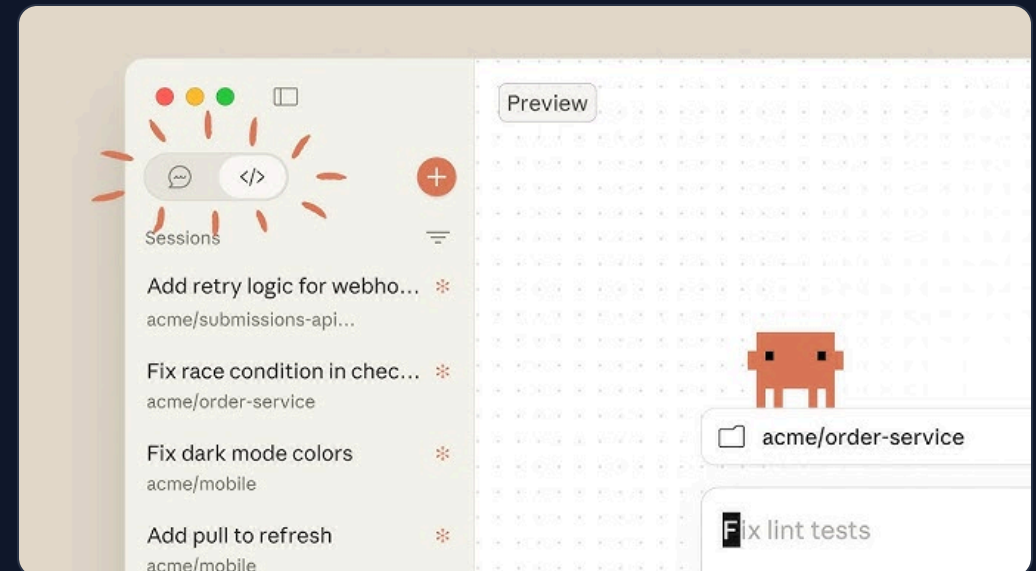
Think of it as "**Claude Agent**".

## WHAT ABOUT CLAUDE COWORK?

Probably where it's heading — **not quite ready** yet.

## HONORABLE MENTION

**pi (OpenClaw)** — the European take.



Claude Code in Claude Desktop

## How the agent uses the **podojo** tools

AGENT

### Claude Code

Plans, decides, acts.

CLI

Command-Line Interface — the agent runs podojo commands like a researcher would.

MCP

Model Context Protocol — a standard "plug" so any agent can call our tools.

TOOLS

### podojo

Projects, usertests, transcripts,  
videos.

Chapter 02

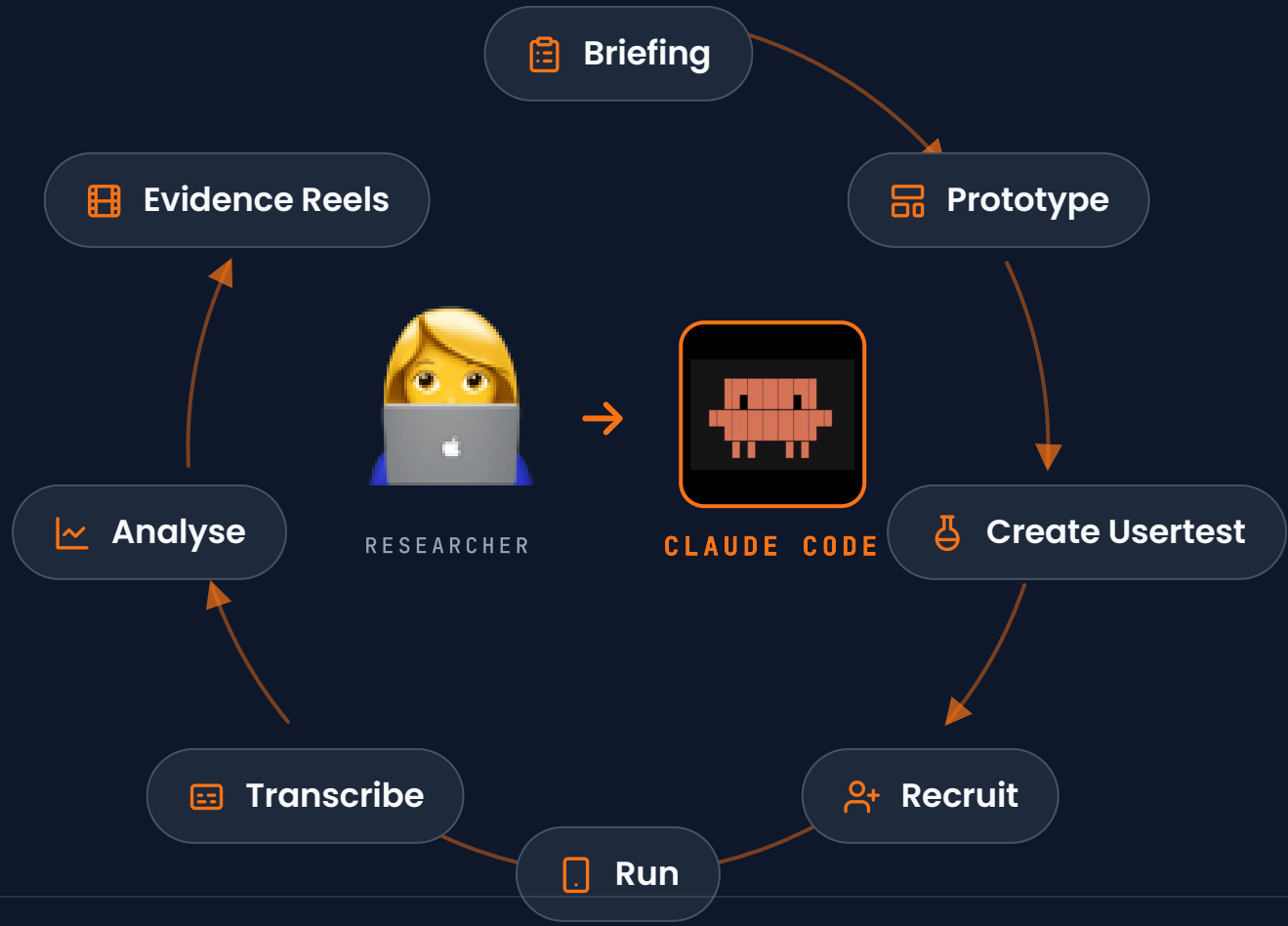
# The flow

Agentic User Research Pilot @ Bolt

---

# From briefing to insight

Agents absorb the **busywork** — researchers do the **actual research**.



## What we've done with Bolt

📅 Started Feb-26 · 📋 3 studies · 👥 ~50 participants (Africa, Europe)

### AGENTIC WORKFLOW

Claude Code drives the whole loop — **no jumping between tools**, no manual handoffs. Its **analytical ability is genuinely amazing**.

### OUR OWN TOOLS, ON PURPOSE

Most testing platforms keep you inside **their webapp** — no API. **podojo is API-first**. Plus: tests run in the **mobile browser** — nothing to install.

### COACHING THE RESEARCHERS

A few sessions in, Bolt's researchers **owned the workflow** — and enjoyed it.

### ACCURACY HOLDS UP

We benchmarked the agentic flow against a fully manual one. **Comparable findings, a fraction of the effort**.

Chapter 02 · Walkthrough

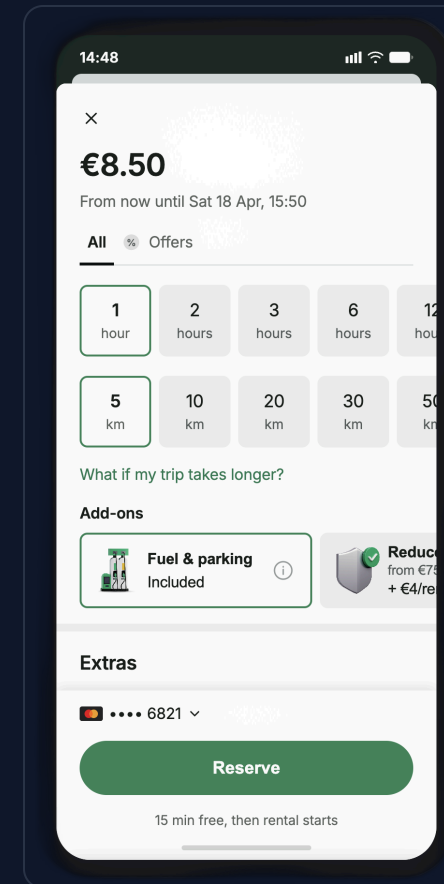
# Steps in detail

How the agentic flow plays out, step by step

---

# Prototype

- Code prototypes over mockups
- **Figma Make**: fast start, has weaknesses
- **Claude Code**: the direction of travel
- **Figma MCP**: not perfect but helpful



Prototype created with Claude Cowork

# Create Usertest

- **podojo Usertesting**: most market tools force you through their webapp — we needed something agentic
- **Testing on Mobile**: runs in the browser — no app install required
- **Brief → test**: research brief flows straight into test setup

The screenshot shows a user test configuration page for 'Bolt Drive new design idea'. At the top, there is a 'Back to User Tests' link and a 'Delete User Test' button. The project name is 'Bolt Drive new design idea' with a 'Live' status indicator. Below the project name, the project ID 'pricing-simplification-bolt-drive' is shown. There are three buttons: 'Preview URL', 'Live URL', and 'Prototype URL'. The main content area is divided into sections for 'Welcome Text', 'Step 1: What to expect', 'Step 2: How to respond', 'Step 3: Incentive reminder', and 'Step 4: Where do you currently live?'. Each step has a 'Screen - Intro' label and a brief description of the task or instruction.

← Back to User Tests

**Bolt Drive new design idea** Live Delete User Test

Project: pricing-simplification-bolt-drive

[Preview URL](#) [Live URL](#) [Prototype URL](#)

**Welcome Text**  
Welcome! Thank you for taking the time to participate in this study. It will take around 15–20 minutes to complete.

**Step 1** Screen — Intro  
**What to expect**  
You will be asked a few questions and then you will see one design idea. There are no right or wrong answers — we just want your honest feedback.

**Step 2** Screen — Intro  
**How to respond**  
During this session you can respond to questions by speaking out loud. Think of it as leaving a voice message.

**Step 3** Screen — Intro  
**Incentive reminder**  
Please keep in mind that you will only receive the incentive if you answer all the questions and do all the activities.

**Step 4** Screen — Question  
**Where do you currently live?**  
Tell us the country and the city of your residence.

# Recruit

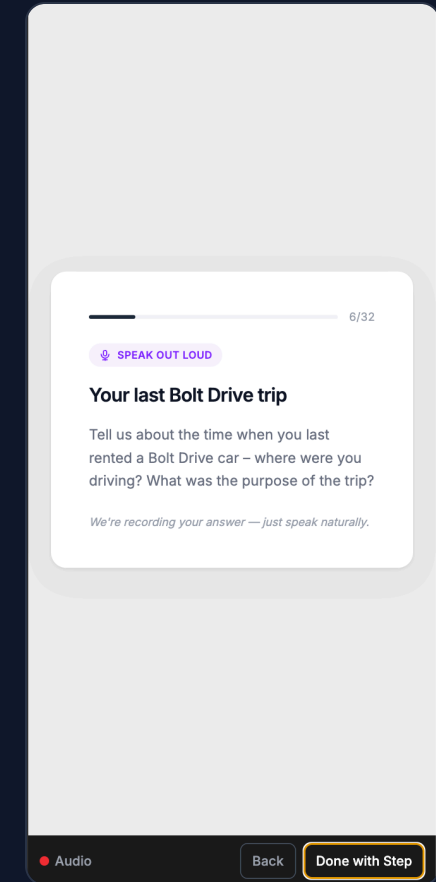
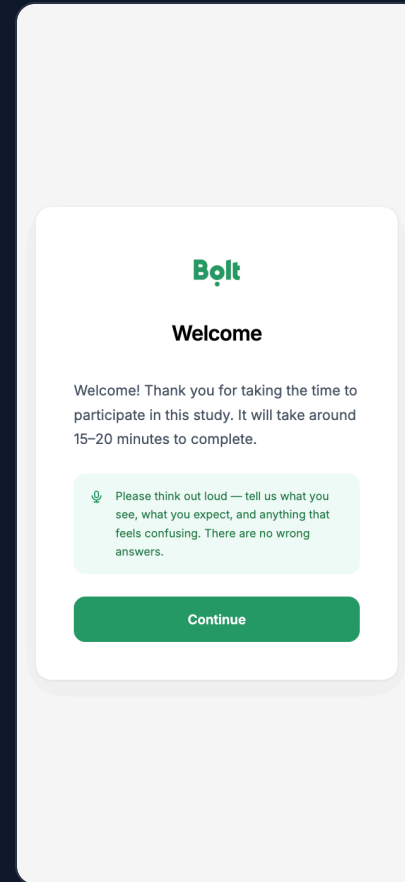
- Research ops built their own tool, with our help
- Large pool of participants
- Easy to filter profiles and send invites

The screenshot displays the 'CYC Process Tool' interface. On the left is a sidebar with navigation options: Home, Participants, Invites: Database (selected), Import New Users, Survey Responses, View Bookings, Check Attendance, and Settings. The main area is titled 'Filters' and includes sections for Text Filters (Country, Age Group, Gender, Booking Status, Primary User Type), Date Filters, and Order Filters (Sliders). A green button labeled 'Send Direct Messages (New Flow)' is visible. Below the filters, there are tabs for 'All Responses', 'Analytics', and 'Individual Response'. The 'All Survey Responses' section is active, showing a table with columns: Infobip ID, Booking Status, Last Contact Date, Last Booking Date, Created Date, Age Group, Gender, Primary User Type, Which country do you live in?, Driver Orders (Last 30 Days), and Courier Orders (Last 30 Days). The table contains four rows of data.

	Infobip ID	Booking Status	Last Contact Date	Last Booking Date	Created Date	Age Group	Gender	Primary User Type	Which country do you live in?	Driver Orders (Last 30 Days)	Courier Orders (Last 30 Days)
0	6169	Booked		03 Feb 2026	27 Oct 2025	25-34	Male	Courier	Poland		0
1	6065	Booked		20 Jan 2026	04 Dec 2025	35-44	None	Rider	Romania		0
2	6203	Booked	25 Feb 2026	25 Feb 2026	04 Dec 2025	18-24	Male	Driver	Romania		793
3	5905	Booked	05 Dec 2025	19 Jan 2026	04 Dec 2025	45-54	Male	Driver	Latvia		397

# Run the test

- Invite links via WhatsApp, SMS, or email
- Unmoderated and mobile-first
- ~15-minute tests



# Transcribe

- Videos transcribed automatically
- Captures audio and visual data
- Solid foundation for analyzing what people say and what they do

🖥️ 03:03  
The view changes to a package selection screen. It shows options for duration (1 hour, 2 hours, etc., up to days) and distance (5 km, 10 km, etc.). Add-on options for "Fuel & parking" and "Reduced liability" are visible below.

👤 **Speaker A** 03:03 - 03:35  
So I need one, one day and 250 kilometers, which is 95 euros. Fuel and parking included. I'm just going to see how much it costs without. It's 57.50. Okay. But to be honest, for 250 kilometers, considering the gas prices, I'll probably include the gas and the parking. And now I'm going to reserve. Okay. That worked out well.

👁️ 03:14  
User taps the "Fuel & parking" add-on, which is initially selected. The price changes from €95 to €57.50, and the add-on now shows "+ €37.50".  
User taps the "Fuel & parking" add-on again. The price changes back to €95, and the add-on is selected again.  
User taps the "Reserve" button.

🖥️ 03:32  
A success screen appears with an illustration and the text "Success. Your rental will start shortly".

👤 **Speaker A** 03:39 - 03:52  
I would say about, I mean, my perception of the packages are that they're always cheapest. So I would say 85 to 90%.

🖥️ 03:55  
A screenshot shows the "Fuel & parking Included" add-on. The question asks if the user noticed it and to explain their understanding.

👤 **Speaker A** 03:57 - 04:29  
Uh, yes, I mean, I addressed this as well when I was on the screen. So, um, I can exclude fuel and parking to lower the price. And, um, then that means that I will have to fuel the car on my own costs and also pay for the parking. But usually, if I go to a different city in which Bolt doesn't operate in, I'll probably have to pay for the parking anyway, so, yeah, the main add-on would be the fuel if I understand this correctly.

# Analyse

The researcher's time lands where it matters: interpretation.

## FILTER

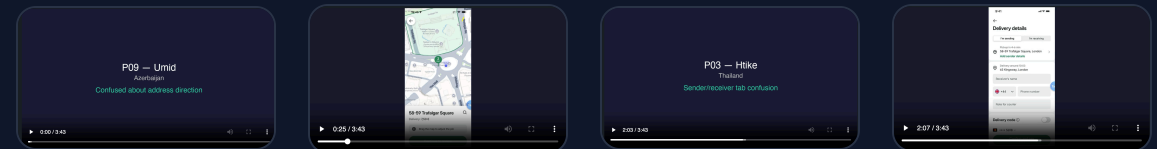
A skill flags **unreliable participants** and unusable feedback — signal in, noise out before analysis begins.

## REPORT

Agent-generated summary of findings: themes, quotes, ranked insights, recommendations.

## Evidence Reels

- One reel per finding, auto-generated
- Claims linked to exact clips
- Quick validation of findings



## How good are the results?

We benchmarked the agent against a researcher-written report on the same raw interviews.

### 01 · BASELINE

A **senior researcher's report** — locked in as the ground truth.

### 02 · AGENT RUN

The **same raw interviews**, processed end-to-end by the agent.

### 03 · BLIND REVIEW

Where **expert and agent disagreed**, a third researcher scored each finding.

### ACCURACY VS. THE HUMAN BASELINE

Human expert



baseline

Agent alone



close, but **below expert**

Agent + showreel check



**matches** fully manual

Based on a small dataset so far — directional, not scientific.

# Run the test

**The knitting-app test we shipped today is live.**

Scan to take it yourself. You get the session recording back when we're done.





# Looking for 5 more design partners.

We help you optimise your product flow with agentic AI and tools.

DISCOVERY CALL · 30 MIN



**Just schedule a discovery call or drop us an email**

[podojo.com/contact](https://podojo.com/contact) - [stefan@podojo.com](mailto:stefan@podojo.com)

*Thank you.*