

AGENCY 150 | WHITE PAPER

Listening at Scale

The science and process behind Vega for strategic advising, and how it compares with surveys and traditional consulting engagements

September 2026

Lamar Pierce & Natasha Swift

Agency 150, Inc.

Executive summary

Every strategy depends on knowledge that leadership does not fully hold. The people who know where customers are being lost, why a process quietly fails, or which initiative the field has already given up on are spread across the organization, and most of them are never asked in a way that lets them answer properly. This paper explains how Vega, Agency 150's AI-led interview platform, addresses that gap in strategic advising, the science behind the approach, and how it compares with the two tools organizations have relied on for decades: employee surveys and consulting engagements.

The argument runs as follows. Strategic knowledge is dispersed (Hayek, 1945), much of it is tacit (Polanyi, 1966), and much of it is withheld. In one well-known study, 85 percent of employees reported having stayed silent about an issue they thought was important (Milliken, Morrison, & Hewlin, 2003). Surveys reach everyone but cannot hear reasons; they measure what leadership already thought to ask, through instruments with well-documented biases (Podsakoff et al., 2003; Nisbett & Wilson, 1977). Consulting engagements hear reasons in depth, but from a small and mostly senior sample, at a cost and on a timeline that limits how often the exercise can be repeated. Figure 1 places the three approaches on the two dimensions that matter: breadth and depth.

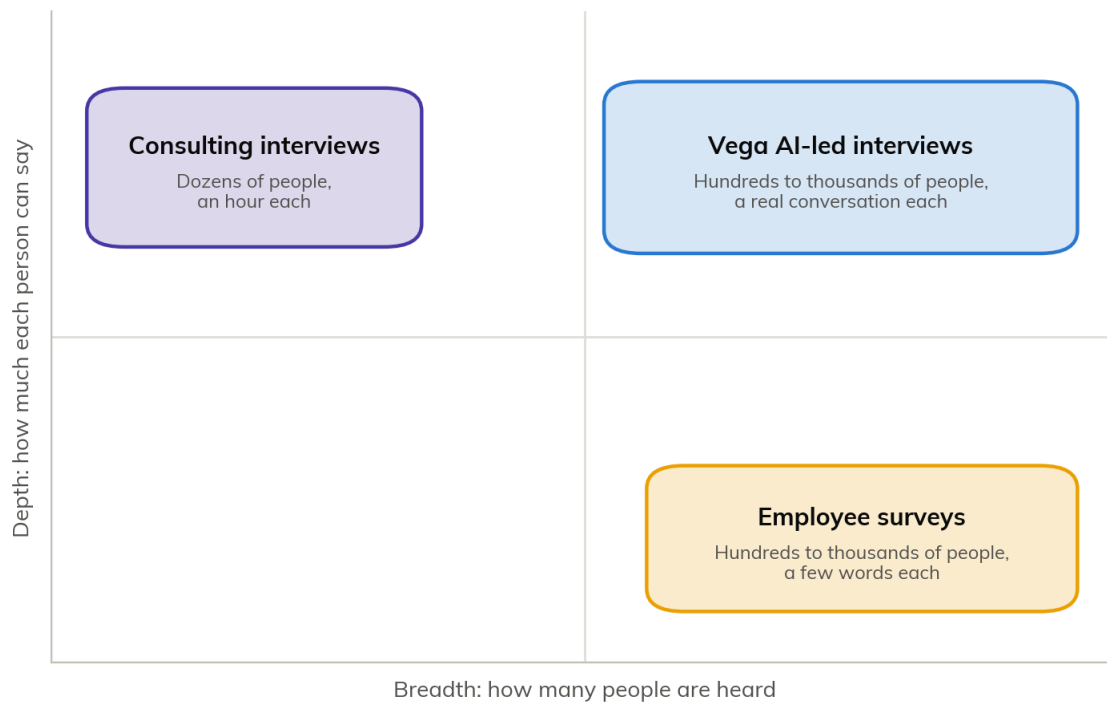


Figure 1. Three ways to learn what an organization knows, placed by breadth (how many people are heard) and depth (how much each person can say).

AI-led interviewing reaches the corner of Figure 1 that human-led interviews and surveys cannot. Peer-reviewed and working-paper evidence now shows that people complete AI-led interviews at high rates, rate the experience positively, and produce materially richer accounts than they give in open-ended survey fields, with transcript quality that trained social scientists rate at parity with an average human interviewer (Chopra & Haaland, 2023; Geiecke & Jaravel, 2024; Xiao et al., 2020). Large language models can then code those transcripts with agreement rates that match or exceed trained human annotators (Gilardi, Alizadeh, & Kubli, 2023; Törnberg, 2025).

Vega turns that evidence into a repeatable advisory process: frame the strategic question, design the interview, interview hundreds or thousands of people in days, code the transcripts against a firm-specific ontology, promote only findings that are linked to specific evidence, and put an advisor in front of the client with the result. The platform does the listening at scale, then provides analysis and judgment with

support of human experts. The paper closes with a section on what Vega does not do, guidance on when each of the three approaches is the right choice, and the design requirements (consent, pseudonymity, evidence links, and retention terms) that make organizational listening trustworthy.

1. The strategic knowledge problem

Strategy work begins with a diagnosis, and a diagnosis is only as good as the information behind it. The difficulty is that the information a strategist needs is not held in any one place. Friedrich Hayek made the point in 1945 for economies as a whole: the knowledge that matters is never given to a single mind "in its totality," but exists as "the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess," and much of it is knowledge of "the particular circumstances of time and place" held by "the man on the spot" (Hayek, 1945). Inside a firm, the same holds: the regional manager knows why the new pricing scheme is being quietly discounted, the engineer knows which integration will take three times as long as the plan says, and the nurse knows why the handoff protocol gets skipped on night shifts.

A second feature of this knowledge makes it harder to collect. Much of it is tacit. Michael Polanyi's formulation that "we can know more than we can tell," describes expertise that people can act on but cannot readily summarize on demand (Polanyi, 1966). Nonaka's theory of organizational knowledge creation treats the conversion of this tacit knowledge into explicit, shareable form as the central act of organizational learning, and it is a conversational act: people articulate what they know when someone asks a good question and then a better follow-up (Nonaka, 1994; Nonaka & Takeuchi, 1995). A checkbox or numerical rating does not do this.

The strategy literature has long recognized that what a firm actually does emerges from the bottom as much as it is directed from the top. Bower's study of resource allocation showed corporate strategy taking shape through the project proposals that middle managers chose to champion (Bower, 1970). Burgelman's work on internal venturing found the same pattern: realized strategy is substantially the aggregate of decentralized choices, later rationalized as intent (Burgelman, 1983). Mintzberg's distinction between deliberate and emergent strategy makes the point general: strategy is "a pattern in a stream of decisions," and a good share of that

pattern is never planned (Mintzberg, 1978; Mintzberg & Waters, 1985). If strategy is partly emergent, then the people who are producing the emergence hold information that the strategist needs and usually lacks.

The final feature of this knowledge is that people withhold it. Milliken, Morrison, and Hewlin (2003) found that 85 percent of the employees they interviewed had been in a situation where they felt unable to raise an issue with a supervisor even though they believed it was important. Detert and Edmondson (2011) traced this silence to implicit, taken-for-granted rules that employees carry from job to job: do not embarrass the boss, do not bring a problem without a solution, do not speak up in front of others. Edmondson's earlier work showed that whether a team surfaces problems depends on psychological safety, a shared belief that candor will not be punished (Edmondson, 1999). Morrison's reviews of the voice literature document how much organizational learning this self-censorship costs, and how rarely organizations build channels to counteract it (Morrison, 2011, 2014). Macnamara's study of organizational communication across government, business, and non-profit organizations found an "architecture of speaking" with almost nothing comparable devoted to listening (Macnamara, 2016).

The consequences show up in execution. Sull, Homkes, and Sull (2015) report that just over half of top-team members say they have a clear sense of how the company's major priorities fit together, and fewer than a third of their direct reports do. Strategy that isn't understood one level down isn't carried out one level down, and leadership usually finds out late.

2. Two Tools, Two Blind Spots

2.1 Surveys: breadth without depth

Employee surveys are the dominant tool for organizational listening because they scale. They can reach every employee at low marginal cost; they produce numbers that can be tracked over time and compared against benchmarks. For monitoring known quantities, such as engagement, intent to stay, or satisfaction with a specific program, that is the right design. The blind spot appears when the question is

strategic rather than evaluative: not "how satisfied are you" but "what is going wrong, and why."

Three limits follow from the instrument itself. First, a survey can only ask what its author already thought to ask. The categories are fixed before the first response arrives, so a survey is well suited to confirming hypotheses and poorly suited to discovering them. Second, the format restricts what a respondent can say. Open-ended fields exist, but the responses are short and get shorter as a questionnaire wears on; Galesic and Bosnjak (2009) found that questions placed later in a web survey received faster answers, more item non-response, shorter open-ended text, and more straightlining. Smyth and colleagues (2009) showed that open-ended answers improve only when the design goes out of its way to invite length, which most organizational surveys do not. Third, the survey is a self-report on a fixed scale, and the psychometrics of self-report are well understood. Podsakoff and colleagues (2003) cataloged the ways in which common method variance, socially desirable responding, consistency motives, and mood inflate or deflate the correlations a survey produces. Paulhus (1991) distinguished deliberate impression management from self-deceptive enhancement; both are present whenever someone rates their own team on a five-point scale. Most fundamentally, Nisbett and Wilson (1977) showed that people are unreliable reporters of the causes of their own attitudes and behavior. Asked "why," they supply a plausible theory rather than an observation. A survey that asks people to rate the drivers of their engagement is collecting theories.

Response rates compound these problems. Baruch and Holtom (2008), reviewing 1,607 published studies, found average response rates of 53 percent for individual-level surveys and 36 percent for organization-level surveys. Anseel and colleagues (2010) similarly found rates drifting downward even as researchers used more techniques to prop them up. Nonrespondents are rarely a random subset. And even when the numbers come back, organizations often do not act on them. A 2021 Gartner survey found that only 29 percent of employees trusted their organization with the data it collected through feedback processes (Zielinski, 2022). Gallup's global engagement figure fell to 21 percent in 2024, only the second decline in twelve years, after more than two decades of survey-driven engagement programs (Gallup, 2025). None of this means surveys are useless. It means that a survey is an

instrument for measurement, not for discovery, and strategic advising is mostly discovery.

2.2 Consulting engagements: depth without breadth

The strategy consulting engagement solves the depth problem the survey cannot. A skilled consultant conducting a one-hour interview with a division head can follow the thread, test a hypothesis, and come away with an account of the business that no questionnaire could produce. Add the firm's analytical toolkit and access to external data, and you get a considered, defensible point of view. The evidence that good management advice creates value is real: in a randomized trial, Bloom and colleagues (2013) found that a consulting intervention in Indian textile plants raised productivity by 17 percent in the first year, and Bruhn, Karlan, and Schoar (2018) found lasting gains in employment and wages for Mexican small firms that received consulting.

The blind spot is the sample. The economics of the consulting model rest on the leverage pyramid described by Maister (1993): a small number of senior partners direct a larger number of junior staff, and every hour of interviewing is billed. The result is that even a large engagement hears from a limited number of people, and those people are disproportionately senior, articulate, and close to the sponsor. The front-line knowledge that Hayek described, the knowledge of particular circumstances, is exactly the knowledge this sample leaves out. Consultants know this and supplement interviews with site visits, workshops, and data pulls, but the constraint is structural: depth is purchased one hour at a time.

Three further limits matter for a client deciding how to spend a strategy budget. The first is time. Engagements run from a couple of months to the better part of a year, which means the diagnosis is a snapshot that has begun to age before it is presented. The second is opacity. Christensen, Wang, and van Bever (2013) observed that clients find it very difficult to judge a consulting firm's performance in advance, so price becomes a proxy for quality; the same authors predicted that the industry would be unbundled by firms that modularize the analytical work that used to require a full team. The third is what the client keeps. The team leaves with its interview notes and its judgment; the client keeps a report. The raw material, which is the organization's own knowledge about itself, is rarely returned in a form the

client can query again. Critics such as Mazzucato and Collington (2023) go further, arguing that heavy reliance on external advisers hollows out the client's own capacity to think, and O'Mahoney and Sturdy (2016) document how consulting firms shape which management ideas spread, well beyond the boundaries of any project. One need not accept the strongest version of those critiques to see the pattern: the engagement model is optimized to deliver a point of view, not to leave the client better able to listen.

3. The Science Behind Vega

Vega is an AI-led qualitative interview platform. A participant receives an invitation, opens a conversation with an AI interviewer, and talks (by text or voice) for as long as the conversation warrants, usually twenty to thirty minutes. The interviewer follows a guide written for the engagement, asks follow-up questions when an answer is thin or interesting, and adapts to what the participant has said. The transcripts are then coded, synthesized, and turned into findings that an Agency 150 advisor brings to the client. This section explains why each step works, drawing on research from economics, psychology, and computer science.

3.1 Conversation as an instrument

The critical question is whether a machine interviewer can elicit the kind of account a good human interviewer gets. Three independent research programs say yes. Chopra and Haaland (2023) built an AI interviewer and used it to conduct 395 interviews on why people do not invest in the stock market. Over 95 percent of participants completed the interview voluntarily; 82 percent rated the experience positively, 74 percent said it felt natural, 53 percent said they would strictly prefer the AI interviewer to a human one, and 95 percent said they would do it again. The median interview ran 27 minutes and produced roughly 300,000 words across about 7,000 open-ended answers. The interviews were thematically richer than single open-ended survey questions on the same topic.

Geiecke and Jaravel (2024) tested an open-source AI interviewer across four domains and five language models. Against an ordinary open-text survey field, the AI interview produced a 142 percent increase in word count. Trained sociologists, asked to rate the transcripts on a scale where 3 represented an average human

expert interviewer, gave the AI interviews a mean score of 2.95, and judged the AI interview more informative than the open-text response in 75 percent of matched comparisons. Among participants, 81 percent preferred the AI interview to a human one or were indifferent, and 76 percent would not have preferred an open-text field instead.

Xiao and colleagues (2020), in a controlled comparison published in *ACM Transactions on Computer-Human Interaction*, ran the same open-ended questions through a conversational AI and through a standard web survey with about 600 participants. Chatbot responses were 39 percent more informative and 26 percent higher on an overall quality index; gibberish responses fell from 9 percent to 2.5 percent; and completion rates more than doubled, from 24 percent to 54 percent. Participants wrote more (90 words versus 64 per answer), spent longer (24 minutes versus 18), and disclosed more about themselves.

The mechanism is the follow-up question, which a survey field cannot supply. When a participant writes "the handoff process is broken," the interviewer asks which handoff, what happens, and what they have tried. That second and third question is where the tacit knowledge Polanyi described becomes explicit, and it is why the qualitative tradition has always relied on interviews rather than questionnaires for discovery. What has changed is that the interviewer no longer has to be a person, so the number of interviews no longer has to be small.

3.2 Why people say more to a machine that listens

It would be reasonable to expect people to hold back from a machine. The disclosure evidence points the other way. Lucas, Gratch, King, and Morency (2014) interviewed participants using a virtual interviewer and told half of them it was controlled by a computer and half that it was operated by a person. Those who believed they were talking to a computer reported less fear of disclosure, engaged in less impression management, and were rated by independent observers as more willing to share personal information. The result fits the voice literature: what suppresses candor at work is the anticipated reaction of another person, particularly one with power over you (Detert & Edmondson, 2011). An AI interviewer has no reaction to anticipate and no memory of who you are.

Vega is built so that this assurance holds in practice. Participants are pseudonymous. Identity is held in a separate vault from the interview record, findings are reported at levels of aggregation that do not identify individuals, and the consent terms state in plain language what will be shared with the client and what will not. This is an ethical requirement, but it is also a practical one. Candor is the raw material, and candor depends on the participant believing the assurance.

3.3 Saturation, heterogeneity, and why breadth still matters

A reasonable objection from a qualitative researcher is that one does not need hundreds of interviews. Guest, Bunce, and Johnson (2006) coded 60 in-depth interviews and found that 73 percent of all the codes that would eventually appear had surfaced by the sixth interview and 92 percent by the twelfth; between the thirtieth and sixtieth interview only five new codes emerged (Figure 2). Hennink and Kaiser's (2022) systematic review of 23 saturation studies found that interviews with homogeneous populations on narrow questions typically saturate within 9 to 17 interviews. If a dozen interviews find the themes, why conduct a thousand?

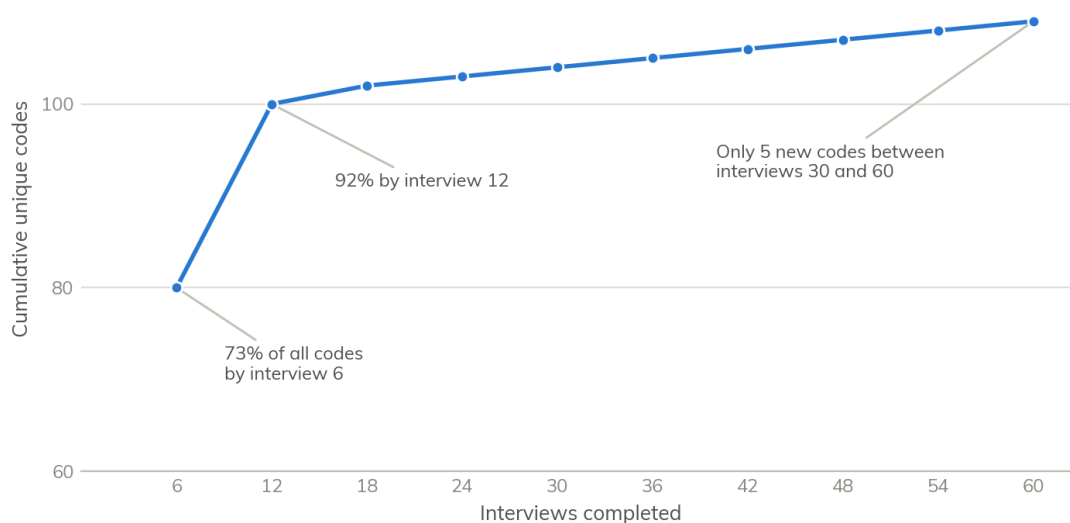


Figure 2. Thematic saturation in a single homogeneous population. Cumulative unique codes across 60 interviews, from Guest, Bunce, and Johnson (2006). Intermediate points are interpolated from the reported milestones.

The answer lies in the word homogeneous. Saturation numbers apply to a uniform population asked a narrow question. An organization is the opposite: it is a set of

overlapping populations (regions, functions, levels, tenures, sites) that experience the same strategy differently, and the strategic question is usually about exactly those differences. Hennink and Kaiser found that meta-themes across heterogeneous samples required 20 to 40 interviews or more, and that is for a single cross-cutting theme, not for a map of where the theme holds and where it does not. Twelve interviews will tell you that a pricing change is unpopular. They will not tell you that it is unpopular in three regions for one reason and in a fourth for a different reason that the head office has never heard. Finding that requires enough interviews in each segment to reach saturation within the segment, which for a large organization means hundreds or thousands of conversations. Hong and Page's (2004) modeling result, that diverse groups of problem solvers can outperform groups of individually stronger ones, points to the same conclusion from a different direction: the value is in the variety of perspective, and variety is a function of breadth.

A second reason breadth matters has nothing to do with statistics. Strategy is implemented by the people who were asked, and asking is itself an intervention. An employee who has spent twenty minutes explaining what is not working, and who then sees the resulting strategy reflect it, has a different relationship to that strategy than one who received it in a town hall. The survey literature's action gap, in which only about half of employees see anything change after giving feedback, is a warning: breadth creates an expectation, and the process has to be designed to meet it.

3.4 From transcripts to findings

A thousand transcripts are not a finding. The analytical step has traditionally been the slowest and least reproducible part of qualitative work, and it is where large language models have most changed the economics. Vega's analysis follows the structure of thematic analysis as codified by Braun and Clarke (2006): familiarization, initial coding, theme development, review, definition, and reporting. What differs is that language models do the coding using a codebook and people review it, rather than people doing it alone.

The reliability evidence for this is now strong. Gilardi, Alizadeh, and Kubli (2023), in *PNAS*, gave the same codebook to trained annotators, crowd workers, and a zero-shot language model across 6,183 texts; the model's inter-coder agreement

was 91 to 97 percent depending on settings, against 79 percent for trained annotators and 56 percent for crowd workers, at roughly one-thirtieth the cost. Törnberg (2025), in *Social Science Computer Review*, found that GPT-4 classified political text with 93 percent accuracy, compared with 83 to 86 percent for expert coders and 80 to 85 percent for crowd workers, with the difference statistically significant across countries and languages. On structured coding tasks, model-assisted coding is more consistent than human coding.

Consistency is not the same as truth, and this is where Vega's design departs from a generic "AI summary." Vega's data model enforces a rule we call evidence gating: no finding is promoted to a client-facing report unless it is linked to specific passages in specific interviews. A theme is a labeled set of coded passages that an analyst and the client can open and read, rather than a paragraph the model wrote. Every code a model applies carries an audit trail, and analysts can override codes, merge themes, and flag risks, with each action recorded. The practical effect is that when an advisor tells a leadership team that the field has lost confidence in a product line, the advisor can show the 140 passages behind that sentence, broken out by region and role, rather than asking the team to trust a summary.

3.5 A firm-specific ontology and memory

The final scientific element is structural rather than statistical. Generic listening tools code every organization against the same categories, which is why their output tends to read the same from one company to the next. Vega instead builds a firm-specific ontology: a typed model of the client's organization (units, roles, initiatives), its research (studies, instruments, sessions, responses), and its analysis (codes, themes, findings, evidence, recommendations). Themes are defined in the client's own vocabulary and refined study by study. This is what makes repeated listening cumulative rather than episodic: the second study on a topic starts from the structure the first one produced, and a finding from March can be compared with a finding from September because both were coded against the same evolving model.

The ontology also governs what you can compare across clients. Client themes belong to the client. A separate layer of canonical themes, owned by Agency 150, allows cross-client benchmarking only where the client's data agreement permits it, and only through an explicit mapping step. The separation is deliberate: it lets a

client benefit from what the platform has learned elsewhere without their own material becoming someone else's benchmark by default.

4. The Vega process for strategic advising

The science above is delivered through a six-step process (Figure 3). Steps one, two, and six are led by an Agency 150 advisor; steps three through five are led by the platform, with analysts in the loop.

Platform-led steps in blue; advisor-led steps in gray

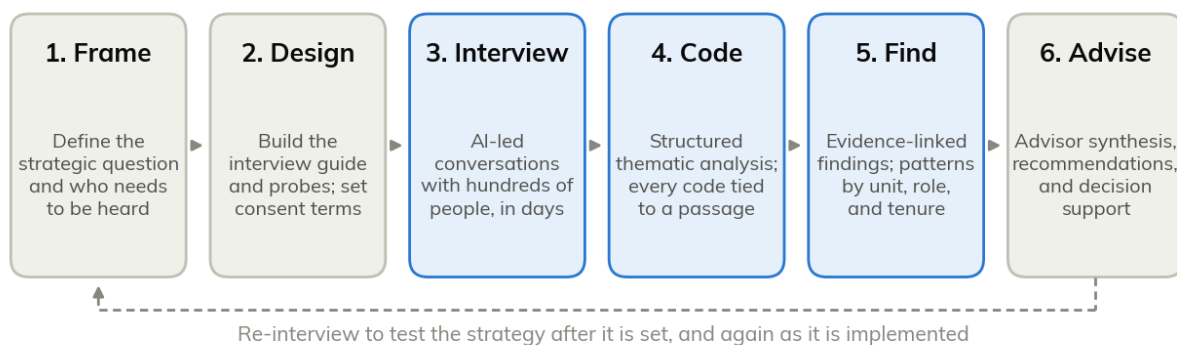


Figure 3. The Vega process for a strategic advising engagement.

Frame. The advisor works with the sponsor to state the strategic question in a form that people across the organization can speak to. "Should we enter the mid-market" is a board question; "what happens today when a mid-market prospect calls" is an interview question, and the answers to the second inform the first. The framing step also settles who needs to be heard: which units, levels, and sites, and in what numbers to allow segment-level comparisons. The advisor's strategy expertise matters most at this step; a poorly framed study produces well-coded answers to the wrong question.

Design. The advisor and the Agency 150 research team build the interview guide using Vega: an opening that establishes purpose and consent, a small number of core questions, and probes the interviewer will use when an answer warrants a follow-up. The guide is tested with a pilot group and revised. Consent terms, pseudonymity arrangements, and data retention are agreed with the client and stated to participants at this stage, not after.

Interview. Participants receive an invitation and complete the conversation at a time of their choosing, by text or voice, on their own device. Because the interviewer is software, a thousand interviews take no longer than ten; fielding is measured in days. Participation is monitored by segment so that under-represented groups can be re-invited before the window closes. Interview quality is monitored as well: the platform tracks length, depth of follow-up, and signs of disengagement, and flags sessions that fall below standard.

Code. Transcripts are coded against a science-based codebook matched to the study design and extended inductively as new themes appear. Language models apply the initial codes; analysts review a sample, resolve disagreements, and refine the codebook, and the models re-code with the refined version. Every code is tied to a passage, and every override is logged.

Find. Themes are assembled from coded passages and examined by segment: which units, roles, tenures, and sites raise a theme, with what frequency and in what terms. Findings that meet the evidence standard are promoted; those that do not are held or discarded. The output is a set of findings, each linked to its underlying passages, along with a distributional picture showing where the finding applies.

Advise. The advisor synthesizes the findings into a strategic point of view and brings it to the leadership team, with the evidence available in the room. This is the step a platform cannot replace. The advisor knows what the findings mean for the strategic question, which of them the leadership team will resist and why, and what the organization is realistically able to do about them. The advisor also closes the loop with participants: what was heard, what will change, and what will not and why. The action gap that undermines survey programs is as much a communication failure as a decision failure, and the process treats it as part of the engagement.

The loop then repeats. Because fielding is fast and the ontology is persistent, a strategy can be tested with the organization after it is drafted and again as it is implemented, at a fraction of the cost of the first study. This turns strategic listening from an annual event into an instrument leadership can use whenever it faces a decision that depends on what its people know.

5. Comparing the Three Approaches

Table 1 summarizes how the three approaches differ on the dimensions that matter for strategic advising.

Table 1. Surveys, consulting engagements, and Vega compared on the dimensions that matter for strategic advising.

	Employee Surveys	Consulting	Vega AI-led interviews
People heard	Hundreds to thousands	Typically dozens, mostly senior, plus workshops	Hundreds to thousands, across levels and sites
Depth per person	Minutes; closed items and short open text	An hour with a skilled interviewer	A 20-to-30-minute adaptive conversation with follow-up probes
What it discovers	Movement in measures already chosen	Reasons, from the people interviewed	Reasons, from the whole organization, mapped by segment
Time to insight	Weeks to field; typically an annual or quarterly cycle	Two months to a year	Days to field; findings in weeks; repeatable on demand
Cost structure	Low per respondent; mostly fixed	High; billed by the hour through a partner-led staffing pyramid	Moderate; low marginal cost per interview
Main bias risks	Common method variance, social desirability, straightlining, self-report on causes	Selection of interviewees; interviewer and firm priors	Model error in coding; interviewer probing limits; both inspectable via evidence links
Evidence traceability	Numbers without reasons	Selected quotes in a deck	Every finding linked to its passages, by segment
Candor conditions	Anonymous but constrained to fixed items	Depends on trust in the interviewer and sponsor	Pseudonymous, non-human interviewer; disclosure evidence favorable
What the client keeps	A dashboard and benchmark	A report; the team and notes leave	Report, advisor judgment, and a structured, queryable evidence base with a firm-specific ontology

	Employee Surveys	Consulting	Vega AI-led interviews
Repeatability	Annual or pulse, same items	Rarely repeated identically	Re-run in days; before/after comparisons on the same model
Best used for	Tracking known measures over time	External analysis, transactions, implementation, facilitation	Diagnosing, formulating, and testing strategy with the people who will execute it

A few rows need explanation. On bias, the three approaches fail in different ways, not just to different degrees. Surveys are vulnerable to the response biases cataloged by Podsakoff and colleagues and to the introspective limits Nisbett and Wilson described. Consulting interviews are vulnerable to selection (who gets interviewed) and to the interviewer's and firm's priors, which shape both what is asked and what is remembered. AI-led interviews are vulnerable to model coding errors and to the limits of a software interviewer's judgment about when to probe. The difference is that the third set of errors is inspectable and resolvable with human expertise: every code is attached to a passage and every finding to its evidence, so an error can be found and corrected in a way that a consultant's recollection of a conversation cannot.

On cost and time, the marginal figure matters more than the headline. A survey's marginal respondent is nearly free, which is its great advantage; a consulting engagement's marginal interview costs a billable hour plus the analysis; a Vega interview's marginal cost is close to the survey's, while its information content is closer to the consultant's. The same logic applies to repetition. Repeating a consulting diagnosis means a second engagement; repeating a Vega study means re-opening the instrument.

On what the client keeps, the difference is strong. After a survey, the client has a dashboard. After an engagement, the client has a report and, if the relationship was good, some transferred judgment. After a Vega engagement, the client has the report, the advisor's judgment, and a structured, queryable body of the organization's own testimony, coded against a model of the organization that improves with each study.

6. What Vega is Not

Vega has three addressable limitations.

Vega hears what people say, not what they do. A consultant on a site visit sees the workaround that no one mentioned. Interview data should be triangulated with operational data and, where the question warrants it, with observation. Vega provides the testimony. The other evidence is still valuable.

Vega does not negotiate, facilitate, or implement. Much of what a consulting team does after the diagnosis, from building the financial model to running the steering committee to sitting with the plant manager while the new process is installed, is work Vega does not do and does not claim to do. Where a client needs that, a consulting engagement is the complementary purchase to Vega.

Vega depends on participation and on candor. The disclosure evidence is encouraging, but an organization in which people do not believe the pseudonymity assurance, or in which leadership has a history of ignoring what it hears, will get thinner interviews. Part of the advisor's job in the framing step is to judge whether the conditions for candor exist and, if not, what has to be said and done before the study opens.

7. Choosing and Combining Approaches

The right question is not which tool is best but which tool fits the job and how they can serve complementary roles.

Use a consulting engagement when the work is analysis or implementation rather than listening: market entry analysis that depends on external data, transaction support, operating-model redesign, a program office to drive execution. The client pays for the engagement model's depth and its capacity to do work on the client's behalf.

Use Vega when the strategic question depends on what the organization's own people know and have not been able to say: diagnosing why a strategy is stalling, testing a proposed direction before it is announced, understanding a merger partner's organization before integration, locating where a culture problem actually lives, harvesting front-line ideas for growth or cost, or checking, six months in,

whether the strategy the board approved is the strategy the organization is executing.

The approaches also combine, and the combinations are often the best design. A survey that shows engagement falling in one region is a reason to open a Vega study in that region to find out why. A Vega study that surfaces a structural problem in the operating model is a reason to bring in a consulting team to redesign it, with the study's evidence as the brief. And a consulting engagement that produces a new strategy is a reason to run a Vega study before launch, to learn how the organization will receive it and what it will need in order to carry it out.

8. Design Requirements for Trustworthy Listening

Because Vega asks people to be candid, the platform and the engagement have to earn that candor. Four requirements are non-negotiable in Agency 150's practice, and a client evaluating any listening approach, including a competitor's, should ask about each.

Transparency. Participants are told, before the conversation begins, what the study is for, who will see what, at what level of aggregation, and for how long the data will be kept. Transparency is specific to the study; a later use requires a later consent.

Pseudonymity. Where designated and stated to participants, identity is separated from the interview record and held in a separate system. Findings are reported at aggregation levels that do not identify individuals, and small segments are suppressed or merged rather than reported. No client receives a transcript with a name on it unless this is explicitly predetermined.

Evidence links. Every finding presented to a client is traceable to the passages that support it. The client can open the evidence. A finding that cannot be traced is not a finding.

Retention terms. The client can decide to delete data after a study upon request, but retention allows later studies to build on earlier ones and makes the firm-specific ontology cumulative; deletion is available for clients whose obligations require it. Cross-client use of any client's material, even in aggregate, happens only under an explicit agreement.

9. Conclusion

For as long as organizations have had strategies, the knowledge those strategies needed has been spread across people who were rarely asked and, when asked, rarely able to answer properly. Surveys reached everyone and heard little; consultants heard a great deal from a few. The research summarized here shows that a machine interviewer can now conduct a conversation that people complete, prefer, and speak candidly in, at a scale that makes it possible to hear an entire organization explain itself, and that the resulting transcripts can be coded more consistently than human teams have managed. Vega packages that capability with the two things the science does not supply on its own: a data model that forces every finding back to its evidence, and an advisor who knows what the findings mean. The result is better-informed judgment, not a substitute for it.

References

- Anseel, F., Lievens, F., Schollaert, E., & Choragwicka, B. (2010). Response rates in organizational science, 1995–2008: A meta-analytic review and guidelines for survey researchers. *Journal of Business and Psychology*, 25(3), 335–349. <https://doi.org/10.1007/s10869-010-9157-6>
- Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human Relations*, 61(8), 1139–1160. <https://doi.org/10.1177/0018726708094863>
- Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., & Roberts, J. (2013). Does management matter? Evidence from India. *Quarterly Journal of Economics*, 128(1), 1–51.
- Bower, J. L. (1970). *Managing the resource allocation process: A study of corporate planning and investment*. Division of Research, Graduate School of Business Administration, Harvard University.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Bruhn, M., Karlan, D., & Schoar, A. (2018). The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in Mexico. *Journal of Political Economy*, 126(2), 635–687. <https://doi.org/10.1086/696154>

- Burgelman, R. A. (1983). A process model of internal corporate venturing in the diversified major firm. *Administrative Science Quarterly*, 28(2), 223–244.
- Chopra, F., & Haaland, I. (2023). *Conducting qualitative interviews with AI* (CEBI Working Paper 06/23; CESifo Working Paper 10666). University of Copenhagen. <https://papers.ssrn.com/abstract=4583756>
- Christensen, C. M., Wang, D., & van Bever, D. (2013, October). Consulting on the cusp of disruption. *Harvard Business Review*.
- Detert, J. R., & Edmondson, A. C. (2011). Implicit voice theories: Taken-for-granted rules of self-censorship at work. *Academy of Management Journal*, 54(3), 461–488. <https://doi.org/10.5465/amj.2011.61967925>
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350–383. <https://doi.org/10.2307/2666999>
- Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, 73(2), 349–360. <https://doi.org/10.1093/poq/nfp031>
- Gallup. (2025). *State of the global workplace: 2025 report*. Gallup, Inc.
- Geiecke, F., & Jaravel, X. (2024). *Conversations at scale: Robust AI-led interviews with a simple open-source platform* (CEPR Discussion Paper 19705). Centre for Economic Policy Research.
- Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30), e2305016120. <https://doi.org/10.1073/pnas.2305016120>
- Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1), 59–82. <https://doi.org/10.1177/1525822X05279903>
- Hayek, F. A. (1945). The use of knowledge in society. *American Economic Review*, 35(4), 519–530.
- Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine*, 292, 114523. <https://doi.org/10.1016/j.socscimed.2021.114523>
- Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, 101(46), 16385–16389. <https://doi.org/10.1073/pnas.0403723101>

- Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37, 94–100. <https://doi.org/10.1016/j.chb.2014.04.043>
- Macnamara, J. (2016). *Organizational listening: The missing essential in public communication*. Peter Lang.
- Maister, D. H. (1993). *Managing the professional service firm*. Free Press.
- Mazzucato, M., & Collington, R. (2023). *The big con: How the consulting industry weakens our businesses, infantilizes our governments and warps our economies*. Penguin.
- Milliken, F. J., Morrison, E. W., & Hewlin, P. F. (2003). An exploratory study of employee silence: Issues that employees don't communicate upward and why. *Journal of Management Studies*, 40(6), 1453–1476. <https://doi.org/10.1111/1467-6486.00387>
- Mintzberg, H. (1978). Patterns in strategy formation. *Management Science*, 24(9), 934–948. <https://doi.org/10.1287/mnsc.24.9.934>
- Mintzberg, H., & Waters, J. A. (1985). Of strategies, deliberate and emergent. *Strategic Management Journal*, 6(3), 257–272. <https://doi.org/10.1002/smj.4250060306>
- Morrison, E. W. (2011). Employee voice behavior: Integration and directions for future research. *Academy of Management Annals*, 5(1), 373–412. <https://doi.org/10.1080/19416520.2011.574506>
- Morrison, E. W. (2014). Employee voice and silence. *Annual Review of Organizational Psychology and Organizational Behavior*, 1, 173–197. <https://doi.org/10.1146/annurev-orgpsych-031413-091328>
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259. <https://doi.org/10.1037/0033-295X.84.3.231>
- Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*, 5(1), 14–37. <https://doi.org/10.1287/orsc.5.1.14>
- Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford University Press.
- O'Mahoney, J., & Sturdy, A. (2016). Power and the diffusion of management ideas: The case of McKinsey & Co. *Management Learning*, 47(3), 247–265. <https://doi.org/10.1177/1350507615591756>

- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). Academic Press.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Polanyi, M. (1966). *The tacit dimension*. Doubleday.
- Smyth, J. D., Dillman, D. A., Christian, L. M., & McBride, M. (2009). Open-ended questions in web surveys: Can increasing the size of answer boxes and providing extra verbal instructions improve response quality? *Public Opinion Quarterly*, 73(2), 325–337. <https://doi.org/10.1093/poq/nfp029>
- Sull, D., Homkes, R., & Sull, C. (2015, March). Why strategy execution unravels—and what to do about it. *Harvard Business Review*, 93, 57–66.
- Törnberg, P. (2025). Large language models outperform expert coders and supervised classifiers at annotating political social media messages. *Social Science Computer Review*, 43(6), 1181–1195. <https://doi.org/10.1177/08944393241286471>
- Xiao, Z., Zhou, M. X., Liao, Q. V., Mark, G., Chi, C., Chen, W., & Yang, H. (2020). Tell me about yourself: Using an AI-powered chatbot to conduct conversational surveys with open-ended questions. *ACM Transactions on Computer-Human Interaction*, 27(3), Article 15. <https://doi.org/10.1145/3381804>
- Zielinski, D. (2022, April 22). *Worker mistrust leads to flawed data in engagement surveys*. SHRM. <https://www.shrm.org/topics-tools/news/technology/worker-mistrust-leads-to-flawed-data-engagement-surveys>