AI Hiring in a Live Company: Technical Report (Part 2 of “AI managers in a live company”)

Companion report to the article “AI Hiring in a Live Company: The Agent Filled the Funnel. The AI Interview’s Labels Did Not Line Up With Who Is Still Working”. Case study under NDA. Version 1.0 (snapshot day 36). Version 1.1 will add 30-day retention.

Evgenii Arsentev, MD, PhD · Founder & CEO, Dad4Kids · ORCID 0000-0002-9120-7298

Technical report, version 1.0, November 2026 · DOI 10.5281/zenodo.23927438 · CC BY 4.0 · https://arsentev.ai/research/ai-hiring

Cite as: Arsentev, E. (2026). *AI Hiring in a Live Company: A 36-Day Case Study of an AI Hiring Agent and an AI Interview* (Technical report). Zenodo. https://doi.org/10.5281/zenodo.23927438

Part 1 of the series: “AI Call Center Supervisor in a Live Company”, DOI 10.5281/zenodo.23926850.

1. Scope

- Company: a US-facing telemedicine company (name under NDA). Role hired: phone role calling partner physicians (two similar job postings). A third posting for a different role is excluded.
- Period: day 1 = the day the first candidate entered the CRM; snapshot = day 36. Calendar dates are not disclosed.
- Unit of analysis: the candidate. One test card is excluded.
- Read-only access. No system was changed for this report.

Sources

code	source	used for
S1	CRM hiring tables: candidates, events (stage changes, notes), human interviews	funnel, verdict labels, who recorded decisions
S2	AI interview session records (one file per session)	completions, verdicts, session content counts
S3	hiring agent’s working files: application list of the main posting, records of applications read in detail	applications listed and read
S4	CRM user accounts and call log (call records by account)	account issued, first and last call, still working
S5	AI interview page (text shown to the candidate)	what candidates were told

2. Funnel

step	count	source	definition
applications listed, main posting (last scan, day 32)	about 2,400	S3	unique application IDs in the list, rounded
applications read in detail	1,973	S3	unique IDs across all detailed-record files (two file formats, overlap 11, removed)
candidates entered into the CRM, both postings	142	S1	excl. 1 test card
human interviews booked	138 for 130 candidates	S1	all bookings
held	81	S1	status done
no-show	37	S1	status no_show
cancelled	10	S1	
past date, status never updated	10	S1	status scheduled, date before snapshot
show-up rate	68.6%		81 ÷ (81 + 37)
result pass / fail	42 / 37	S1	2 held interviews have no result

Notes: - No “candidate replied” step exists. The first message proposed a time and the booking was made before any reply. - Applications of the second posting were not counted in S3; only its CRM entries are included. - Show-up rate by calendar week of the interview: 27% no-show in week 1, 39% in week 2. - Decision lag: on day 15, 37 candidates had a human pass and no hiring decision recorded in the CRM (a few already had CRM accounts); the oldest had waited 13 days (AI supervisor’s log of open items, cross-checked with S1). - Selection rules were changed 8 times in days 2–15 (hiring agent’s run book). - Losses compared: no-shows 37 of 118 held or missed bookings; human interview fails 37.

3. AI interview

3.1 Process (as disclosed)

1. Personal link, no login, microphone only, about 15 minutes; the answer can be re-recorded.
2. The five questions, the opening, transitions and closing are pre-recorded clips of a company executive, produced as an AI avatar. A voice AI asks at most one follow-up per question and plays the physician in the role play.
3. Part one: five fixed questions (past outbound sales with numbers, working day, other commitments, full-time schedule, start date).
4. Part two: a five-minute role play in which the AI plays a physician with objections and the candidate seeks a first commitment; then the candidate’s own questions.
5. A separate model scores the transcript on the content of the answers only and returns HIRE, MAYBE or NO with a reason.

The CRM stores 90 for HIRE, 50 for MAYBE and 10 for NO. These are labels, not measurements. The “90/100” in the part 1 weekly report is this label.

Vendors and models of the avatar, the voice AI and the scorer are not disclosed. They are not Claude.

3.2 Counts

measure	value	source
candidates invited	41	S1 (22 with a verdict + 19 still “invited”)
completed	22 (54%)	S1, S2 (22 scored sessions, 22 candidates)
window of completions	day 16 to day 26	S2
completions, days 27–36	0 (invitations still sent up to day 32; cause not visible in the records)	S2, S3
verdicts HIRE / MAYBE / NO	7 / 12 / 3	S1 final verdict
re-scored after a scoring rule correction	1 (NO → MAYBE)	S1 events
sessions with no role play in the transcript	2 of 22	S2 scorer notes
human interviews booked in the week of the day-15 rule to stop them	4	S1
human interviews booked, calendar weeks 4–6 (days 21–36)	23	S1

3.3 Agreement with the human interview

Candidates with both an AI verdict and a human result: 16.

AI verdict	human pass	human fail
HIRE (5)	5	0
MAYBE (9)	6	4
NO (2)	2	0

One MAYBE candidate has both a pass and a fail (two interviews), so MAYBE rows sum to 10 results for 9 candidates. The order varies: for some candidates the human interview came first. Both NO candidates with a human result had passed the human interview 15 or more days before the AI verdict.

4. Outcomes

4.1 Definitions

- Account: a CRM user account created for the candidate (for a trial day or a start). Matched to the candidate by name (exact full name; two by first-name match, checked by

hand).
- Started: at least one call logged by the account (S4).
- Working on day 36: account active and calls logged on day 36.
- Trial day: one scheduled working day with real calls, judged by the team lead and recorded the same day. Trial days started on day 21. A candidate scheduled for a trial may make no call at all.

4.2 By AI verdict

group	candidates	account	started	working day 36
AI HIRE	7	7	4	2
AI MAYBE	12	9	8	6
AI NO	3	0	0	0
invited, not completed	19	1	1	0
no AI interview	101	13	6	2
total	142	30	19	10

- Hired at any point (stage “hired” recorded): 12 candidates (15 hire stage changes; 3 are re-records of the same decision). Working on day 36: 10.
- Never made a call after getting an account: 11 of 30.
- Stopped after starting: 9 of 19, after 1 to 4 days with calls. Company ended it after a trial day: 5; left on their own: 1; not recorded who ended it: 3.
- Working share among those who started: HIRE 2 of 4, MAYBE 6 of 8. Among all with the verdict: HIRE 2 of 7 (29%), MAYBE 6 of 12 (50%). With these group sizes the difference is not statistically meaningful (two-sided Fisher exact test: $p \approx 0.63$ among all with the verdict; $p \approx 0.55$ among those who started).
- The three candidates of the part 1 “90/100” claim (scheduled for a trial day on day 21): 1 working, 1 never called, 1 called for one day.

4.3 Time to first call

Equal-window comparison (all CRM entrants, each with at least 7 days of follow-up):

CRM entry	entrants	first call within 7 days of entry	within 14 days	first call by day 36
before day 16	112	1	3	12
days 16–29	28	6	6	6

Among the 19 who started: median 14 days from CRM entry to first call (range 2–25); entrants from day 16 on: median 4 (n=7, range 2–7); before day 16: median 20 (n=12, range 6–25). The medians count only people who started, and late entrants had less time to be counted, so they overstate the gap; the equal-window table does not have this bias.

Account creation → first call: 1 to 4 days for all 19. The CRM entry date is about one day after the application; the application date itself is not stored.

The cohorts differ in more than the AI interview: the earlier one includes the backlog of passed-but-undecided candidates, and trial days started only on day 21. Several AI-track candidates (re-invited from the backlog) are in the earlier cohort.

4.4 Retention

Nobody has reached 30 days. On day 36 the longest-working person had 22 days since the first call. Version 1.1 will report 30-day retention for all 19 who started, by AI verdict.

5. Who recorded the people decisions

decision recorded in the CRM	by a person	by the agent
hire	11	4
reject	25	8
total (stage changes; some candidates have more than one)	36	12

- All 4 agent-recorded hires cite the team lead’s decision in the team chat; one also records the agent’s check of the trial day against CRM activity. The hiring agent is a different agent from the part 1 supervisor and may record decisions.
- Agent-recorded rejections: no-shows (with the team lead’s message cited), departures, rule-based screening (no phone-sales experience), and two AI NO verdicts.
- All 3 AI NO candidates were rejected on that verdict: 2 recorded by the agent, 1 by a person citing “score 10”. Two of them had passed the human interview. There is no record that a person read the transcript before any of the 3 rejections.

6. Disclosure to candidates (S5), measured against two US laws

Neither law applies to this case; they are used as a yardstick. In this interview the candidate is heard, not filmed.

- New York City Local Law 144 (automated employment decision tools that substantially assist or replace discretionary decisions): bias audit (impact ratios by sex and race/ethnicity) within one year before use, public summary of results, notice to candidates at least 10 business days before use, naming the qualifications and characteristics assessed and how to request an alternative process or accommodation. Source: NYC Department of Consumer and Worker Protection, https://www.nyc.gov/site/dca/about/automated-employment-decision-tools.page
- Illinois Artificial Intelligence Video Interview Act (820 ILCS 42), for positions based in Illinois where the applicant records a video: before the interview, notify that AI may be used, explain how it works and what general characteristics it evaluates, obtain consent; share videos only with people needed to evaluate; delete on request within 30 days; employers relying solely on AI to select for an in-person interview report race and ethnicity data. Source: https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=4015&ChapterID=68 (statute text checked via secondary sources)

- Illinois Human Rights Act amendment (HB 3773), in force since 1 January 2026: bars AI use with a discriminatory effect on protected classes and requires notice when AI is used in recruitment and hiring. Source: Illinois General Assembly, HB 3773 (103rd General Assembly).

requirement (yardstick)	**this AI interview**
tell the candidate AI is used	partly: the page says an AI plays the physician in the role play; the avatar video, the voice AI follow-ups and the model verdict are not disclosed
explain how it works and what it evaluates	no
separate consent before the interview	no
record kept and its use stated	partly: the page says answers are saved as text for the team
bias audit	no record of one
alternative or human review	no alternative offered instead of the AI interview; 14 of 19 HIRE/MAYBE candidates also had a human interview; no record of a person reading the transcript before the 3 NO rejections

One candidate asked during a session whether the AI interview was part of the hiring process.

7. Formulas

- Show-up rate = held ÷ (held + no-show).
- Completion rate = completed ÷ invited.
- Days to first call = date of first call logged by the account − CRM entry date of the candidate.
- Working on day 36 = active account with calls on day 36.

8. Assumptions and limitations

- One company, two similar postings, 36 days, no control group; the AI interview was not randomized.
- Small groups; no claim of predictive validity in either direction.
- Name matching of candidates to accounts can fail; two matches rest on first name only.
- Some CRM statuses were not updated (10 past interviews still “scheduled”), so held and no-show counts are lower bounds.
- Recruiter time saved (65–100 hours for 1,973 applications at 2–3 minutes each) is the author’s estimate, not a measurement.
- 30-day retention is not observable yet.

9. What is disclosed and what is not

Disclosed: counts and shares in this report, formulas, how the AI interview works at the level of section 3.1, how decisions were recorded, what candidates were told.

Not disclosed: company name and market details; calendar dates; names or identifying details of candidates and staff; locations; business results and pay; question texts, role-play script and scoring prompt; vendors and models other than “Claude agents” for the hiring agent; page addresses and job posting identifiers; internal operations. Raw data and scripts stay with the company.