ANTHROP\C

Appendix to “What 81,000 people want from AI”

March 2026

Corrections

Mar 19, 2026: Updated to reflect that incomplete interviews were not included in sentiment analysis.

Methods

For a full description of the Anthropic Interviewer methodology, see here. What follows is a brief overview of how we collected and analyzed data for this study.

What we asked. Our interview had four core questions: (1) What's the last thing you used an AI chatbot for? (2) If you could wave a magic wand, what would AI do for you? (3) Has AI ever taken a step towards that vision for you? (4) Are there ways AI might be developed that would be contrary to your vision or what you value? Anthropic Interviewer followed up on each, probing for the underlying values and experiences behind people's answers.

Data collection. We received 112,846 interviews over the course of one week in December 2025, of which 80,508 met our quality threshold—we filtered out conversations that were spammy, unserious, or extremely minimal (e.g. a very short initial response, then disengagement). We did not filter out people who engaged earnestly and then dropped out partway. Respondents came from 159 countries and wrote in 70 languages.

Representativeness. We validated the representativeness of our Claude.ai user base by usage patterns, region, and tier. The first question we asked people, as a warm-up question, was what they last used an AI chatbot for. We found that this matched our Anthropic Economic Index and other similar prior research well. For example, software development and programming accounted for 34% of interviewees' last usage, comparable to the 36% of Claude.ai conversations classified under computer and mathematical occupations in the September 2025 Anthropic Economic Index report. Learning and education (10%) and business strategy and operations (9%) fell within the ranges reported across the Economic Index and other prior analyses, which placed education at 7–12% and business strategy at ~6%.

We also found that the distribution of interviewees closely matched our weekly active user distribution for that week by Claude tier (Free/Pro/Max) and region (APAC, EMEA, etc).

How we analyzed responses. Interview transcripts were processed through a suite of hand-validated Claude-powered classifiers, each designed to extract a specific dimension of a respondent's answers. Each interview was classified across multiple dimensions,

including: visions (how they answered the question about what they'd want AI to do if they had a magic wand), experiences (how they answered the question about whether AI had taken a step towards their vision) and concerns (how they answered the question about how AI might be developed contrary to their values or vision). We also classified job domain (inferred from self-description, e.g. legal or healthcare work) and job structure (what kind of employment setup they had, e.g. employee or entrepreneur). 72% of respondents stated some job information.

Concerns were multi-label, i.e. a single interview could receive multiple codes, since interviewees tended to state multiple concerns. Some classifiers (e.g. the "concerns" classifier) were designed first by analyzing answers with a bottom-up clustering algorithm—categories were derived from clusters that surfaced during initial analysis—to ensure comprehensive yet cleanly separated categories. Each classifier was validated for at least 90% agreement with a human on 25 labels. Transcripts where interviewees did not reach the relevant part of the interview were excluded from the relevant analysis (e.g. those who did not answer the concerns question were excluded from the concerns classifier, the sentiment analysis and the light and shade analysis).

Example of classifier prompt for extracting sentiment score:

```
None
Rate this person's overall sentiment toward AI on a 1-7 scale based on the FULL
interview.

1–Extremely negative.
2–Negative: Concerns clearly dominate.
3–Lean negative: Overall more negative than positive, e.g. concerns weigh
heavier.
4–Mixed/neutral: Genuinely in the middle–real benefits AND real concerns,
roughly balanced, hard to find that this person leans towards one side or
another, or they are simply neutral.
5–Lean positive: Overall more positive than negative.
6–Positive: Clearly enthusiastic. Concerns are minor, hypothetical, or
mentioned only when directly asked.
7–Extremely positive.

Guidance:
- Note and weight what they emphasize and return to.
```

```
- Someone who lists benefits then pivots to a long passionate concern monologue
is probably a 3-4, not a 5-6.
- Most people have mixed feelings—don't default to extremes. Use the full 1-7
range.

<interview>
{TRANSCRIPT}
</interview>

Return ONLY a single integer (1-7) in <answer> tags.
```

Example of classifier prompt for classifying job context:

```
None
Classify this person's professional context along TWO dimensions based on what
they say about themselves in the interview.

Respond with exactly two labels, comma-separated, no spaces:
employment_structure,professional_domain

Use the MOST SPECIFIC label you can confidently support. When you have LIMITED
information, use BROADER categories (e.g. employee, independent) or
UNCLEAR/OTHER. When you have CLEAR information, use the more SPECIFIC
categories.

DIMENSION 1 — EMPLOYMENT STRUCTURE (how is your work organized?):

BROAD:
employee - works for someone.
independent - works for themselves.

SPECIFIC:
entrepreneur_founder — Building or scaling a business. If unsure vs
solopreneur, choose entrepreneur_founder.
solopreneur — One-person operation without scaling ambitions. Solo builders,
indie hackers, side hustles.
freelancer_independent_contractor — Sells labor/skills to clients. Talks about
"my clients," "gigs," rates, finding work.
```

```
small_business_owner – Runs a small business (shop, restaurant, practice,
agency) focused on steady operations rather than growth/scaling (as
entrepreneurs tend to be).
employee_large_org – Works for someone else in a medium-to-large organization
(~50+ people).
employee_small_org – Works for someone else in a small company or startup (<~50
people).
employee_with_side_hustle – Has a steady job but also runs a serious AI side
project (not just hobby use).
academic_institutional – Employed by a university or research institution, e.g.
tenure-track, lab, or grant-funded structure.
student – In school at any level. (This generally means their domain is
not_working)
not_working – Retired, unemployed, on disability, stay-at-home parent, between
jobs.

FALLBACKS:
unclear_structure – Not enough information to classify employment structure.
other_structure – Some other structure not described.

DIMENSION 2 – PROFESSIONAL DOMAIN (what kind of work do you do?):

BROAD:
white_collar_office – Office/desk-based professional work but unclear which
specific type (e.g. mentions having lots of emails).

SPECIFIC:
technical_software – Software engineering, web/app development,
blockchain/Web3, data engineering, DevOps, IT. Building or troubleshooting
technical systems.
science_research – Scientific research in any discipline – bench science,
computational science, social science, clinical research, R&D. Oriented toward
knowledge production, experiments, publications.
creative_arts – Writing, visual art, music, film, design, photography, game
design, content creation. The creative craft is the primary identity (e.g.
content creation for instrumental marketing doesn't count).
education_teaching – Teaching, tutoring, curriculum development
healthcare_medical – Doctors, nurses, therapists, pharmacists, medical
technicians; clinical or patient-facing healthcare work.
legal – Lawyers, paralegals. Licensed or specialized professional services in
law.
```

```
finance_accounting – accountants, financial analysts, auditors. Licensed or
specialized professional services in finance.
sales_marketing_communications – Sales, marketing, PR, communications,
advertising, growth, account management, social media management.
administrative_operations – Office management, executive assistance, data
entry, bookkeeping, HR administration, office coordination, project
coordination.
trades_physical – Electricians, construction, mechanics, manufacturing,
trucking, logistics, agriculture, maintenance, dispatching, etc.
government_policy_military – Government administration, policy analysis, civil
service, military, law enforcement. Distinguished by public-sector
institutional context.
community_social_services – Social work, nonprofit work, community organizing,
religious leadership, counseling (non-clinical), advocacy.
management_executive – Primary identity is managing people or organizations
rather than doing domain-specific work. Executives, general managers, C-suite.
not_working - student (non-PhD i.e. not doing research) or not currently
working for any reason.
other_domain – Doesn't fit neatly into the above.
unclear_domain – Not enough information to classify professional domain. If
they don't mention their job this is likely the best choice, since not_working
is meant for people who are explicitly not working

EXAMPLES:
- Software engineer at Google → employee_large_org,technical_software
- Freelance graphic designer → freelancer_independent_contractor,creative_arts
- Physics professor → academic_institutional,science_research
- Stay-at-home mom using AI for personal projects → not_working,unclear_domain
- High school teacher → employee_large_org,education_teaching
- Medical doctor at a hospital → employee_large_org,healthcare_medical
- PhD student in biology → student,science_research
- no mention of work -> unclear_structure,unclear_domain
- "I'm in the tech sector" → employee,white_collar_office
- Retired accountant → not_working,not_working (since retired, not currently
working in accounting)
- Undergrad student learning ML -> student,not_working

<interview>
{{TRANSCRIPT}}
</interview>
```

```
Think for 2-3 sentences in <thinking> tags first. Don't confuse hobbies/side
projects with their actual profession.
Then output: <answer>structure,domain</answer>
```

Limitations

A few limitations are worth naming:

- *Self-reported occupational categories.* Occupational categories are inferred from what respondents said about themselves, not from verified metadata. 72% did mention some occupational details, but people who didn't mention any such details weren't coded. Our job-level analyses perhaps overrepresent people who use AI at work, and thus have more reason to mention their work in the context of this interview.

- *Label ambiguity.* Classifying open-ended natural language text is often inherently subjective and requires judgment calls.

- *User sample.* All respondents were existing Claude.ai users who opted in to the interview. This skews toward people who have found enough value in AI to keep using it, and likely toward more positive visions than a general population sample would produce.

- *Ordering effects*. The interview structure creates natural priming: by the time respondents reach the question about AI being developed contrary to their values, they've already articulated their hopes. Someone who just described wanting emotional companionship may be primed to voice concerns about authentic connection. This could inflate the co-occurrence between light and shade that we discovered. The interview's length also produced some dropout and possibly less elaboration on later questions; we corrected for this where relevant (e.g., removing those who didn't reach the concerns question from that specific analysis).

Reflections on method. One thing we didn't fully anticipate was how candid people would be. Respondents shared things—grief, mental health crises, financial precarity, relationship failures—that our human user researchers rarely encounter in traditional interviews. While this might be due to the nature of the questions we asked, we also think this reflects something real about the AI interviewer format: there's little social cost to vulnerability when the “someone” on the other end isn’t a person. The same qualities that make people turn to AI for emotional support in their daily lives appear to make them more forthcoming in an AI-led interview. This increased shamelessness may be a double-edged sword: it both leads to sharing more authentic details, and can lead to people acting in a more hostile or dismissive manner than they would be in a human interview.

Response quality was generally high. Few interviews were minimal, incoherent, or clearly disengaged—see breakdown below—which we attribute partly to the conversational format and partly to novelty.

Over the 80,508 interviews, these were the quality breakdowns of the answers. "Not reached" means they dropped off before finishing:

Quality label	"If you could wave a magic wand, what would AI do for you?" (%)	""Has AI ever taken a step towards that vision for you?" (%)	"Are there ways AI might be developed that would be contrary to your vision or what you value?" (%)
Substantive	97.6%	92.5%	88.1%
Refused	1.1%	0.5%	1.1%
Minimal	0.2%	0.3%	0.7%
Rushed	0.1%	0.2%	0.1%
Confused	0.1%	0.2%	0.3%
Not reached	0.9%	6.3%	9.7%

Additional analysis

Geographical breakdown

This was the geographical breakdown of interviewees:

Region	N	%
North America	23,727	29.5%
Western Europe	15,261	19.0%
East Asia	10,250	12.7%
Southern & Eastern Europe	9,385	11.7%
Latin America & Caribbean	8,139	10.1%
South Asia	4,579	5.7%
Southeast Asia	2,828	3.5%
Middle East	1,929	2.4%
Oceania	1,832	2.3%
Sub-Saharan Africa	1,651	2.1%
North Africa	574	0.7%
Central Asia	315	0.4%

Light and shade co-occurrence statistics

For each tension, we measured whether mentioning the benefit predicts mentioning the harm. The answer is yes, consistently: people who raise the positive side are substantially more likely to also raise the negative side. Emotional support and dependence is particularly elevated here; mentioning one makes the other roughly 3× more likely than baseline.

A more revealing split is between experienced and anticipated accounts. We measured co-occurrence separately for each, using φ (a correlation metric that adjusts for base rates, used here because hypothetical mentions are much rarer and would otherwise inflate any ratio-based measure). When people speak from experience, benefit and harm co-occur strongly (avg $\varphi = +0.20$); when they speculate, the link is more than twice as weak (avg $\varphi = +0.07$). The tensions, in other words, are discovered through use—people don't forecast that the thing helping them will also cost them, they learn it.

Pairing	Lift	Lift (Experienced)	Lift (Anticipated)	φ	φ (Experienced)	φ (Anticipated)
Time-saving / Illusory productivity	1.58×	1.59×	1.66×	0.277	0.207	0.047
Economic empowerment / Displacement	1.55×	1.88×	1.37×	0.163	0.089	0.056
Learning / Cognitive atrophy	1.64×	1.63×	1.87×	0.206	0.122	0.070
Emotional support / Dependence	3.04×	4.69×	2.73×	0.326	0.337	0.098
Better decision-making / Unreliability	1.71×	1.78×	1.94×	0.286	0.244	0.059
Average				**+0.252**	**+0.200**	**+0.066**

In the Limitations section above, we discuss "ordering effects" that might prime some of these results or inflate the co-occurrence between light and shade to some degree. However, we believe that the strength of the associations we found (e.g. 3× above baseline) is beyond what priming alone could explain. The priming story is also not entirely obvious—you might expect enthusiasts to defend their desired use case, and articulate concerns about e.g. being *prevented* from engaging in emotional partnership with AI rather

than articulate the first-order risks of what they desire (i.e. emotional dependence). That they disproportionately did the latter is itself interesting.