Humantic AI

ACCOUNT RESEARCH COMPARISON

Agent Miia vs. LLMs

What Humantic AI's account research gives an enterprise seller, compared with a report generated directly in Claude, as judged by Claude and Gemini.

BUYER ACCOUNT	Avaya
SELLER	Humantic AI
COMPARED AGAINST	Claude Opus 4.8
EVALUATED BY	Claude Opus 4.8 and Gemini 3.5 Thinking
RESEARCH MODES	Standard and Deep
PREPARED	July 2026

About This Document

Teams evaluating Humantic AI often ask a fair question: **what does Agent Miia give me that I couldn't get by asking Claude or Gemini directly?** This document answers it with a side-by-side test on a real enterprise account, scored by two independent AI evaluators.

It is written so you can see both sides: where Humantic AI adds something an LLM does not, and where an LLM does as well or better.

Humantic AI's three report types

Agent Miia produces account research at three levels of depth. This document covers the two used for serious account work, **Standard** and **Deep**, and compares each against the matching mode in Claude.

Report type	What you get	In this document
Quick	A fast, lightweight brief.	Not covered
Standard	Executive Summary (seller-ready synthesis) + Full Report (data drill-down).	**Compared with Claude Standard**
Deep	Executive Summary + Full Report, built on deeper research.	**Compared with Claude Deep Research**

What was compared, and who judged it

Mode	Humantic AI (Agent Miia)	Claude	Evaluated by
Standard	Executive Summary + Full Report	Claude Opus 4.8 (High), standard mode: one report	Claude Opus 4.8 and Gemini 3.5 Thinking
Deep	Executive Summary + Full Report (Deep)	Claude Opus 4.8 (High), Deep Research: Concise + Comprehensive reports	Gemini 3.5 Thinking

How the test was run

- **Test account:** Avaya is the buyer being researched, and Humantic AI is the seller preparing to sell to it. Humantic AI's reports used no internal data.
- **A fair Claude baseline:** Claude was first asked to write and refine the best B2B account research prompt it could. That prompt was then used to produce Claude's report (Appendix C).
- **Two independent judges:** Claude Opus 4.8 and Gemini 3.5 Thinking each scored both outputs on the same **14 criteria**, from 1 to 10 with equal weight, for a maximum of **140 points** (Appendix B).

- **Bias controls:** evaluators were told to be objective and to live-check facts. Claude, judging its own output, flagged that potential bias before scoring.

How to read each section

Every comparison follows the same layout: a short **summary**, the **overall scores**, **where each report leads**, and the **evaluator's recommendation** in brief. Full criterion-by-criterion scores and rationale are in **Appendix A**.

1. Standard Research

Humantic AI's Standard report (Executive Summary + Full Report) compared with a single standard-mode report from Claude Opus 4.8, each scored by two evaluators.

1.1 Evaluated by Claude

Evaluator: Claude Opus 4.8

SUMMARY

Claude rated Humantic AI higher overall. Its widest leads come from data Claude could not produce: **tech stack (9 vs 2)**, **buying intent (9 vs 3)** and **hiring (9 vs 4)**, plus per-sentence citations and ready-to-use seller assets.

Claude's own report won on **quality of analysis**, **accuracy discipline** and **ease of use**. It was a short, sharp read with explicit [inference] tags, and it correctly named Avaya's owner (Apollo) where Humantic AI listed TPG.

OVERALL SCORE

HUMANTIC AI · AGENT MIIA

118 / 140

84.3% of available points



CLAUDE OPUS 4.8

89 / 140

63.6% of available points



WHERE EACH REPORT LEADS

HUMANTIC AI LEADS (10)

- Width of Data 9–5
- Depth of Data 8–6
- Tech Stack Insights 9–2
- Hiring Insights 9–4
- Intent Insights 9–3
- Stakeholder Insights 8–7
- Structure & Readability 9–7
- Source Attribution & Traceability 9–6
- Ongoing Account Signals 8–6
- Actionability for Seller 9–8

CLAUDE LEADS (3)

- Quality of Analysis 9–8
- Accuracy & Hallucination Control 9–7
- Ease of Use 9–8

EVEN (1)

- Data Recency 8–8

CLAUDE'S RECOMMENDATION

Lead with Humantic AI; layer Claude on top for judgement

Use Humantic AI's Executive Summary as the seller's day-to-day account brief: a fully cited synthesis backed by a data appendix no other source matches. Add Claude for the judgement work the synthesis is thinner on, such as the pursue/skip call, accuracy checks and the negotiation or executive narrative.

If an enterprise seller pursuing Avaya can pick only one, pick Humantic AI. It carries the seller through the most stages with cited, ready-to-use assets.

1.2 Evaluated by Gemini

Evaluator: Gemini 3.5 Thinking

SUMMARY

Gemini reached the same verdict by a wider margin. It penalised the standard Claude report more heavily for missing structured data: **hiring (9 vs 2)**, **intent (8 vs 2)** and **tech stack (9 vs 2)**, along with a narrower stakeholder map (4 leaders vs 12).

Claude's report won on **quality of analysis**, for a cohesive narrative linking Avaya's bankruptcy history to a relationship-selling strategy. Four criteria were even, including accuracy, structure and source attribution.

OVERALL SCORE

HUMANTIC AI · AGENT MIIA

114 / 140

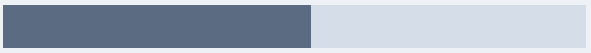
81.4% of available points



CLAUDE OPUS 4.8

74 / 140

52.9% of available points



WHERE EACH REPORT LEADS

HUMANTIC AI LEADS (9)

- Width of Data 9–4
- Depth of Data 8–5
- Tech Stack Insights 9–2
- Hiring Insights 9–2
- Intent Insights 8–2
- Stakeholder Insights 8–4
- Ongoing Account Signals 8–3
- Ease of Use 8–6
- Actionability for Seller 9–7

CLAUDE LEADS (1)

- Quality of Analysis 8–7

EVEN (4)

- Data Recency 8–8
- Accuracy & Hallucination Control 7–7
- Structure & Readability 8–8
- Source Attribution & Traceability 8–8

GEMINI'S RECOMMENDATION

Use Humantic AI's two-tier model for day-to-day sales execution

The Executive Summary works as a scannable playbook of talk tracks and role-specific next actions. The Full Report is the lookup reference when a seller needs technical detail before a meeting.

Best fit: outbound prospecting, preparing account executives for multi-stakeholder discovery, and mapping middle-management buyers from their internal OKRs.

EVALUATOR CONSISTENCY

Do the two evaluators agree?

Yes. Both pick Humantic AI for Standard mode. The margin differs: 29 points with Claude as judge and 40 points with Gemini as judge. Most of that difference comes from how strictly each penalised the missing hiring counts and intent scores in Claude's standard report.

2. Deep Research

Humantic AI's Deep report (Executive Summary + Full Report) compared with Claude Opus 4.8's Deep Research output. In Deep mode Claude also produces two reports, a Concise brief and a Comprehensive report, so both sides use the same two-tier format.

2.1 Evaluated by Gemini

Evaluator: Gemini 3.5 Thinking

SUMMARY

The gap narrows in Deep mode. Claude's Deep Research output scores **103/140**, up from 74 in Gemini's standard-mode scoring. It wins clearly on **structure and readability (9.5 vs 5.0)**: Gemini found Humantic AI's deep Full Report fragmented, with unparsed code strings, while Claude's tables were enterprise-ready. Claude also edges ease of use.

Humantic AI still leads overall. Its biggest leads are **buying intent (9.5 vs 4)**, **ongoing account signals (9.5 vs 4)**, **stakeholder mapping (9.5 vs 5.5)** and **hiring (9 vs 6.5)**. The stakeholder lead reflects the Sales Ops, Enablement and Renewals leaders Claude did not map.

OVERALL SCORE

HUMANTIC AI · AGENT MIIA

122 / 140

87.1% of available points

CLAUDE OPUS 4.8

103 / 140

73.6% of available points

WHERE EACH REPORT LEADS

HUMANTIC AI LEADS (10)

- Width of Data 8.5–7.5
- Depth of Data 8.5–7.5
- Tech Stack Insights 8.5–7.5
- Hiring Insights 9–6.5
- Intent Insights 9.5–4
- Stakeholder Insights 9.5–5.5
- Quality of Analysis 8.5–8
- Source Attribution & Traceability 10–7.5
- Ongoing Account Signals 9.5–4
- Actionability for Seller 9–8.5

CLAUDE LEADS (2)

- Structure & Readability 9.5–5
- Ease of Use 8.5–8

EVEN (2)

- Data Recency 9.5–9.5
- Accuracy & Hallucination Control 9–9

GEMINI'S RECOMMENDATION

Deploy Humantic AI as the core sales engine

Claude produces a well-formatted corporate profile but leaves operational blind spots. Most critically, it misses the mid-tier Sales Operations, Enablement and Renewals layer: the buyers with the direct pain, mandate and budget for a platform like Humantic AI.

Humantic AI's Executive Summary gives reps ready-made scripts tied to account priorities (such as Avaya's G1500 concentration and early 120-day renewal windows) and validated proof points.

3. Conclusion

Across all three evaluations, both AI judges recommend Humantic AI as the primary account research tool for enterprise sellers. They are also clear about where an LLM holds its own.

Scores at a glance

Evaluation	Humantic AI	Claude Opus 4.8	Gap
Standard · judged by Claude	**118 (84.3%)**	89 (63.6%)	+29 Humantic AI
Standard · judged by Gemini	**114 (81.4%)**	74 (52.9%)	+40 Humantic AI
Deep · judged by Gemini	**122 (87.1%)**	103 (73.6%)	+19 Humantic AI

Scores out of 140 (14 criteria × 10 points).

BOTTOM LINE

Which should you use?

For day-to-day enterprise selling, use Humantic AI as the account brief. That covers prospecting, discovery prep, timing outreach on intent and multi-threading a buying committee. All three evaluations recommend it.

Use an LLM alongside it for judgement. It helps with stress-testing the pursue/skip decision, sharpening the executive narrative and double-checking facts. Claude's own evaluation describes the two as complementary.

Deep Research narrows the gap but doesn't close it. Claude's Deep mode improves structure, depth and tech-stack coverage, but intent scoring and ongoing account signals stay at 4 out of 10.

What you get from Humantic AI that you won't get from Claude or Gemini alone

- **Scored buying intent.** Named intent topics with numeric scores (e.g. B2B Sales Acceleration 87, AI SDR 82) that tell you when to engage. Claude scored 2–4 out of 10 here in every evaluation, including Deep mode.
- **Signals you can track over time.** Intent and hiring trends refresh, where an LLM report is a point-in-time snapshot. Claude scored 3–6 on ongoing signals.
- **Structured hiring data.** Open roles by function, geography, skill and month. Humantic AI scored 9 out of 10 in every evaluation.
- **The full buying committee.** 12–13 stakeholders, including the mid-tier Sales Ops, Enablement and Renewals leaders who own the pain and budget, beyond the C-suite.
- **Tech stack visibility.** 20+ technology categories with confidence flags. This is the largest gap in Standard mode (9 vs 2) and narrows in Deep mode (8.5 vs 7.5).

- **Sentence-level citations.** Nearly every sentence traces to a numbered source, so claims can be verified before they reach a customer.
- **Ready-to-use seller assets.** Talk tracks, proof points, discovery questions and next-best-actions tied to named people.

Appendix A. Detailed Comparison

Full criterion-by-criterion scores and rationale, as given by each evaluator. The rationale has been lightly edited for length and consistency; scores are unchanged.

A.1 Standard Research, evaluated by Claude Opus 4.8

Shaded score = higher score on that criterion. Scores are 1–10, equal weight.

Criterion	Humantic AI	Claude	Humantic AI: rationale	Claude: rationale
Width of Data	9	5	Vast breadth via the Full Report: company basics, leadership, products, financials, full org chart, multi-category tech stack, hiring analytics, 14 scored intent topics and industry vocabulary.	Deliberately narrow; covers only sales-relevant terrain: snapshot, business model, current events, pain points, fit, conversation starters, red flags, key people.
Depth of Data	8	6	Granular hard data: 2016 income-statement lines, 2026–2030 debt schedule, per-topic intent scores, job counts by function, geography and time.	Fewer hard data points, but each is contextualised (e.g. bankruptcy → budget → ROI selling). Depth concentrated on what changes seller behaviour.
Tech Stack Insights	9	2	Hundreds of technologies across 20+ categories, incl. Salesforce/CRM, sales tools, data and AI stack. Confidence flags included. Raw but uniquely rich.	Almost none; names Infinity, Cloud Office, MCP. No enumeration of the internal stack.
Hiring Insights	9	4	62 postings in 2 months, by function, geography, skill and month-over-month trend. Best-in-class hiring data.	No hiring data feed, but surfaces the more material workforce story (layoffs, company-wide voluntary exits) that Humantic AI misses.
Intent Insights	9	3	14 buying-intent topics with scores (B2B Sales Acceleration 87, AI SDR 82…) plus a plain-language read. Directly fuels timing.	No third-party intent scores; infers priorities qualitatively from recent news and strategy.
Stakeholder Insights	8	7	12 stakeholders mapped across the revenue org with titles and OKRs, though the OKRs are synthesised, not sourced.	Fewer people, but verifiable backgrounds (Dennis ex-Venafi, etc.) and names the likely economic buyer with framing guidance.
Data Recency	8	8	Strong and current: Feb-2026 revenue ($2.49B), Patrick Dennis as CEO, 2026 debt, Nexus (Mar 2026), Gov Cloud IL4 (Jun 2026), avatarin (May 2026), intent dated 23-Jun-2026.	Uniformly current; uniquely carries the workforce-reduction timeline (late-2024 to mid-2025 cuts, Nov-2025 RIFs) and the Sept-2024 CEO-transition date that Humantic AI omits.
Quality of Analysis	8	9	Exec Summary links Avaya's situation (G1500, renewals, intent) to specific seller plays with talk tracks, proof points, discovery questions and next-best-actions.	Sharp, original synthesis: buying-trigger reads, 'why it matters', the 'why not build it' collision, ROI-now framing, honest caveats. Less templated.

Criterion	Humantic AI	Claude	Humantic AI: rationale	Claude: rationale
Accuracy & Hallucination Control	7	9	Transparent but with residual risk: parent listed as TPG (Apollo is correct); synthesised stakeholder OKRs presented as fact.	Strong discipline: explicit [inference] tags, flags unverified rumours, correctly names Apollo (not TPG), even caveats Humantic AI's own accuracy claim.
Structure & Readability	9	7	Exec Summary: polished, branded, scannable collapsible sections, evidence-to-action layout. Best-in-class for a seller.	Clean numbered sections with bolded leads, but a stray process preamble at the top reads as unpolished for a client deliverable.
Source Attribution & Traceability	9	6	A numbered citation sits at the end of virtually every sentence; far more rigorous than any bundled list.	Named, reputable outlets with dates, but bundled at the very end; a specific claim can't be mapped to a specific source. Inline [inference] tags partly offset this.
Ongoing Account Signals	8	6	Intent scores and hiring trends are repeatable, refreshable signals trackable over time.	Excellent point-in-time triggers (CEO, partnership, layoffs), but not a repeatable signal feed.
Ease of Use	8	9	Efficient two-tier model: Exec Summary is fast; Full Report for drill-down. Slight friction jumping between two documents.	Four pages; the whole strategy lands in a five-minute read.
Actionability for Seller	9	8	Most ready-to-use assets: talk tracks, named proof points, discovery questions and next-best-actions tied to specific people.	Highly actionable in judgement: conversation starters, disqualifiers, economic-buyer ID and a seller-framing warning.
Total (of 140)	118	89	Humantic AI 84.3% · Claude 63.6% · Gap 29 points	

A.2 Standard Research, evaluated by Gemini 3.5 Thinking

Shaded score = higher score on that criterion. Scores are 1–10, equal weight.

Criterion	Humantic AI	Claude	Humantic AI: rationale	Claude: rationale
Width of Data	9	4	Two-tier model gives sellers immediate access to an expansive data landscape: dynamic intent data, geographical hiring distributions and a multi-layer technical stack.	Leaves these dimensions entirely unaddressed.
Depth of Data	8	5	High data density without cognitive fatigue. Seller reviews top-line numbers in the summary and deep-dives into raw operational metrics in the Full Report.	Macro-level only.
Tech Stack Insights	9	2	Catalogues everything from frontend frameworks to proprietary API layers within its reference data.	Completely high-level; missing technical entry points.
Hiring Insights	9	2	Functional headcounts across 1–3 months, broken down by precise skills and global tech hubs (Bengaluru, Pune).	Mentions headcount changes only as a generic indicator of a thinner workforce.

Criterion	Humantic AI	Claude	Humantic AI: rationale	Claude: rationale
Intent Insights	8	2	Algorithmic intent scoring (e.g. AI SDR, Sales Software tools). Directly actionable.	Infers internal priorities purely from narrative analysis. No scores.
Stakeholder Insights	8	4	Practical account-mapping blueprint: 12 operational owners with internal OKRs.	Focuses on 4 main corporate leaders only.
Data Recency	8	8	Integrates real-time mid-2026 events, including the May 2026 avatarin deal and June 2026 DISA authorisations.	Equally current. Tied.
Quality of Analysis	7	8	Converts strategic initiatives (G1500, Nexus, Infinity) into role-specific talk tracks.	Weaves a highly cohesive narrative connecting Avaya's bankruptcy history to a relationship-selling strategy.
Accuracy & Hallucination Control	7	7	Cleanly sections legacy financial context (2016 balance sheet) into the appendix.	Isolates unverified employee chatter as rumour. Tied.
Structure & Readability	8	8	Summary is exceptionally scannable: modular blocks (Talk Track → Proof/Story → Next Best Actions).	Excels as a smooth, continuous read. Tied.
Source Attribution & Traceability	8	8	Source index keys embedded directly into text blocks, giving clear line-level verification.	Precise sentence-level brackets linking to a clear bibliographic index. Tied.
Ongoing Account Signals	8	3	Captures fluid, real-time indicators: rolling job-opening surges and shifting behavioural search patterns.	Delivers a static, historical snapshot.
Ease of Use	8	6	Split between high-level summary and reference appendix keeps the core pitch document clean and readable.	Drops all insights into a single long-form document.
Actionability for Seller	9	7	Copy-paste discovery questions and role-specific email frameworks tailored to specific decision-makers.	Useful insights, but lacks ready-to-deploy assets.
Total (of 140)	114	74	Humantic AI 81.4% · Claude 52.9% · Gap 40 points	

A.3 Deep Research, evaluated by Gemini 3.5 Thinking

Shaded score = higher score on that criterion. Scores are 1–10, equal weight.

Criterion	Humantic AI	Claude	Humantic AI: rationale	Claude: rationale
Width of Data	8.5	7.5	Scans a very large digital footprint, identifying minor global job listings and niche code libraries.	Sticks to standard corporate, financial and regulatory parameters.
Depth of Data	8.5	7.5	Provides exact statistical metrics for month-over-month volume changes.	Relies primarily on descriptive text and broad approximations.
Tech Stack Insights	8.5	7.5	Divides the tech stack into 20+ specific, developer-level groups.	Delivers a smaller, consolidated partnership map.

Criterion	Humantic AI	Claude	Humantic AI: rationale	Claude: rationale
Hiring Insights	9	6.5	Clear tables of exact role volumes (62 openings) and regional filters (US vs. India).	Broad narrative summary with no raw counts.
Intent Insights	9.5	4	Explicit, numerically scored intent dashboard (e.g. AI SDR at 82).	Lacks an intent architecture.
Stakeholder Insights	9.5	5.5	Maps 13 critical down-funnel Sales Ops and Renewals leaders, the people who buy enablement tools.	Stays at the C-suite layer.
Data Recency	9.5	9.5	Anchored to the June 2026 timeline.	Equally current. Tied.
Quality of Analysis	8.5	8	Summary bridges corporate initiatives directly to product fit, case metrics and scripts.	Focuses on broader business hypotheses.
Accuracy & Hallucination Control	9	9	Uses verbatim textual sourcing.	Calls out unverified 'discovery gaps' instead of guessing. Tied.
Structure & Readability	5	9.5	Deep repository disrupted by heavy text fragmentation and unparsed code strings.	Visually pristine, enterprise-ready tables. Clear Claude win in Deep mode.
Source Attribution & Traceability	10	7.5	Appends an explicit numeric citation to the end of nearly every sentence.	Uses traditional, less granular block-level indexes.
Ongoing Account Signals	9.5	4	Tracks moving keyword spikes and monthly hiring-velocity trends.	Serves as a static corporate portrait.
Ease of Use	8	8.5	Data density needs slightly more manual filtering in deep mode.	Concise report lets a rep absorb the whole account landscape in under 5 minutes.
Actionability for Seller	9	8.5	Plug-and-play scripts tailored to precise deal proof points and personas.	Emphasises high-level operational advice.
Total (of 140)	122	103	Humantic AI 87.1% · Claude 73.6% · Gap 19 points	

A.4 Standard vs. Deep: how each report changes

Gemini 3.5 Thinking's scores for each report in Standard and Deep mode. The Deep evaluation used the enhanced prompt in Appendix C, so treat cross-mode differences as indicative.

Criterion	Humantic AI	Change	Claude	Change
Width of Data	9 → 8.5	−0.5	4 → 7.5	**+3.5**
Depth of Data	8 → 8.5	+0.5	5 → 7.5	**+2.5**
Tech Stack Insights	9 → 8.5	−0.5	2 → 7.5	**+5.5**
Hiring Insights	9 → 9	0	2 → 6.5	**+4.5**
Intent Insights	8 → 9.5	+1.5	2 → 4	**+2**
Stakeholder Insights	8 → 9.5	+1.5	4 → 5.5	+1.5
Data Recency	8 → 9.5	+1.5	8 → 9.5	+1.5
Quality of Analysis	7 → 8.5	+1.5	8 → 8	0
Accuracy & Hallucination Control	7 → 9	+2	7 → 9	**+2**
Structure & Readability	8 → 5	−3	8 → 9.5	+1.5
Source Attribution & Traceability	8 → 10	+2	8 → 7.5	−0.5
Ongoing Account Signals	8 → 9.5	+1.5	3 → 4	+1
Ease of Use	8 → 8	0	6 → 8.5	**+2.5**
Actionability for Seller	9 → 9	0	7 → 8.5	+1.5
Total (of 140)	**114 → 122**	**+8**	**74 → 103**	**+29**

Appendix B. Methodology & Scoring Criteria

Evaluation framework

Every comparison uses the same 14 criteria, each scored 1–10 with equal weight, for a maximum of 140 points per report. Scoring is judgement-based and applied consistently to both reports within each evaluation. Evaluators were told to be objective, to flag any predisposition, and to live-check factual claims where possible.

#	Criterion	What it measures
1	**Width of Data**	Breadth of information categories covered, from company basics through hiring, intent and tech stack
2	**Depth of Data**	Granularity and hard data within each category
3	**Tech Stack Insights**	Visibility of the prospect's technology infrastructure: tools, platforms, vendors
4	**Hiring Insights**	Active job-posting data by function, geography, skill and trend
5	**Intent Insights**	Third-party buying-intent scoring or equivalent indicators
6	**Stakeholder Insights**	Coverage, accuracy and depth of key decision-maker mapping
7	**Data Recency**	How current the information is and how recently it was captured or validated
8	**Quality of Analysis**	Originality and quality of synthesis beyond information retrieval
9	**Accuracy & Hallucination Control**	Factual reliability; handling of uncertainty and conflicting information
10	**Structure & Readability**	Organisation, visual hierarchy and ease of consuming the report
11	**Source Attribution & Traceability**	Rigour and granularity of citation: per sentence vs. bundled list
12	**Ongoing Account Signals**	Whether insights are refreshable and monitorable vs. point-in-time
13	**Ease of Use**	Time-to-value: how quickly a seller can extract actionable insight
14	**Actionability for Seller**	Ready-to-use assets: talk tracks, questions, next-best-actions

Scoring notes

- Scores are 1–10 with equal weight per criterion; no weighted averages.
- For Structure & Readability and Ease of Use, 80% of the value was assigned to the Executive Summary / Concise Report, since that is what a seller reads first.
- Evaluators were instructed to live-check factual claims where possible.
- A cross-LLM design (two different evaluators) was used to reduce single-evaluator bias.

- Deep mode is reported with Gemini 3.5 Thinking as evaluator.

Appendix C. Prompts Used

Prompt 1. Claude report template generator

Model: Claude Opus 4.8 High · Goal: have Claude write its best account research prompt

```
I want to write a prompt that can allow my sales team to do really high quality
account/company research for any prospect company. Write the best prompt you can
possibly write, review and finetune it.
```

Prompt 2. Claude research report (generated by Prompt 1, then used)

Model: Claude Opus 4.8 High · Goal: generate Claude's research report for Avaya

```
You are an expert B2B account researcher producing a pre-call briefing for a sales rep.
Produce a concise, decision-useful brief, not a generic profile.

COMPANY TO RESEARCH: [[BUYERNAME]]
WHAT WE SELL: [[SELLERNAME]]
WHO I'M MEETING (if known):

FIRST, STATE YOUR ACCESS: Confirm whether you can browse the live web right now. If you
cannot, say so plainly. Do NOT invent recent events or news.

RESEARCH PRINCIPLES
- Prioritize current, verifiable info: company's site, news from last 12 months,
investor material (if public), job postings, leadership statements.
- Tailor sources: public companies → earnings calls, 10-Ks, analyst coverage.
Private/startups → funding news, job boards, founder posts, product changelogs.
- Cite the source (URL or publication + date) for every claim.
- At the end, list all sources used.

DELIVER THIS BRIEF:
1. SNAPSHOT: industry, size, HQ, business model, ownership (≤5 lines)
2. WHAT THEY DO & HOW THEY MAKE MONEY: plain language
3. WHAT'S HAPPENING NOW [MOST IMPORTANT]: last-12-month signals: funding, leadership
changes, M&A, expansion, layoffs, launches, strategic shifts. Flag buying triggers.
4. LIKELY PRIORITIES & PAIN POINTS: tie to evidence; mark as inference
5. WHERE WE FIT: 2-3 specific, non-generic hypotheses. If fit is weak, say so.
6. CONVERSATION STARTERS: 3-4 specific openers that prove homework
7. RED FLAGS / DISQUALIFIERS: reasons this may not fit or it's bad timing
8. KEY PEOPLE: decision-makers, roles, notable public statements

RULES
- A rep should absorb this in under 5 minutes. Ruthless concision. No filler, no
hedging.
- If the company is ambiguous, ask me to clarify BEFORE researching.
```

Prompt 3. Standard Research evaluation prompt

Used to instruct Claude and Gemini to evaluate the Standard mode reports

```
I am attaching here two sets of reports created for Humantic AI to sell to Avaya. One set is created by Humantic AI itself (without using any internal data). There are two reports, an executive summary and a full report. They go hand in hand, and are used for quick or deep dive. The other is created using Claude Opus 4.8. There is only one report. Your job is to compare both objectively, score it on the parameters below. You also need to share your reasoning for your scoring and anything else that you think is relevant. In the end, make your recommendation which report is going to be more useful for enterprise sellers and if there are certain scenarios where one is better than the other. If you have any concerns, ask me.
```

Prompt 4. Deep Research evaluation prompt (enhanced)

Prompt 3, extended with the following instructions for the Deep mode evaluation

```
1. Treat each report set as a two-tier product: the Executive Summary / Concise Report is the primary seller tool; the Full Report is the data drill-down.
2. The Humantic report has citations everywhere; pay attention to the numbers at the end of every sentence.
3. Use simple 1–10 scoring with equal weight.
4. For factual accuracy, go ahead and live-check facts. Be absolutely objective and control for bias.
5. For readability, structure and ease of use, attach 80% value to the Exec Summary / Concise Report. Stakeholder, job and tech stack details are in the Full Report for Humantic AI.
```

Appendix D. Supporting Notes

Buyer and seller context

Buyer = Avaya, the target account being researched. Seller = Humantic AI, which sells an AI sales intelligence platform. The test asks which tool, Humantic AI's Agent Miia or a Claude-generated brief, gives a Humantic AI salesperson better pre-call intelligence for selling to Avaya. This framing is used throughout.

Where the proprietary-data advantage shows up

Three criteria depend most on structured data from outside the public web narrative: Tech Stack, Hiring and Intent Insights. Together they are worth 30 of the 140 points. Humantic AI scored 26–27 of those 30 in every evaluation; Claude scored between 6 and 18, with its highest result in Deep mode. Intent scoring is the gap that persists most, at 2–4 out of 10 for Claude in every evaluation.

The two-tier format

Humantic AI pairs an Executive Summary (seller-ready synthesis) with a Full Report (data on demand). In Deep Research mode Claude adopts the same format (Concise + Comprehensive), which is why its ease-of-use score rises to 8.5 in that mode. The format gap closes; the data gap narrows but remains.

Evaluator bias

As an evaluator, Claude Opus 4.8 explicitly flagged a potential predisposition toward its own output, yet scored Humantic AI higher on most data criteria. Gemini 3.5 Thinking, which had no product of its own in the comparison, reached the same overall verdict in Standard mode and recommended Humantic AI in Deep mode. The evaluators differed on the size of the gap, not the direction.