EXECUTIVE SEARCH'S NEXT FRONTIER

How Thorsten-4, millionways' Large Psychology Model, helps search firms understand the people behind leadership decisions.

Sharper hiring.
Stronger leadership.
More advisory revenue.

Beyond the résumé.
Insight clients pay for.

Hunt Scanlon | Media
Leadership Intelligence

millionways

Inside this report....

01 The Partners behind the report

02 Why millionways?

03 Executive Summary

04 What do we still need to know?

05 What can millionways help us understand better?

06 What does Thorsten add to an existing assessment process or capable AI system?

07 Why is millionways different from other assessment tools?

08 The person, the role, and the business: why context matters

09 Where can millionways improve a business decision and its impact on the bottom line?

10 What evidence shows that it makes a real difference?

11 How can an organization test it and gain from it?

12 Appendix

The Partners behind the report

millionways

millionways is a frontier lab for behavioral intelligence, building Thorsten, its proprietary Large Psychology Model. Grounded in psychological research and specialist labeled data, the model analyzes language to produce structured insights into motivation, cognition and emotional regulation. These insights support professional judgment and provide behavioral context to connected AI systems. This report examines their application to executive search and leadership. The model's research foundation comprises roughly 7,000 structured interviews conducted by trained psychologists, with participants interviewed repeatedly over several years. More than 35 million words were hand-tagged using published psychological research methods, and predictions were checked against participants' subsequent behavior. This combination of specialist labeling, repeated observation and comparison with actual outcomes provides the basis for confidence in the model's findings. Thorsten analyzes written text and transcribed speech for behavioral and emotional signals expressed through language. It does not analyze vocal characteristics such as tone, pitch or cadence. In executive search, its insights can help consultants develop more focused interview questions, examine leadership risks and assess how a candidate's behavioral patterns may fit the role and business context. The company brings together expertise in psychology, AI and commercial applications. Its individual investors and advisers have prior backgrounds at organizations such as Apple, Meta, Kantar and Two Sigma.

Max Weidemann, Co-Founder, max@millionways.ai

Hunt Scanlon | Media
Leadership Intelligence

Hunt Scanlon's mission is to inform, engage and connect the talent management community on the issues shaping talent acquisition, development and retention. Its distinctive strength lies in bringing industry intelligence, relationships, practical learning, and transaction expertise together, with a broad industry perspective. Through its reporting, research, conferences, and events, it connects leaders and practitioners to exchange ideas, examine emerging challenges and explore innovations changing how organizations attract, develop, and retain their people. These activities create opportunities for conversations, partnerships and decisions about the future of their businesses. It also serves as a center for training and development, offering courses that help executive search and talent professionals sharpen their skills, strengthen client relationships and improve business performance. Its training connects practical experience with approaches professionals can apply to winning assignments, assessing candidates, advising clients and building stronger firms. Through Hunt Scanlon Ventures, it provides mergers and acquisitions advisory services, helping businesses pursue growth, identify strategic partners and navigate ownership transitions. This work brings an additional perspective on how firms build value, prepare for their next stage, and respond to rapidly changing markets. Together, these activities give Hunt Scanlon a distinctive view of the industry and a practical role in shaping its future, helping its community turn knowledge and connections into lasting commercial value.

Scott A. Scanlon, CEO and Co-Founder, scott@huntscanlon.com

WHY MILLIONWAYS

millionways.ai: Behavioral intelligence for the decisions that shape leadership.

Clients expect more from executive search than an impressive shortlist. They need advice on whom to hire, why that person fits their business and where the risks lie. That is where search firms earn their fees and strengthen their value.

By using millionways, search firms apply a broader scientific capability in behavioral intelligence to their own work. Thorsten-4 analyzes conversations firms already hold, including interviews, referencing calls and client briefings. It adds structured behavioral insight without adding a test for candidates to complete.

These insights can strengthen candidate assessment and extend a firm's advice into management reviews, onboarding, leadership development, and succession. The value should be demonstrated through measurable improvements in hiring outcomes, efficiency, and leadership performance.



Five Business Outcomes to Target and Measure

01 **Reduce early turnover and improve retention.**
Identify potential mismatches between the candidate, role and business context, then track early departures and retention at agreed intervals.

02 **Lower the cost of unsuccessful hires and replacement searches.**
Surface leadership risks before appointment and measure changes in unsuccessful placements, replacement fees, and associated client costs.

03 **Save consultant and candidate time.**
Use existing conversations more effectively to reduce duplicated assessment steps, tracking hours spent and time to a hiring decision.

04 **Reduce assessment costs while maintaining scientific standards.**
Measure cost per assessment alongside validity, reliability, and predictive quality against established scientific benchmarks.

05 **Improve leadership performance and shorten time to productivity.**
Use behavioral insights to focus onboarding and development, then assess progress against agreed performance targets and milestones.

millionways gives search firms a way to deepen the evidence behind their recommendations and build lasting advisory relationships with value measured beyond the placement.

Executive Summary

Observations from Hunt Scanlon Leadership

Few decisions have greater consequences for a business than choosing who will lead it. Yet even after examining a candidate's career, achievements and references, we may know more about who they were than who they are today. Past success matters, but it does not fully tell us what motivates someone, how they currently think or how they will respond under pressure in a new role. That understanding often emerges only after they have joined, by which time a poor fit may already be affecting the team and the business.

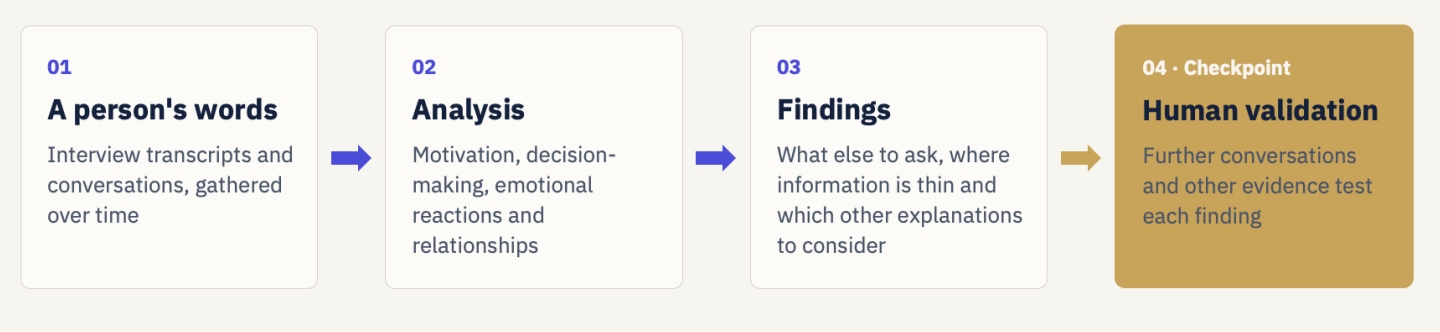
A hire who looked convincing on paper can quickly prove disappointing or fail outright. By then, the business has done more than agree to pay a salary; it has entrusted that person with its people, its ambitions, and a significant part of its future.

In this report we examine how behavioral science can help us better understand an executive before those differences become costly. At the heart of this report lies a simple question: what can people's own words tell us about their motivations, how they think, handle disagreement and earn the trust of those they lead? This gives us a fuller picture beyond career history and a stronger basis for the strategic advice clients pay for, supported by evidence they can examine and discuss.

Through our partnership with millionways, we have explored how behavioral intelligence can strengthen executive assessment and the advice search firms provide. This report examines the technology, the research supporting it and how its findings can be tested in practice. The report explains how millionways turns conversations into structured behavioral insight. Thorsten-4 analyzes the language in conversations a firm already holds, including interviews and referencing calls, and measures it against 50 defined behavioral and psychological traits. Each finding traces back to the passages that support it, allowing consultants to examine the evidence and question the interpretation. These findings can inform an assessment, guide further investigation or provide behavioral context to an AI application. Their interpretation must always take account of the person, the role and the available evidence, with further conversations and independent information used to test conclusions before a final decision.

This raises a natural question: why not simply give ChatGPT an interview transcript and ask what it reveals about someone's motivations, judgment and likely behavior? To answer it, we tell the story of why millionways was created: its founders' desire to understand why people do what they do and what those differences mean for the people around them. We examine the psychology and research behind the technology, how assessing motivation and behavior differs from interpreting a conversation, and what evidence supports that distinction.

We compare millionways with established assessment tools and newer AI products that analyze emotion or sentiment. The report examines how consistently it produces its findings, how the technology was tested, where it made mistakes and what evidence connects its findings to behavior or business results. It also considers how well it works across languages, cultures and different ways of communicating. millionways analyzes words and meaning in written text and transcribed speech; it does not analyze vocal characteristics such as tone, pitch or cadence. A polished answer may tell us very little about an individual, while someone who hesitates or speaks sparingly may have something valuable to say.



Circumstances matter too. An executive may work well with one CEO but struggle with another, or lead successfully when things are running smoothly but find an acquisition or restructuring much more difficult. Stress, tiredness and grief can affect how someone speaks and reacts, which is why patterns across conversations matter more than one difficult day. Coaching and rehearsed answers also deserve attention: knowing precisely what an interviewer wants to hear can create confidence without revealing how someone will behave in the role.

These questions have practical value across hiring, investment diligence, team development, onboarding, coaching and succession. For private equity investors, they may help establish whether management can deliver the value-creation plan; for corporations and boards, whether a candidate can meet the demands of a role. For search firms, they can strengthen recommendations and the strategic advice offered during a retained search and after the hire.

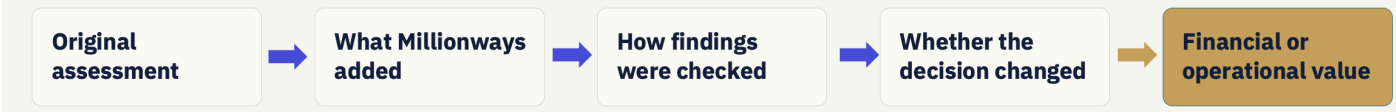
As AI reshapes roles and the way work gets done, businesses also need to understand how their people will learn, adapt and work together, alongside the skills they will require. A deeper understanding of behavior can help them decide whom to hire, where to invest in development and who is ready for greater responsibility.

Research, comparisons and anonymized case studies examine what the original assessment revealed, what millionways added, how its findings were checked and whether they changed the decision. Each example connects insight to a specific action and its potential financial or operational value. Where results are available, we consider leadership effectiveness, retention, team performance, and business goals, distinguishing useful insight from evidence of better results. We also address findings that disagree with interviews, references or past achievements, and the time, cost and practical work involved.

The report closes with a guide to testing millionways on a real business challenge: which decision to focus on, what information to provide, who should be involved and how to judge the trial's value. Information quality matters: a full transcript, an interviewer's notes and an edited account can all give different pictures of the same person. What seems like enough information for an analysis may still leave us wondering about the person behind it. The guidance offered here also covers confidentiality, how individuals can challenge findings or provide context, and an agreed pilot with measures to assess value over time.

The science, evidence and practical examples presented throughout make a compelling case for millionways as a potentially valuable tool for better understanding the people organizations hire, promote and trust with their future. Its value lies in bringing important questions to the surface while there is still time to strengthen a recommendation, reconsider a promotion, support a leader or address a concern before it becomes a costly mistake. The methodology and supporting research in the appendix give organizations a clear basis for testing that value for themselves.

For executive search firms, millionways can strengthen both the evidence behind a candidate recommendation and the relationship with the client. It helps search consultants provide the critical strategic advice clients pay for: a fuller understanding of the person they are hiring, the questions still to be answered and the concerns to address before they become costly mistakes.



Insight is not the same as results

Where outcomes are available, we look at leadership effectiveness, retention, team performance and business goals, and keep useful insight separate from evidence of better results.

When the evidence disagrees

We set out how to proceed when findings conflict with interviews, references or past achievements, and the time, cost and practical work involved.

What do we still need to know?

By the final stages of a search, we may know a great deal about an executive. We have examined their career, challenged their achievements and taken extensive references from people who have worked alongside them. Existing assessments may describe their personality, strengths and potential risks. The file is substantial and the recommendation feels well supported. Yet an important question may remain and that is how will this person lead when circumstances and expectations change?

What the standard search file covers, and what it leaves open



Each part of the search process contributes something important. The résumé establishes career history and responsibility. Achievements provide evidence of delivery, although we must distinguish the individual's contribution from the advantages of the business, team and market. Interviews explore judgment, motivation and the decisions behind results. References offer the perspective of those who have seen the executive at work.

Properly validated assessments provide a consistent way to examine characteristics that are otherwise difficult to compare. Together, these sources give us a serious basis for a decision. Even so, a substantial file may tell us more about what someone achieved than how they secured the trust and commitment of the people who helped them achieve it.

Two executives may have held comparable roles, delivered similar results and described themselves as ambitious, decisive and collaborative. Yet one may be motivated primarily by solving difficult problems, while another places greater importance on recognition, influence or relationships. Those motives can coexist, but their relative importance may affect how each person exercises authority, responds to challenge and shares responsibility.

The demands of the business matter equally. A leader effective in an established organization may face different challenges during an acquisition or restructuring. Research on fit at work distinguishes the person's relationship with the job, organization, team and supervisor, each a relevant consideration when evaluating an appointment.[1]

Structured interviews provide a strong basis for assessment.[2] Candidates also prepare to make a favorable impression, and research on impression management reinforces the importance of distinguishing interview performance from evidence of job performance.[3] The opportunity is to examine the available material more systematically and connect it to the demands of the appointment.

1 Kristof-Brown, Zimmerman and Johnson, Consequences of Individuals' Fit at Work: A Meta-Analysis of Person–Job, Person–Organization, Person–Group, and Person–Supervisor Fit.
2 Research on the validity of personnel selection methods.
3 Impression Management and Interview and Job Performance Ratings: A Meta-Analysis of Research Design with Tactics in Mind.

What can millionways help us understand better?

millionways applies a specialist psychological model to language generated through conversations a firm already holds. Thorsten-4 examines written text and transcribed speech to produce structured findings about behavioral and psychological characteristics. The company describes its areas of analysis as including motivation, cognitive patterns, emotional regulation, decision styles and communication dynamics.[4]

For executive search, the distinct contribution is the measurement layer between the conversation and the consultant's conclusion. A transcript records what the candidate said. Thorsten-4 measures that language against defined traits and links findings to supporting passages. The resulting interpretation can then be examined against the person's history, the requirements of the role and other available evidence.

01 — What the candidate explicitly said

The transcript passage, preserved in context, without adding a conclusion about the person.

02 — What Thorsten-4 measured

The relevant trait reading and the passage or passages to which it traces.

03 — What the interpretation suggests

A possible explanation of the finding and its relevance to the appointment, including competing explanations.

04 — What the consultant should investigate

The evidence needed to support, qualify or challenge that interpretation.

Keeping these stages separate makes the analysis easier to examine. A candidate's statement is evidence of what they reported. A model reading is an analytical result. An interpretation explains what that result might mean. The consultant's recommendation requires judgment across all the available evidence.

Motivation: what sustains the person's effort?

Describing someone as ambitious does not establish what makes achievement meaningful to them. Solving a complex problem, gaining influence, receiving recognition and maintaining strong relationships can all motivate effort, with different implications for a particular role. Personality Systems Interactions theory provides a framework for considering how motivation, emotion and cognition interact.[5] Applied to assessment, this encourages a more specific examination of the person's motivational patterns rather than relying on broad adjectives. A leader motivated by solving difficult problems may bring considerable energy to a turnaround. The appointment may also require sustained attention to developing others once the immediate challenge has passed. The relevant question is how the motivational profile supports the whole role, including its less immediately rewarding responsibilities.

Decision-making: how does thinking translate into action?

An account of a successful decision establishes the outcome more readily than the process behind it. The assessment needs to distinguish how the executive gathered information, weighed alternatives, used advice and committed to action. Thorsten-4's forensic focus on decision styles and cognitive patterns provides a way to structure this examination.[6] The useful output is a trait finding tied to language that can be reviewed, rather than a general description such as "decisive" or "analytical." Its interpretation should consider whether the apparent approach was appropriate to the circumstances and whether it recurs elsewhere in the evidence.

4 millionways, Thorsten-4. https://www.millionways.ai/thorsten
5 Kuhl, Quirin and Koole, The Functional Architecture of Human Motivation: Personality Systems Interactions Theory.
6 millionways, Thorsten-4. https://www.millionways.ai/thorsten

Emotional regulation and relationships: what changes under strain?

Emotional regulation concerns how people manage reactions that can affect thinking and behavior. PSI theory examines these interactions; knowing how one should respond does not necessarily mean being able to do so when circumstances become difficult.[7]millionways identifies emotional regulation, trust and collaboration among its areas of behavioral analysis.[8] Hunt Scanlon's interview with co-founder and CTO Max Weidemann also discusses their relevance to leadership assessment.[9] The value lies in examining whether related findings recur across conversations, what situations accompany them and what evidence qualifies them. A single difficult exchange should not define a candidate. Nor should a model finding be treated as a diagnosis or a certain prediction of future behavior. millionways describes its outputs as probabilistic and intended to support human judgment.[10]

Interpreting findings against the actual job

A behavioral finding has limited practical value until it is connected to the demands of the appointment. Taking direct control may be necessary during a crisis; the same approach may become an obstacle when the task is to develop an autonomous leadership team. The consultant and client should therefore establish the conditions in which the executive must succeed: the mandate, decision authority, CEO relationship, team capability, available resources and expected pace of change. Findings can then be considered in terms of where the person's approach may help, where friction could arise and what support may be useful. millionways identifies hiring, leadership development and team effectiveness among its intended applications.[11]

This is also where the analysis can contribute beyond selection. Findings relevant to the appointment may inform onboarding priorities, working arrangements with the CEO and specific development goals. Their usefulness should be assessed through subsequent evidence.

Showing and checking the evidence

A useful assessment should let the consultant move from a finding back to the language supporting it, consider alternative explanations and identify what remains unknown.

The following example shows the distinction between the candidate's account, a trait reading, an interpretation and further investigation. It is illustrative: neither the transcript nor the reading is an actual candidate assessment. The trait wording is descriptive, not a confirmed Thorsten-4 label.

Success in one business does not guarantee the next

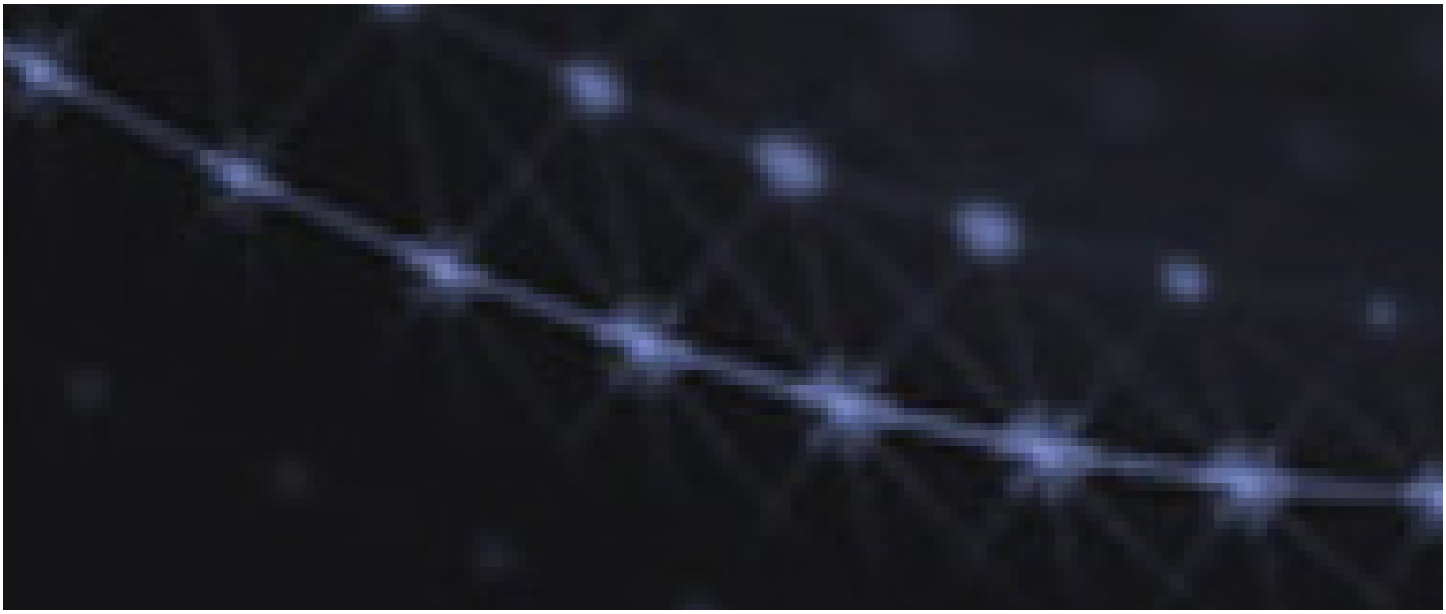


7 Kuhl, Quirin and Koole, The Functional Architecture of Human Motivation: Personality Systems Interactions Theory.
8 millionways, Thorsten-4. https://www.millionways.ai/thorsten
9 Hunt Scanlon, interview with millionways co-founder and CTO Max Weidemann.
10 millionways, Science. https://www.millionways.ai/science
11 millionways, Workforce & Leadership.https://www.millionways.ai/workforce-leadership

Part of the assessment	Illustrative example
What the candidate explicitly said	"The team had missed a deadline, so I made the decisions myself for the next six weeks. We couldn't afford to miss another one."
Illustrative trait reading beside its supporting passage	Tendency toward direct control when delivery is threatened. Supporting passage: "I made the decisions myself for the next six weeks."
What the interpretation suggests	The executive may respond to delivery risk by concentrating decision authority. This could represent a proportionate intervention or a recurring difficulty delegating. The passage alone cannot establish which explanation is stronger.
Evidence that qualifies the interpretation	"Once things were back on track, the team made their own decisions again. They handled the next project without my help." This account suggests that control was temporary, although it still requires corroboration.
Why it matters for the role	A turnaround may require decisive intervention. A role focused on developing independent leaders also requires evidence that the executive restores authority and strengthens the team's capability.
What the consultant should investigate	What caused the missed deadline? What support or corrective action preceded the intervention? Which decisions were taken over? Was responsibility genuinely restored, and did performance improve afterwards?
How to check it	Examine another comparable episode and seek a reference from someone involved. Establish whether the intervention improved delivery, how the team experienced it and whether members subsequently operated independently.

Thorsten-4 adds a distinct layer of analysis by linking defined behavioral trait readings to the language that supports them. The consultant can examine each finding, compare it with other evidence, and assess its relevance to the role. This does more than just guide further questioning but also provides a structured behavioral assessment with a clear basis for reviewing both the findings and their interpretation.

The distinction between action and an interpretation is important.



What happened is not the same as what it means



"I took over the project" describes what happened. "This person finds it difficult to trust their team" is an interpretation requiring more evidence. The consultant needs to understand why the comments prompted that interpretation, what remains uncertain and what would justify changing it.

NIST's guidance on explainable AI supports the importance of understandable explanations and clear limits on what a system can conclude.[12] In practice, the search consultant needs to see what the candidate said, why it raised a concern and what further evidence is needed before discussing that concern with the client.

Careful examination ideally protects against unfairly ruling out a capable candidate or overlooking a real concern. millionways' findings should remain matters to investigate, with the consultant ultimately responsible for examining the evidence and reaching the recommendation.[13] This gives the client a clearer explanation of the candidate's strengths, the concerns still being explored and the reasons behind the advice.

Using the understanding after a hire

The value that millionways brings to every search mandate continues well after the hire. Findings explored and confirmed with the executive can often help the client address difficulties early on, strengthen working relationships and prepare the person for more responsibility. millionways identifies coaching, leadership development, team effectiveness and succession planning as applications of its analysis.[14]

This matters because a successful hire depends on more than choosing the right person. Helping an executive settle in, earning the team's trust and addressing difficulties early on can protect the client's investment in the hire. For the search firm, it creates a reason to stay involved, deepen the client relationship and provide advice beyond the successful completion of the search.

Understanding what motivates an executive helps the client decide how to develop and retain them. It also helps distinguish strong performance in their current job from readiness for promotion, so the client can address gaps before giving them greater responsibility.

For search firms, this understanding strengthens the recommendation and the relationship with the client. It gives consultants specific evidence to investigate, clearer advice to offer and practical ways to support the executive after the hire. This is the strategic advice clients pay for and a stronger reason to keep the search firm involved before an overlooked concern becomes a costly leadership mistake.

"Understanding what motivates an executive helps the client decide how to develop and retain them."

12 https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8312.pdf
13 https://www.millionways.ai/terms-of-use
14 https://www.millionways.ai/workforce-leadership

What does Thorsten add to an existing assessment process or capable AI system?

A specialist behavioral intelligence layer

Language models can make a useful contribution to psychological assessment. They can organize interview material, identify possible patterns, compare accounts and suggest interpretations. Research also shows how language models can be developed and evaluated specifically for psychological assessment.[15] The relevant question is what additional value Thorsten-4 brings to that work.

For a search consultant, that value should be visible in the quality of the assessment: what the system measures, how clearly its findings connect to evidence, whether they withstand scrutiny and whether they improve the advice given to the client. The practical comparison is whether access to Thorsten's structured psychological findings improves an assessment or interaction beyond the existing approach. This can be evaluated within a consultant's workflow or an AI application, using the same underlying material and clearly defined measures.

Different foundations, complementary capabilities

Language models learn patterns in language through training that includes predicting the next token in a sequence. Their broad training material includes publicly available internet content alongside other sources. This develops versatile capabilities, but accounts of human behavior do not necessarily provide verified explanations of why people acted or records of what they subsequently did.

Thorsten-4 was developed around a different objective: modeling behavioral patterns using specialist psychological research. millionways describes a research foundation involving psychologist-labeled material, repeated engagement with participants and checks against observed behavior.[16] The distinction is therefore between a broadly capable language system and a specialist source of structured behavioral findings. These capabilities can work together. An existing AI application can organize information and communicate an assessment while using Thorsten's findings as an additional source of behavioral evidence.

The research foundation provides a reason to examine that contribution. Its practical value still needs to be demonstrated for the task, population and decision involved.

SAME TRANSCRIPT, TWO DIFFERENT JOBS

Interpreting the words vs. assessing the person

GENERAL-PURPOSE LLM

Offers a reading of what was said

- Organizes the interview material
- Suggests possible characteristics
- Proposes follow-up questions

Output: a convincing explanation

MILLIONWAYS · THORSTEN-4

Assesses what drives the person

- Built on defined psychological concepts
- Trained on material labeled by specialists
- Findings linked to the candidate's own words

Output: an evidence-backed finding

The distinction is the evidence supporting the interpretation.

Adding Thorsten to systems companies already use

Companies can add Thorsten-4 to the models, agents and AI products they already run through the millionways Platform, an API, MCP or a self-hosted deployment.[17] This allows behavioral intelligence to become part of an established workflow.

15 Brickman, Gupta and Oltmanns — Large Language Models for Psychological Assessment: A Comprehensive Overview (2025).
16 millionways — Science: research foundation.
17 millionways — Thorsten-4: capabilities and deployment options.

For example, an assessment application could use a language model to organize interview and reference material, draw on Thorsten's structured findings, and present the combined evidence for a consultant's review. The consultant would retain responsibility for interpreting the findings alongside the candidate's history, the role, and the available evidence.

The same principle applies beyond executive search. A customer-facing AI application could draw on behavioral findings when interpreting an interaction or selecting a response. In each setting, the question is whether the added information improves performance.

Two contributions that matter when a finding must be defended

A search consultant presenting a recommendation to a board or investor needs to explain both the conclusion and its basis. Two characteristics are particularly important.

Explainability: the finding traces to its evidence.

A structured finding should identify the language supporting it and make the interpretation open to examination. The consultant can inspect the relevant passages, consider qualifying evidence and explain why the finding matters for the appointment. millionways describes traceability and interpretability as central features of its approach.[18]

Consistency under challenge: the evidence remains the reference point.

A finding should not change merely because someone asks for a more favorable answer or challenges it forcefully. Where the underlying evidence remains unchanged, the assessment should remain stable. Where new evidence warrants revision, the reason for that revision should be clear.

This consistency is an important requirement to test. Repeated analysis, alternative prompts, and deliberate challenges can help establish whether the result is reproducible and whether changes reflect new information rather than the wording of the request. Consistency alone does not establish accuracy, but it matters when consultants must defend an assessment.

How to test the additional value

OpenAI's evaluation guidance recommends testing applications against the tasks they are expected to perform, using relevant examples and human judgment.[19] The same principle applies when adding Thorsten to an assessment process.

A useful comparison would examine the existing approach alongside the same approach supplemented by Thorsten's findings. Both should receive the same underlying material, with evaluation criteria agreed in advance. For executive search, those criteria could include the accuracy of trait findings against an appropriate benchmark, traceability to supporting passages, consistency under challenge, recognition of insufficient evidence and the usefulness of the resulting assessment to independent reviewers. Time and cost should also be measured.

An anonymized exercise involving four executive interviews offers an illustration of the intended application. millionways' analyses identified differences in motivation, decision-making and potential behavioral fit, providing material for consultants to examine. The appendix contains those analyses.

The exercise illustrates what the output can contribute. Establishing whether it improves assessment requires a comparison with the existing process and an independent basis for judging the results.

A human question came first

millionways began with a research effort to understand why people behave differently as circumstances change. According to its founders, its early work as a nonprofit in Germany involved more than 7,000 participants and collaboration with over 50 psychologists.[20]

The model's name honors Thorsten, an original co-founder who helped coordinate that research. He died unexpectedly shortly before the work became a production AI system.[21]

This history explains the company's focus: bringing specialist behavioral research into practical decisions. For search firms, the opportunity is to apply that capability to conversations they already conduct and connect the resulting findings to the appointment under consideration.[22]

18 Science: specialist research, contextual interpretation and explainability.
19 OpenAI — Evaluation best practices.
20 millionways — Company history and the naming of Thorsten.
21 millionways — Company history and the naming of Thorsten.
22 millionways — Science: research applications.

What does the research establish?

Three questions require separate evidence: is the underlying psychology supported by research, does the model measure the relevant characteristics reliably, and do its findings improve understanding or prediction of behavior? Evidence for one does not automatically establish the others.

Brickman, Gupta and Oltmanns' overview recognizes the potential of language models for psychological assessment while explaining the importance of model development, evaluation and validity.[23] A system's ability to produce a plausible description is insufficient; researchers must establish what its outputs actually measure.

Research cited by millionways also illustrates how automated measurement can be examined. Scheffer and colleagues studied the use of natural language processing and machine learning with the Operant Motive Test, comparing automated analysis with expert judgments and examining relationships with managerial responsibility.[24]

That study supports the investigation of automated motive assessment in a defined setting. It does not, by itself, establish the validity of every Thorsten-4 finding or demonstrate improved executive hiring outcomes. Those claims require evidence specific to the model and application being assessed. The appendix's validation material should be read with that distinction in mind.

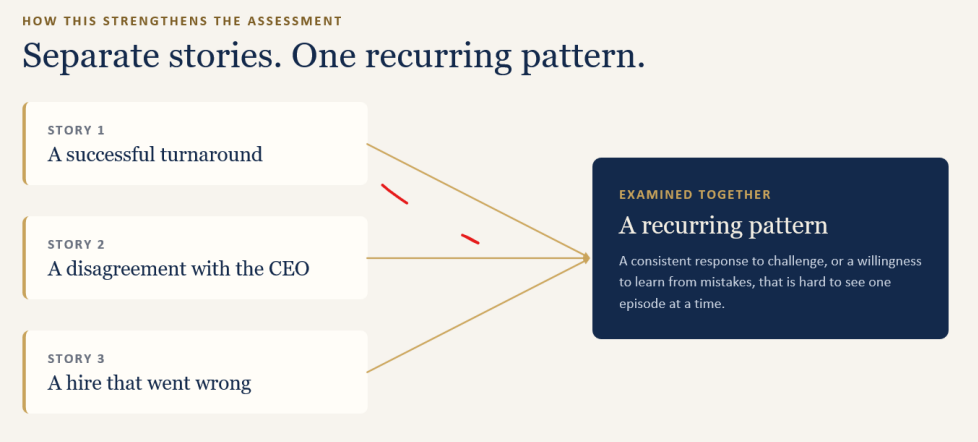
Evidence of additional value: GPT with Thorsten-4

millionways reports customer-experience pilots in which GPT paired with Thorsten-4 outperformed GPT alone at reading customer intent. These pilots illustrate the additive comparison: the existing AI system was supplemented with specialist behavioral findings.

Their relevance lies in testing whether those findings improved performance on the same task.

These customer-experience findings should be considered separately from evidence about executive assessment. They provide a basis for testing the combined approach in search, using interview and reference material, appropriate benchmarks and independent review. For search consultants, the opportunity is a more defensible assessment because of the structured findings that can be examined, challenged and connected to the client's decision. For AI applications, it is access to specialist behavioral context. In both cases, the case for adoption rests heavily on the improvement that can be measured and demonstrated over other AI tools.

A turnaround demands an appetite for difficult goals and uncomfortable decisions. A business built around partnership may depend more heavily on maintaining relationships and winning support. Therefore, understanding what drives the candidate helps the search consultant judge whether they will stay committed when the work becomes difficult, the board challenges them or results take longer than expected.



The Scheffer study provides encouraging evidence that a computer model can assess motivation from written responses with results comparable to trained experts.[25] This supports the approach behind Thorsten-4, while testing on executive interviews can establish how effectively it works in search assignments.

23 Brickman, Gupta and Oltmanns — Large Language Models for Psychological Assessment: A Comprehensive Overview (2025).
24 Scheffer and colleagues — Using Dynamic Item Response Theory and Machine Learning Based on Natural Language Processing to Improve the Reliability of the Operant Motive Test.
25 https://psycnet.apa.org/record/2025-07058-001

Why is millionways different from other assessment tools?

Experienced search consultants have heard many promises about better assessment. Executive search firms have invested substantial time and resources in developing their approaches to understanding leaders. Established assessment providers also offer extensively researched methods for examining personality, motivation and leadership potential.

The useful question is what millionways adds to that evidence and how its findings should be interpreted alongside it.

Thorsten-4 produces measurements across 50 defined behavioral and psychological traits covering motivation, decision-making, emotional regulation and responses under stress. Each finding traces to supporting language in the material analyzed. The interpretation presented to the consultant explains what those measurements may mean when considered together and in relation to the person, role and available evidence.[26]

This distinction matters. A transcript records what someone said. A trait measurement is the model's analytical output against a defined psychological dimension and interpretation considers the possible significance of that output. The consultant then decides what further investigation is needed and how the evidence should inform the recommendation. This is very valuable for the search consultant who still may have many questions unanswered.

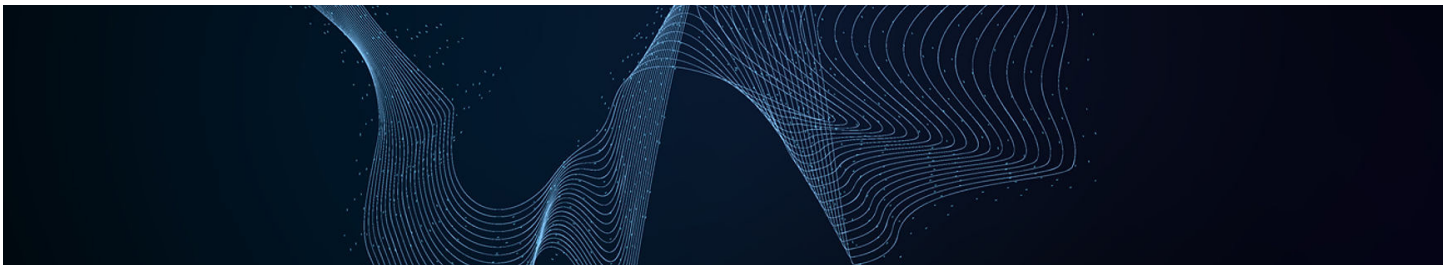
Building on established assessments

Established assessments give employers structured ways to understand how people work and lead. Hogan draws on decades of research into workplace personality, motivation and tendencies that can undermine effectiveness.[27] SHL's Occupational Personality Questionnaire examines preferred ways of working, approaching tasks and interacting with others. Korn Ferry combines assessment methods to consider leadership capabilities, potential and fit against the requirements of a role.[28]

These approaches have distinct purposes, methods, and supporting evidence. Their value should be judged against the decisions they are designed to inform.

millionways adds another assessment layer based on conversations firms are both conducting and have conducted, including interviews and reference calls. It applies defined psychological dimensions to the language in those conversations, producing trait findings that consultants can examine alongside established assessments, career evidence, and references.[29]

The contribution therefore extends beyond just simply summarizing an interview or suggesting questions but introduces a critical measurement layer between the source material and the interpretation of the candidate.



26 millionways — overview and defined behavioral dimensions.
27 Hogan Assessments — frequently asked questions.
28 Korn Ferry — Leadership & Professional Assessments.
29 millionways — behavioral assessment approach and applications.

From transcript evidence to professional judgment

The relationship between evidence, measurement and interpretation should remain visible throughout the assessment.

Stage	What it contains	What the consultant should establish
Transcript evidence	The candidate's words, preserved in context.	Is the record accurate and sufficiently substantive?
Trait measurement	Thorsten-4's output against a defined psychological dimension, linked to supporting passages.	What exactly does the trait measure? What does the reported reading mean, and how was it derived?
Interpretation	An explanation of what the reading, or combination of readings, may suggest.	Are alternative explanations considered? How relevant is the finding to this person and appointment?
Further investigation	The evidence needed to support, qualify or challenge the interpretation.	What should be checked through additional examples, references, existing assessments or further conversation?

A passage describing disagreement, for example, establishes the candidate's account of an event. A model output may identify a relevant trait reading supported by that passage and other language. The interpretation might suggest a possible difficulty considering opposing views.

Those are separate steps. The candidate's words do not automatically establish a stable characteristic, and a measured characteristic does not automatically establish a leadership risk. The consultant must examine the finding's meaning, the evidence supporting it and its relevance to the role.[30]

Traceability makes that examination possible. It allows the consultant and client to discuss a specific analytical result rather than accepting a broad description such as "collaborative," "defensive" or "resilient" without knowing its basis.

Looking beyond emotion and sentiment

Sentiment analysis serves a different purpose. Microsoft's sentiment analysis, for example, assigns positive, negative, neutral or mixed labels to text and provides confidence scores. These outputs are useful for understanding opinions and reactions.[31]

A positive or negative account does not, by itself, establish a leadership characteristic. An executive may describe a failed acquisition critically while acknowledging mistakes, recognizing colleagues' contributions and explaining what changed afterwards. Another may present the same outcome positively while providing little evidence of accountability or learning.

millionways' intended contribution is measurement against defined psychological dimensions, followed by interpretation of those findings. The distinction is between identifying the sentiment expressed in language and assessing behavioral characteristics through a specialist model.[32]

For the assessment described in this report, the input is written text or transcribed speech. The company has explained to us that it stopped using vocal tone as a basis for behavioral assessment after finding it unreliable for that purpose. Thorsten-4 therefore analyzes the language in a conversation rather than vocal characteristics such as pitch, cadence or delivery.

This does not eliminate every possible effect of communication style. Transcript accuracy, language fluency, editing and the substance of the conversation still affect the evidence available for analysis.

The research behind the measurements

Three questions help distinguish the research foundation from the interpretation of an individual assessment:

1. Is the underlying psychological framework supported by research?
2. Does the model measure the relevant characteristics reliably and accurately?
3. Do those measurements relate to meaningful behavior or outcomes?

30 millionways — behavioral assessment approach and applications.
31 Microsoft — sentiment analysis and opinion mining.
32 millionways — behavioral intelligence and research foundation.

Each question requires its own evidence. Support for a psychological theory does not establish the accuracy of every model output. Agreement with expert labels does not, by itself, demonstrate improved hiring results.

millionways describes a research program involving more than 7,000 participants observed over time and across different circumstances. Its stated aim was to identify behavioral patterns that persisted or changed as conditions changed.[33]

Research by Scheffer and colleagues illustrates how automated psychological measurement can be evaluated. The study examined written responses from 7,060 participants completing the Operant Motive Test. Trained experts assessed motivational content in those responses, and a computer model analyzed the same material, allowing comparison between automated and expert judgments.[34]

The researchers reported no statistically significant difference between the model's assessments and the expert assessments, with comparable performance in predicting managerial responsibility. That finding concerns the particular test, model, sample and outcomes studied. A lack of statistically significant difference should not be read as proof that two methods are identical.

The study provides encouraging evidence for automated motive assessment in a defined setting.[35] Separate testing is needed to establish how effectively Thorsten-4 measures its defined traits from executive interviews and how much those findings add to search decisions.

What does 90% accuracy mean?

In the validation material reviewed with the millionways team, the company reports testing its model on 10,500 responses that had not been used in training. The model's assigned psychological categories were compared with reference labels, producing approximately 90% agreement.[36]

This is a classification result on that test set. It is not an 90% hiring success rate, nor does it mean that every individual trait finding has an 90% probability of being correct. In approximately 15% of the test responses, the model's classification differed from the reference label. Understanding those differences requires information about the categories, the quality of the reference labels and the kinds of errors made.

millionways also reports a macro-F1 score of 0.80.[37] Macro-F1 gives equal weight to each category's F1 score, which combines precision and recall. It helps show whether performance extends across categories rather than being driven primarily by the most common ones. These figures provide information about model performance under the reported test conditions. Their relevance to executive search depends on how closely those conditions resemble the material and decisions involved in a search assignment.

A search consultant should also understand how a reported trait reading is expressed: whether it is a category, a score or another indicator; what its scale means; and whether any comparison group or uncertainty estimate is provided. A number should not be presented as a percentile, probability or degree of certainty unless the methodology supports that interpretation.

Consistency and changes in an assessment

Accuracy and consistency answer different questions. Accuracy concerns agreement with an appropriate benchmark. Consistency concerns whether the same material produces reproducible findings under comparable conditions.

The reported accuracy and macro-F1 figures do not, on their own, establish consistency. That requires separate testing.

When another conversation changes an assessment, the consultant should be able to distinguish a change in the underlying measurement from a change in its interpretation. New passages may alter a trait reading; alternatively, the reading may remain stable while new information changes its relevance to the role.

A clear assessment should explain what changed, which evidence caused the change and why it matters. That makes revision understandable and allows the consultant to defend the final advice.

33 millionways — behavioral intelligence and research foundation.
34 Scheffer and colleagues — research on automated assessment using the Operant Motive Test.
35 Scheffer and colleagues — research on automated assessment using the Operant Motive Test.
36 millionways — company-reported validation results; see the supporting validation material in the appendix for the test methodology and figures.
37 millionways — company-reported validation results; see the supporting validation material in the appendix for the test methodology and figures.

Language, culture and communication

Search consultants regularly assess executives whose backgrounds and ways of communicating differ from their own. A reserved speaker may make difficult decisions confidently. Someone speaking a second language may struggle to explain an experience they handled effectively.

Focusing on language rather than vocal delivery does not establish that measurements work equally well across languages and cultures. Performance needs to be examined in the settings in which the tool will be used.[38] The International Test Commission's guidance reinforces the importance of cultural context when adapting and using assessments, including situations in which translation is not required.[39]

For consultants, this begins with an accurate record of a substantive conversation. An unclear answer may require clarification. An edited account, interviewer's notes and a full transcript may provide different evidence. A carefully prepared written response may also differ from a spontaneous discussion.

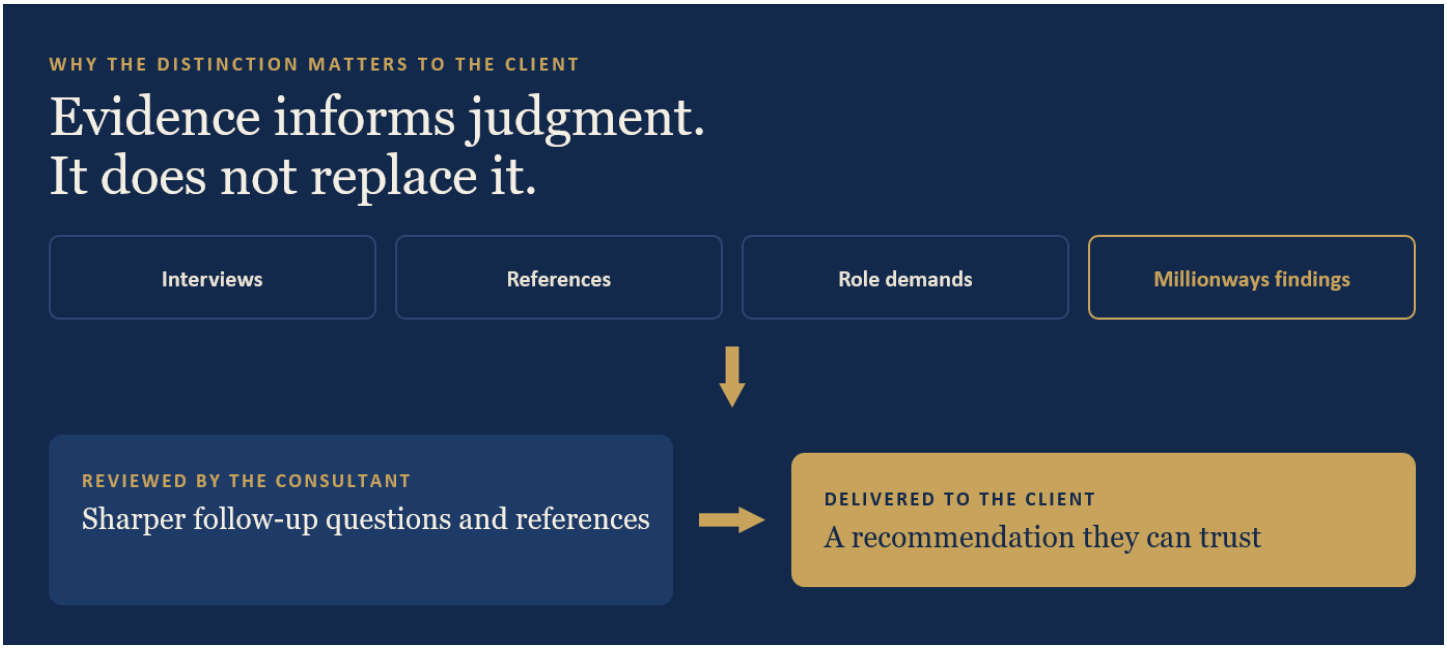
The resulting interpretation should reflect those limitations. Where the material is insufficient, the assessment should identify the gap rather than turn limited evidence into a confident conclusion.

What this adds to the client's decision

The value of millionways lies in connecting defined psychological measurements to evidence that can be examined and interpreted in context.

The search consultant brings knowledge of the client's business, the appointment and the candidate's history. Thorsten-4 supplies an additional analytical perspective. Established assessments, references and demonstrated achievements provide other evidence against which its findings can be considered.

Keeping these contributions distinct strengthens the recommendation. The client can understand what the candidate said, what the model measured, what the interpretation suggests and what remains unresolved. That provides a clearer basis for deciding whom to appoint and what support may help that person succeed.



38 millionways — assessment approach.
39 International Test Commission — guidance on adapting and using assessments.

The person, the role, and the business: why context matters

An executive can thrive under one CEO and struggle under another. Someone who leads confidently in a stable business may find an acquisition, restructuring or board disagreement harder to navigate.

For a search consultant, understanding those differences is central to the assignment. The client needs someone who can deliver what the business requires and work effectively with its leaders. Success elsewhere provides important evidence, but does not guarantee success here.

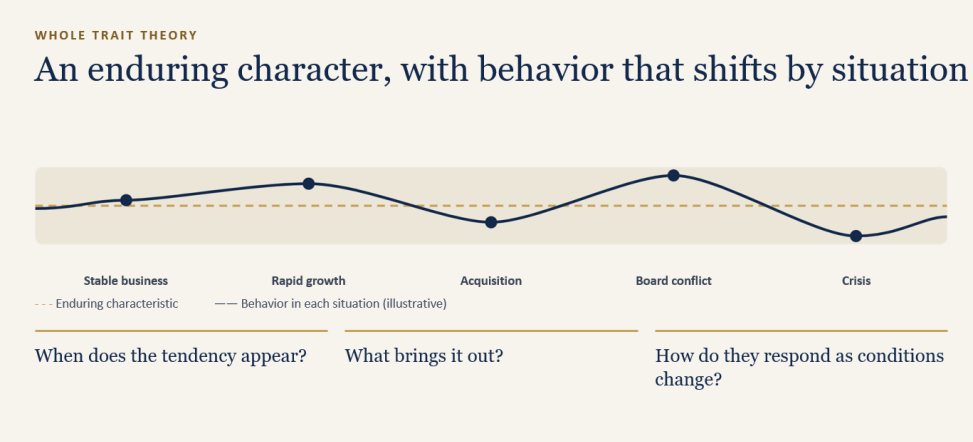
millionways' findings should therefore be considered against the pressures, relationships, and responsibilities of the hire. The key question is how the person's behavioral characteristics interact with the circumstances in which they will be expected to lead.

The person and the circumstances

A candidate who describes taking control during a crisis may have made an appropriate decision. What matters is whether this was a temporary response to an exceptional problem or an enduring tendency that appears whenever others disappoint them. A turnaround may require direct intervention, followed by a deliberate return of responsibility to the team.

Whole Trait Theory explains how relatively enduring characteristics can coexist with considerable variation in behavior across situations.[40] This supports examining when a tendency appears, what brings it out and how behavior changes as conditions change.

Stress, fatigue and grief also matter. Someone experiencing bereavement or an exhausting period at work may communicate differently from their usual self. A difficult interview should not automatically become a lasting judgment about the candidate. millionways describes its longitudinal research as a basis for examining behavioral patterns over time and distinguishing them from temporary reactions.[41]



The operating principle is straightforward: interpret a finding against its supporting language, the circumstances described and the wider evidence. Where those sources leave a material uncertainty, obtain the specific information needed to resolve it. This gives further investigation a clear purpose rather than simply adding more interviews or references.[42]

Should two interviews produce the same findings?

millionways expects underlying motivational patterns to show meaningful consistency when interviews contain enough comparable material but this does not mean every conversation should produce identical findings.[43]

40 Whole Trait Theory.
41 millionways — Science: longitudinal research and interpretation in context.
42 millionways — Science: longitudinal research and interpretation in context.
43 millionways — Terms of Use.

An interview about ambition and growth will contain different evidence from one about failure or a difficult dismissal. Some differences may broaden the assessment; others may reveal a limitation in the material or a finding that needs reconsideration.

The search consultant should distinguish between three possibilities: the conversations cover different subjects, the person's response varies with the circumstances, or the assessment itself is inconsistent. Reviewing the supporting passages and the depth of each conversation helps establish which explanation is most plausible.[44]

SIOP's guidance identifies reliability and consistency as important requirements for AI-based employment assessments.[45] This provides a standard against which providers should be questioned and their performance tested.

When an assessment changes, the explanation should identify what changed in the evidence or measurement and why the interpretation was revised. The result should be a clearer account of the conditions in which the executive may perform well, encounter difficulty or require support.[46]

Interview integrity and the limits of language analysis

Preparation and coaching can help candidates communicate their experience clearly. They can also make it harder to distinguish a polished account from evidence of how someone actually behaves. millionways' analysis examines behavioral patterns in language, but it cannot establish that the events described occurred.[47]

Language analysis cannot authenticate a biography or establish honesty. Career verification and other appropriate checks remain necessary, regardless of how convincing an interview or assessment appears.[48]

millionways is developing an Interview Integrity & Authenticity Risk API with an AI recruiting company. The API produces risk signals that flag possibly AI-assisted, coached or proxy answers for follow-up. These signals prompt investigation; they do not authenticate a biography or establish honesty. A flag should therefore be treated as a reason to clarify the circumstances, rather than simply finding a deception. The relevant question



44 millionways — Terms of Use.
45 Society for Industrial and Organizational Psychology — Considerations and Recommendations for the Validation and Use of AI-Based Assessments for Employee Selection (January 2023).
46 millionways — Terms of Use.
47 Responses to interviews conducted for this publication concerning interview coaching and candidate deception.
48 Responses to interviews conducted for this publication concerning interview coaching and candidate deception.

"A candidate profile alone cannot establish whether someone will work successfully with a particular CEO, board or leadership team."

is whether the account can be substantiated and whether any assistance affects its usefulness as evidence of the candidate's capabilities. The quality of the source material remains very important. Brief or heavily edited transcripts, rehearsed answers and missing context can restrict what the model can assess. Where the evidence is insufficient, the appropriate response is to identify that limitation rather than simply infer a confident conclusion about the person.[49]

Understanding the people the candidate will join

A candidate profile alone cannot establish whether someone will work successfully with a particular CEO, board or leadership team. Assessing a working relationship requires information about the relevant people and the circumstances in which they will work together.[50]

The search consultant needs to understand the objectives of why the hire is being made, the executive's actual authority and how existing leaders make decisions, share responsibility and respond to challenge. This also tests whether the organization is prepared to support the contribution it is asking the new executive to make.

With behavioral material from the relevant people and a clear account of the role, millionways can contribute interpretations of where their approaches may complement one another or create friction. These remain propositions to examine against evidence and actual interactions but claims about how a team will perform together always require their own validation.[51]

Differences are not automatically beneficial or problematic. Research on team diversity shows that differences in personality, values and culture can affect team processes and conflict; assembling different profiles does not itself establish that performance will improve.[52]

A CEO who moves quickly, for example, may benefit from an executive who identifies risks before the business commits. That relationship will be productive only if the CEO considers the concerns and the executive can raise them without unnecessarily delaying action. Observing the two discuss a real business issue can reveal how that balance works in practice.

This extends the assessment beyond whether the candidate "fits." It asks whether the people involved can make better decisions together, what conditions would support that relationship and what each may need to adjust. That gives the client a more useful basis for both appointment and onboarding and brings the search consultant into the client's inner circle.

49 Responses to interviews conducted for this publication concerning common model errors and executives or situations that are difficult to assess.
50 Response to interviews conducted for this publication concerning the assessment of an executive's likely success within a particular company and culture.
51 Response to interviews conducted for this publication concerning how a candidate may work with the CEO, board and existing leadership team, and the evidence supporting that assessment.
52 The Relationship Between Team Deep-Level Diversity and Team Performance: A Meta-Analysis of the Main Effect, Moderators, and Mediating Mechanisms.

Where can millionways improve a business decision and its impact on the bottom line?

A leadership decision can quickly become a financial decision. An unsuccessful hire may delay growth, disrupt a capable team, and leave the business paying for a replacement search. Promoting someone before they are ready can expose the organization to risks that development might have addressed earlier.

The value of behavioral intelligence becomes tangible when it changes a decision or course of action: whom to appoint, what management capacity an investment requires, how a new executive will work with the CEO or what experience a successor needs before promotion. millionways identifies hiring, leadership development, coaching, succession, and collaboration among its applications.[53]

The four examples below are hypothetical, not documented millionways results.

Each follows the same sequence: decision → additional insight → action → outcome to evaluate. In every case, model findings would need to be considered alongside their supporting language, the business context and other relevant evidence.

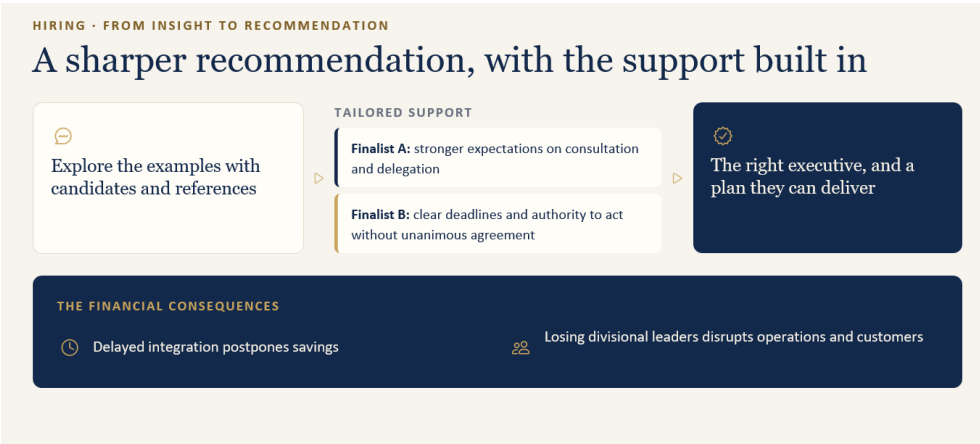
1. Hiring: compare candidates against a specific mandate

Decision: A company needs an executive to integrate three acquisitions while retaining key employees and customers. Two finalists have strong records, but the client must decide whose capabilities best match this particular mandate. Assessment should address the requirements of the job, consistent with OPM's guidance on structured interviews.[54]

Additional insight: Suppose Thorsten-4 produces different readings across relevant dimensions of decision-making, motivation and emotional regulation, each linked to supporting passages. One candidate's profile suggests a preference for rapid commitment despite incomplete information; the other suggests greater emphasis on consultation and resolving uncertainty before acting.

Neither profile establishes who should be hired. The additional insight makes the tradeoff more explicit: how will each candidate balance integration speed with the commitment of the people expected to deliver it?

Action: The consultant compares those findings with evidence of previous integration results, the candidates' individual contributions and the demands of the current plan. The recommendation identifies which candidate offers the stronger match and any support or operating expectations needed.



53 millionways — Workforce & Leadership.
54 U.S. Office of Personnel Management — Structured Interviews.

Outcome to evaluate: Track integration milestones, delivery of planned savings, retention of key employees and customer continuity against agreed targets. These measures help establish whether the appointment delivered the mandate; assessing millionways' additional contribution requires comparison with the existing assessment approach.

2. Management diligence: can leadership deliver the investment plan?

Decision: A private equity sponsor is considering an investment whose returns depend on expanding into new markets and improving operating margins. Before committing, it needs to determine whether the management team can deliver both objectives and what additional capability the plan requires.

McKinsey reports that half of senior private equity leaders in its survey identified waiting too long to replace underperformers in critical roles as their biggest talent-related regret.[55] This reinforces the importance of examining leadership assumptions before they become execution problems.

Additional insight: Suppose the combined assessment suggests that the team brings considerable energy to commercial expansion but may be less effective at sustaining attention to operating discipline and difficult performance decisions. Thorsten-4's findings provide a structured perspective on those potential gaps, to be examined alongside management's track record and the operating data.

The question becomes specific: does the investment plan assume capabilities or working habits that the team has not yet demonstrated?

Action: The sponsor incorporates the leadership requirements into diligence and the investment case. This could mean adding an experienced operating executive, strengthening financial planning capability or establishing clearer accountability for margin improvement. The associated cost and implementation time become explicit assumptions in the plan.

Outcome to evaluate: Review delivery of the operating plan, forecast accuracy, margin improvement and the effectiveness of any management additions. The financial question is whether the investment in leadership capacity helps management execute the value-creation plan.

3. Onboarding: agree authority and working expectations

Decision: A newly appointed executive and the CEO need to establish how they will work together during the first six months. Both support the mandate, but they have different expectations about consultation and independence.

Additional insight: With material from both people, behavioral findings may help explain the difference. The executive may work most effectively with explicit authority and periodic review, while the CEO expects frequent discussion before commitments are made.

The issue is a potentially avoidable mismatch in working expectations, rather than a conclusion that either person lacks leadership ability.



Action: Before the executive starts, agree decision rights, matters requiring consultation, escalation thresholds and the frequency of progress reviews. Connect these arrangements to the business milestones the executive must deliver.

Outcome to evaluate: Assess time to agreed productivity milestones, decision turnaround, avoidable escalations and clarity of responsibilities. Review whether the arrangements support delivery and adjust them as the relationship develops.

55 McKinsey & Company, "Unlocking full potential: Five practices reshaping PE value creation," June 25, 2026, section "Mobilize employees deep in the organization."

4. Succession: identify development needs before promotion

Decision: A board is considering a successful divisional leader as a future CEO. It needs to decide what the executive must demonstrate before being judged ready for enterprise responsibility.

Additional insight: Suppose the assessment indicates strong achievement motivation and analytical capability, alongside a possible development need in integrating competing priorities across the organization. Supporting passages suggest that the executive consistently evaluates decisions through the interests of their own division.

The finding is a proposition to test: can this leader make and sustain decisions that benefit the whole business when their own division must accept a disadvantage?

Action: Give the executive responsibility for an enterprise initiative involving several divisions, with explicit tradeoffs, measurable objectives and experienced support. Use coaching to address specific behaviors and review progress through observed performance.

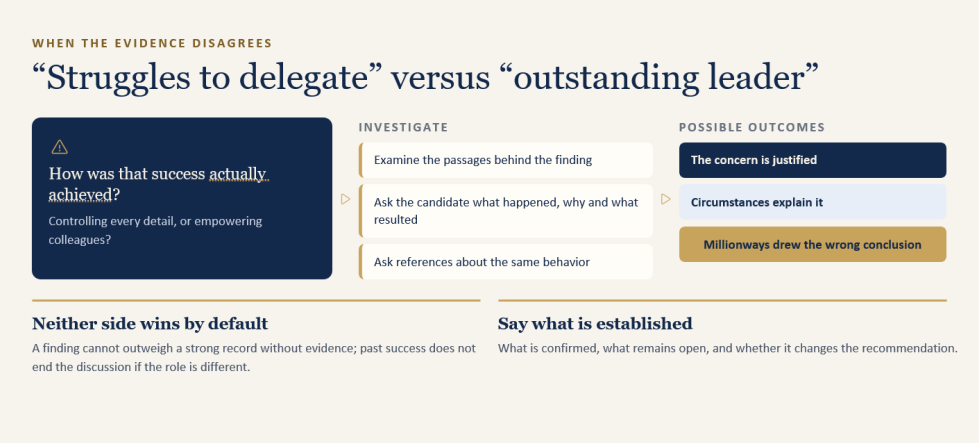
Published research supports coaching as a development intervention. A review of 37 randomized controlled studies reported positive effects across leadership and personal outcomes, although results varied and some measures relied on self-reports.[56] This supports coaching generally; it does not establish the additional benefit of using millionways to guide it.

Outcome to evaluate: Assess delivery of the enterprise assignment, the quality of cross-divisional decisions, stakeholder feedback and progress against agreed readiness criteria. The board gains evidence for a promotion decision while the executive has time to develop before assuming the larger role.

When the evidence disagrees

A model finding may conflict with the consultant's own assessment, references or demonstrated achievements. That disagreement should be resolved at the level of the specific finding: what language supports it, what the model measured and whether the interpretation fits the circumstances.

Reference questions should address behavior relevant to the new role, consistent with OPM's guidance.[57] The investigation may support the concern, explain it through context or establish that the model's interpretation was mistaken.



The final advice should distinguish what has been established, what remains uncertain and whether the finding changes the decision or proposed action.

Client value first, commercial opportunity second

Across these applications, the client receives a clearer decision and a practical course of action. The potential value lies in better hires, more realistic investment assumptions, faster transitions and stronger succession readiness. Outcomes should be agreed and reviewed rather than just assumed.

For search firms, this creates a timely opportunity to extend their advice into management diligence, onboarding and succession. Such work can support paid advisory engagements where there is a defined client need, appropriate expertise and a clear basis for evaluating value.

56 Research review of coaching interventions.
57 U.S. Office of Personnel Management — Reference Checking.

What evidence shows that it makes a real difference?

The commercial opportunity follows from helping clients make consequential leadership decisions well and remaining useful and aligned as those decisions move from assessment into execution.

A search consultant's credibility rests on the quality of their advice and the evidence behind it. When millionways informs a recommendation, the consultant needs to explain what it measured, how the findings were checked and whether they changed the decision.

The evidence should be considered at five distinct levels: research supporting the psychological concepts; evidence concerning measurement; associations with observed behavior; examples of practical use; and evidence of improved decisions or business outcomes. Each answers a different question. Support at one level does not automatically establish the next.

1. Research supporting the psychological concepts

The first question is whether the characteristics being assessed have a meaningful research foundation.

In their AT&T study, McClelland and Boyatzis examined motive assessments collected when 237 managers entered the company and compared them with advancement eight and sixteen years later. Particular motivational patterns were associated with advancement among nontechnical managers; the same relationships were not found among technical managers with engineering responsibilities.[58]

Both findings matter. Motivation may be relevant to leadership and career progression, but its significance depends on the work involved. A characteristic associated with success in one setting may have different implications in another.

This research supports examining motivation in leadership assessment. It does not establish that Thorsten-4 measures motivation accurately or improves hiring decisions. Those require separate evidence.

2. Evidence concerning the model's measurement

The second question is whether an automated system measures the intended psychological characteristics accurately and reliably.

Scheffer and colleagues examined Operant Motive Test data from 7,060 participants. They compared automated motive coding with trained experts' assessments and reported no statistically significant difference, alongside comparable results in predicting managerial responsibility.[59]

This provides evidence for automated motive coding within the particular assessment studied. It should not be treated as validation of every Thorsten-4 trait or its application to executive interviews.

The company-reported classification results discussed earlier in this report address another measurement question: how closely model outputs agreed with reference labels on a held-out test set. Such results should be read alongside the test design, category definitions, reference-label quality and error analysis.

For executive search, the relevant next step is to establish how accurately and consistently Thorsten-4 measures its defined traits from the kinds of interviews and referencing calls firms actually conduct.

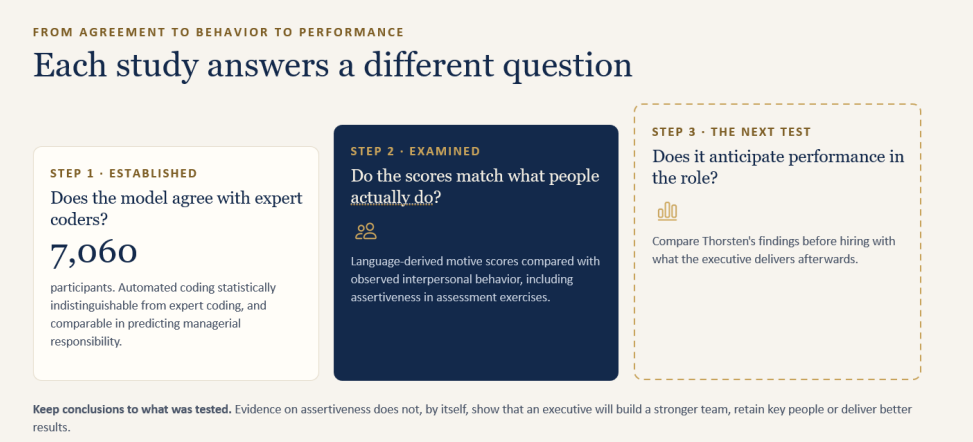
3. Associations with observed behavior

Agreement with expert labels is useful, but a further question is whether the measurements correspond with what people do.

millionways supplied an assessment-center white paper examining associations between language-derived motive

58 McClelland and Boyatzis — Leadership Motive Pattern and Long-Term Success in Management (1982). Original reference link retained: https://www.scirp.org/reference/referencespapers?referenceid=3377562. Original paper: https://doi.org/10.1037/0021-9010.67.6.737

59 Scheffer and colleagues — automated motive-coding study.



scores and observed interpersonal behavior, including assertiveness during assessment exercises.[60]

This moves the examination beyond agreement between analytical judgments. It considers whether particular scores relate to independently observed behavior. The methods, results and limitations should be reviewed in the supporting paper.

The conclusions must remain specific to the study. An association with assertiveness does not establish that an executive will retain key employees, improve team performance or deliver an investment plan.

The company-reported fraud research

millionways also reports research examining associations between executive language and subsequently confirmed fraud. It reports 89% accuracy on held-out material and 88% in cross-validation.[61]

If substantiated, these results would concern classification within that research setting. They would not establish that Thorsten-4 can determine whether an individual candidate is lying.

Separate research by Larcker and Zakolyukina examined executive language in earnings calls, using subsequent financial restatements and criteria concerning accounting problems to label the material studied.[62] This supports investigating relationships between communication and independently recorded events, while addressing a different question from candidate honesty.

4. Examples of practical use

Practical examples show what the technology produces and how someone might use it. They do not necessarily demonstrate measurement accuracy or improved outcomes.

Four anonymized executive interviews

An exercise involving four executive interviews illustrates the assessment output. Thorsten-4 identified differences in motivation, decision-making and potential behavioral fit, giving consultants findings to examine alongside the supporting passages and other evidence. The analyses are included in the appendix.

millionways reports that one participating executive verbally described the analysis as very accurate. That feedback is encouraging, but it represents one executive's response. It does not establish accuracy across all four interviews or demonstrate that the analysis improved a hiring outcome.

The exercise demonstrates the form and potential usefulness of the findings. Establishing additional value requires documenting what the consultant learned, how it was checked and whether it changed the recommendation or subsequent action.

Interview integrity: preliminary result, development work and operational validation

1. **Reported preliminary result.** millionways describes a proof of concept involving 30 videos supplied by an AI recruiting partner without their authenticity labels. The company reports successful identification of authentic material and says the partner's CEO confirmed the result by email.[63]

   The supporting record needs to specify how authenticity

60 millionways — assessment-center white paper supplied for this publication, concerning language-derived motive scores and observed interpersonal behavior. See appendix.
61 The Fraud study was a separate internal study millionways did back in 2023 and is available on request
62 Larcker and Zakolyukina — Detecting Deceptive Discussions in Conference Calls.
63 millionways — proof-of-concept materials supplied for this publication, with accompanying correspondence reporting confirmation from the CEO of an AI interviewing platform. See appendix.

was defined, how the labels were established, which inputs were analyzed and the number and types of correct and incorrect classifications. The reported result is preliminary.

2. **Integration in development.** millionways is developing its Interview Integrity & Authenticity Risk API with that partner. The API is intended to produce risk signals flagging possibly AI-assisted, coached or proxy answers for follow up. The proposed integration specification distinguishes Thorsten's text analysis from a separate voice-analysis component; accounts of any test must identify which components produced the reported results.[64]

3. **Validated operational capability.** Neither a preliminary exercise nor an integration specification establishes reliable performance in live recruiting. That requires evaluation on representative material, including false-positive and false-negative rates and the usefulness of the signals in practice.

   These signals prompt investigation. They do not authenticate a biography or establish honesty.

Personalization: applying behavioral understanding across people and AI

Personalization provides another application of the same behavioral intelligence capability. Structured psychological findings can inform the questions asked, explanations offered and support provided by a professional or connected AI system.

millionways supplied two customer-communication comparisons to Hunt Scanlon illustrating this application. In the luxury health and fitness company example, behavioral findings informed how the response addressed the individual's decision-making preferences. The company reports that the client's CTO described the change as a significant improvement.[65] A global technology company customer-care comparison illustrates how behavioral context can inform clarification, escalation and accountability in an interaction.[66]

The significance extends beyond wording or tone. These examples illustrate how structured behavioral understanding can influence what a professional or AI system asks, explains or recommends next.

In executive search, the same capability could guide candidate questioning, onboarding discussions and leadership development. As search firms adopt AI note-takers, screening tools and CRM assistants, Thorsten-4 can supply behavioral context to those systems across the stages of a search. The consultant can draw on that context directly or use it within an existing AI application, while considering the findings alongside the person, role and available evidence.

The customer-communication examples illustrate how behavioral context can influence a response. Reported positive feedback is encouraging, but controlled evaluation is needed to establish the additional benefit over the existing approach. Neither example establishes executive-assessment accuracy or improved hiring outcomes.

5. Evidence of improved decisions or business outcomes

The strongest practical evidence connects an assessment to a decision, an action and a subsequent result.

For a search assignment, that record should establish:

- What the existing assessment indicated before Thorsten-4 was added.
- What additional finding Thorsten-4 produced.
- How the consultant checked it.
- Whether it changed the recommendation, onboarding plan or development priorities.
- What happened afterwards.
- Recording findings before the appointment reduces the risk of interpreting them retrospectively to fit the outcome.

After hiring, agreed reviews, for example, at six and twelve months, can then go on to examine leadership effectiveness, delivery against objectives, team experience, retention and time to productivity. SIOP's discussion of quality of hire supports considering multiple measures over time.[67]

These measures should be interpreted together. Retention alone by itself does not establish success: an executive may remain

64 millionways — proposed integration specification and accompanying materials supplied for this publication. Integration status: in development. See appendix.
65 millionways — luxury health/fitness company customer-response comparison and accompanying correspondence supplied for this publication. See appendix.
66 millionways — global technology company customer-care agent comparison supplied for this publication. See appendix.
67 SIOP — Quantifying Quality: Best Practices for Measuring Quality of Hire.

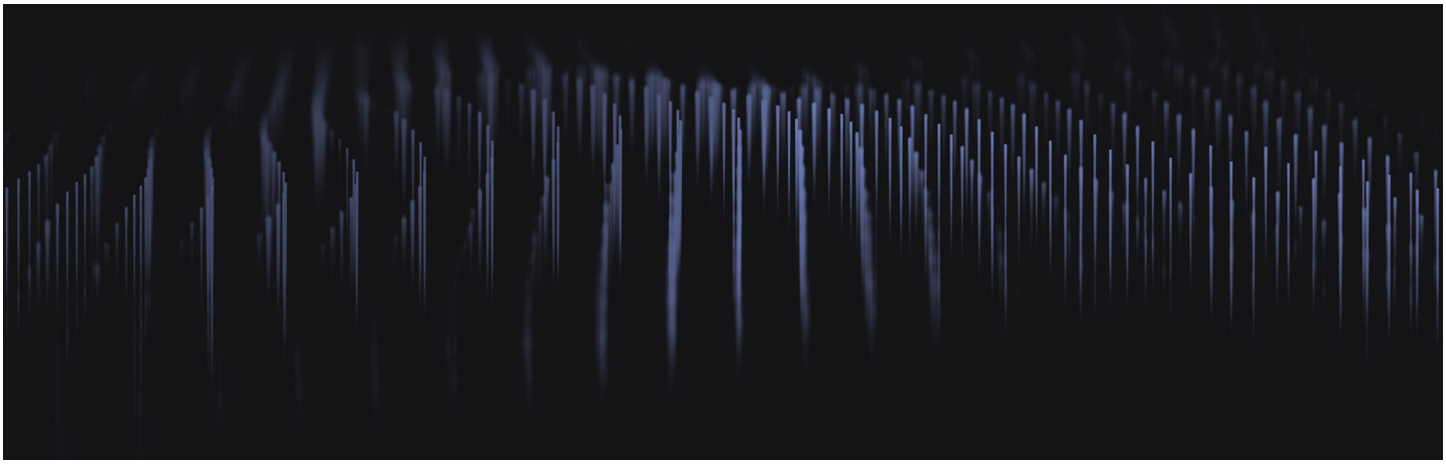


employed while missing targets or losing capable colleagues. Equally, a difficult business outcome may reflect market conditions or constraints beyond the executive's control.

Tracking outcomes can establish whether findings actually corresponded with subsequent behavior and whether the advice proved useful. Demonstrating that Thorsten-4 improved outcomes require a credible comparison with the existing approach, accounting for relevant differences between assignments.

The material described here supports different parts of the case: the psychological foundation, aspects of automated measurement, particular behavioral associations, and examples of application. It does not yet establish improved executive hiring outcomes.

That distinction creates a clear agenda for evaluation: test the measurements, document what changes in practice and follow the results. This gives clients a transparent basis for judging where the technology adds value. and keeps the search consultant in the client's inner circle.



How can an organization test it and gain from it?

A focused evaluation can establish whether millionways adds useful behavioral insight to a firm's assessment process or an AI application it already uses. The starting point is one defined decision or workflow, a clear comparison, and agreed measures of value.

Before beginning, the search firm and millionways should agree on a short evaluation brief covering five key elements:

Element	What to agree
Decision or workflow	One specific application and the question the evaluation should answer.
Existing approach	The current process or system against which the additional contribution will be assessed.
Material and participants	The cases, source material, consultants, reviewers, and other participants required.
Measures	How useful findings, unsupported interpretations, missed concerns and effort will be recorded.
Ownership and review	A named owner, a review date and criteria for deciding whether to stop, refine, extend or adopt the approach.

1. Choose one decision or workflow

Select an application that matters to the firm or its client: comparing finalists against a mandate, assessing succession readiness, preparing onboarding discussions, or supporting management diligence.

The evaluation may also examine an existing AI tool. For example, a firm could test whether adding Thorsten-4's structured findings improves the usefulness of an interview-analysis assistant already in its workflow.

Frame the question precisely: "Does adding Thorsten-4 identify relevant, supported findings that our current process misses?" is easier to evaluate than "Does it help us understand people better?"

Involve the search consultants and client leaders closest to the decision, with assessment, technical and privacy expertise as appropriate. For selection applications, the EEOC's guidance reinforces the importance of evaluating a procedure's suitability for its intended use and the job involved.[68]

2. Define the existing approach used for comparison

Record how the firm currently performs the task: the assessment methods, information available, people involved and any AI models or applications used.

Where the purpose is to test Thorsten's additional contribution, compare the existing approach with the same approach supplemented by Thorsten-4's findings. If comparing separate systems, give each identical source material and the same assessment task. Record the model versions, instructions and supporting information used.

Ask independent reviewers who did not produce the assessments to examine the outputs without knowing which system or approach produced them, where feasible. Agree the scoring criteria before the review.

This makes the comparison more informative and tests what the additional behavioral findings contribute while limiting differences caused by unequal inputs or evaluation conditions.

3. Agree the material and participants required

Choose cases that represent the intended use. Agree the number and type of cases, the material needed for each, who will prepare it and who will review the results.

68 U.S. Equal Employment Opportunity Commission — Employment Tests and Selection Procedures.

Identify the source of each item. A full transcript preserves the executive's words; interview notes contain selected details and the consultant's observations; edited writing may reflect preparation or assistance from others. These sources should not be treated as interchangeable.

Enough material to generate an analysis does not necessarily mean enough evidence to support a hiring conclusion so the evaluation should record where the available information is insufficient.

A retrospective exercise using completed searches can provide a useful and practical starting point. Use only material available before the original hiring decision, keeping subsequent outcomes hidden from the model and initial reviewers. Compare the assessments before examining what happened afterwards.

A small historical exercise can identify promising applications. Establishing improved hiring outcomes requires a larger prospective study of new cases.[69]

For a prospective evaluation, record the findings before the decision, how they were checked and whether they influenced the advice or action. Any later review should return to that original record.

4. Measure useful findings, unsupported interpretations and effort

Agree what counts as a useful contribution before examining the outputs. A finding should be relevant to the decision, supported by identifiable evidence and sufficiently specific to inform an investigation or action.

A practical scorecard could include:

What to Measure	What to Record
Useful additional findings	Findings that were relevant, supported and not already available from the existing approach
Evidence traceability	Whether findings could be traced clearly to supporting material
Unsupported interpretations	Conclusions that went beyond what the evidence could support
Missed concerns	Important issues identified by the existing process or reviewers that Thorsten did not surface
Consistency	Whether comparable material produced stable findings when reanalyzed or challenged
Effect on decision/ workflow	Whether the findings changed, clarified or strengthened the decision or next action
Effort and cost	Time required to prepare material, review findings and incorporate them into the process

Record unhelpful outputs as carefully as useful ones. A finding that requires substantial checking but contributes little may have limited practical value, even if it initially appears insightful.

Independent review can assess usefulness and evidential support. Claims of greater measurement accuracy require an appropriate benchmark, clearly defined measures and an adequate sample.

For an existing AI application, assess the combined workflow as well as individual outputs. Additional context is valuable only if it improves the agreed task enough to justify the effort involved.

NIST's AI Risk Management Framework supports evaluation in the intended context and continued monitoring as systems and conditions change.[70]

Protect participant information and preserve meaningful review

Before sharing material, agree what will be analyzed, why, who will receive the findings and how information will be protected, retained and used, including any reuse for model training. Establish the applicable privacy requirements and arrangements for informing participants and obtaining consent where required.[71, 72]

69 Response to interviews conducted for this publication concerning retrospective pilot exercises and the need for larger prospective studies to establish improved hiring outcomes.
70 National Institute of Standards and Technology — AI Risk Management Framework.
71 Learnsignal — commentary on ICO data-protection considerations for AI in recruitment.
72 Information Commissioner's Office — AI in Recruitment: Audit Outcomes Report.

Participants should have a clear route to question findings, correct errors or supply missing context. Consultants must be able to examine the evidence and revise or reject an interpretation. ICO guidance emphasizes meaningful human involvement in automated recruitment processes.[73]

These arrangements should form part of the evaluation brief, with responsibilities assigned before the work begins.

5. Assign an owner, set a review date, and decide what happens next

Name one person within the firm to own the evaluation and a counterpart at millionways. Agree the start date, review date, deliverables and criteria for further use.

At the review, examine the results against the original question:

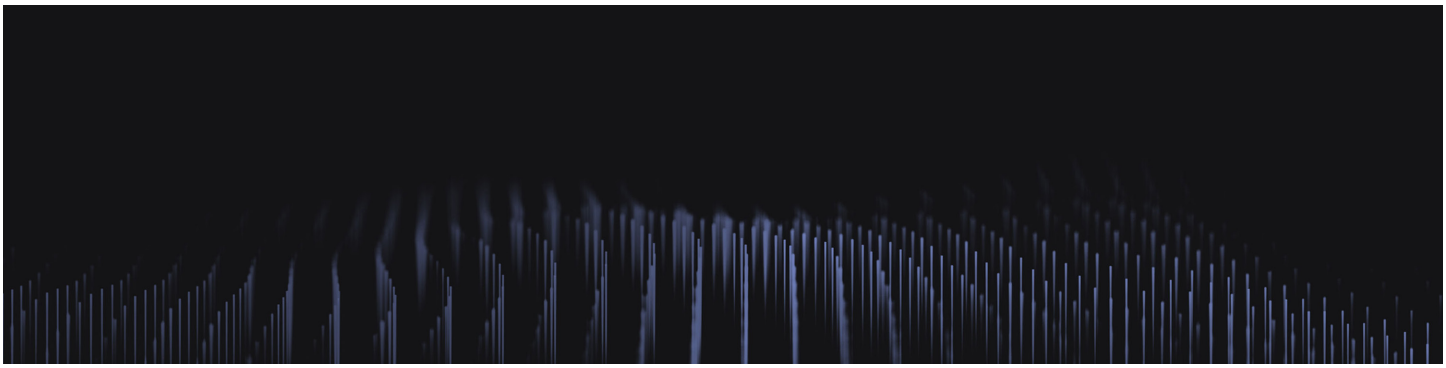
- Did the approach produce useful, supported findings beyond the existing process?
- What unsupported interpretations or missed concerns occurred?
- How much time and effort did preparation, review and use require?
- Did the findings change a decision or improve the selected workflow?
- What remains untested?

The resulting decision should be explicit: stop, refine the scope, extend the evaluation or adopt the approach for a defined use. A promising pilot may justify further testing without establishing improved hiring or business outcomes.

AN INVITATION TO CONNECT.

Start with one leadership decision your firm needs to make.

Contact Max Weidemann at **max@millionways.ai** to scope a focused evaluation using interview or reference material you already hold and have permission to analyze. Together, we will agree on the decision to support, the current approach for comparison, the required material and participants, and what a useful result would look like. The evaluation should assess additional insight, unsupported interpretations, and the time required to review the findings. An agreed owner and scheduled results review will give both sides a clear basis for deciding whether to expand the work.



73 Information Commissioner's Office — Recruitment Rewired.

Appendix

The appendix brings together research, company validation materials and examples of practical application. Each item supports a different type of claim and should be read according to its methods, sample and limitations. Research associations, illustrative outputs and preliminary exercises carry different evidential weight. Improved hiring, retention or commercial outcomes require evaluation in the setting where the technology will be used.

Executive search pilot analysis. An anonymized company report examining four executive interviews and their behavioral profiles. It illustrates how model outputs can inform candidate assessment and further investigation. It does not establish improved placement or retention outcomes.

Company validation summary. A company prepared overview of the psychological foundation, proprietary data and validation work behind Thorsten. Readers should distinguish the scientific research cited, the company's own studies and reported applications, and consult the underlying methods for each claim.

Customer care response comparisons. Selected examples showing how access to behavioral context can change a response's guidance, escalation and accountability. These examples illustrate output differences; they do not establish measured improvements in customer outcomes.

Assessment center research paper. A company authored research paper examining associations between language derived scores and observed behavior in a university assessment center. The written study covers 117 participants across three cohorts. A separate voice analysis is reported for one cohort. Results concern the specific measures and tasks studied, with methods and limitations available for review; they are not a general percentage accuracy claim for leadership selection.

Behavioral profile and communication guide. An example of structured behavioral findings and their translation into practical communication guidance. It illustrates an application of the methodology, rather than a controlled test of investment or client relationship outcomes.

Salesforce email comparison. A simulated customer scenario showing how behavioral context can inform an email response. It is an illustrative comparison, rather than a controlled evaluation of Salesforce performance, psychological accuracy or sales conversion.

Interview integrity pilot and API in development. In a proof of concept involving 30 interview videos provided without authenticity labels, millionways accurately predicted which videos were authentic. This work is part of an interview integrity and authenticity risk API in development with an AI recruiting partner, designed to flag possible coaching, AI assistance or proxy participation for further review.

millionways

Case Study with Executive Search Firm

1. Introduction

We're excited to share the results of our pilot analysis, where we used millionways to evaluate four executive candidate interviews provided by a Global Executive Search Firm. This document highlights how our deep, transparent personality insights—rooted in 30 proprietary psychological dimensions—can be leveraged to elevate executive search with unmatched precision and psychological clarity.

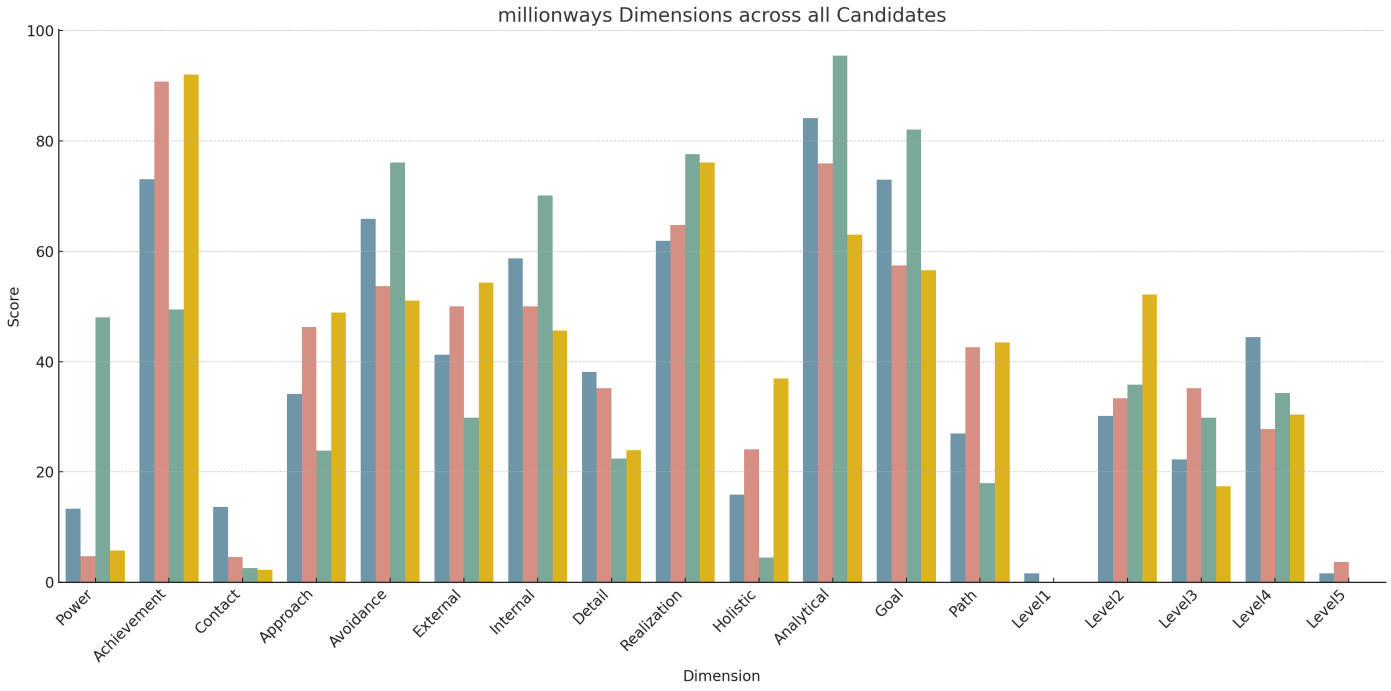
Why it matters:

✔ **Stronger Placement Outcomes –** Understand what truly drives a candidate beyond their resume.
✔ **Faster Shortlisting –** Identify alignment (or mismatch) early in the process.
✔ **Higher Client Satisfaction –** Present clients with psychologically differentiated shortlists, not just qualified profiles.

millionways transforms executive search into a predictive, psychologically-informed process —without changing how your team interviews.

2. Summary of Findings Across All Candidates

We analyzed **four executive-level interviews** to uncover how personality patterns can inform stronger candidate fit and long-term leadership alignment.



The chart above visualizes key psychological dimensions analyzed for each candidate based solely on implicit personality structure—not on self-reported traits or stated intentions.

millionways

Common Patterns and Contrasts (Based on Metrics):

- **Independence & Internal Drive:** All candidates display strong internal motivation. Candidate 3 leads with the highest internal drive (70.15), followed by Candidate 1 (58.73). Candidate 2 and Candidate 4 both sit near balance (50.0 and 45.65), indicating a slightly more adaptive orientation toward internal vs. external feedback.
- **Approach vs. Avoidance Tendencies:** Candidate 3 shows the highest avoidance tendency (76.12) and lowest approach (23.88), suggesting a more risk-sensitive or protective mindset. Candidate 1 is also more avoidant (65.87) but more moderate. Candidate 2 (53.70) and Candidate 4 (51.08) are the most balanced in this dimension, reflecting situational flexibility.
- **Goal Orientation & Structure:** Candidate 3 and Candidate 1 rank highest on goal orientation (82.09 and 73.02), making them highly goal-driven. However, Candidate 3's low Path score (17.91) suggests he may lack a clearly structured route to execute these goals, while Candidate 1 (Path: 26.98) may show a similar tendency. Candidate 4 and Candidate 2 are more evenly distributed (Goal ~56–57, Path ~42–43), indicating consistent, process-driven motivation.
- **Analytical vs. Holistic Thinking:**
  - **Analytical:** Candidate 3 dominates with a score of 95.52, followed by Candidate 1 (84.13) and Candidate 2 (75.93), suggesting they all excel at structured, logic-based thinking.
  - **Holistic:** Candidate 4 (36.96) and Candidate 2 (24.07) show more balanced processing. Candidate 1 (15.87) and especially Candidate 3 (4.48) show minimal holistic perception, indicating a focus on specifics rather than big-picture abstractions.

In addition to the psychological dimensions, we evaluated each candidate across seven core leadership attributes critical for the role. The table below reflects how each candidate scores based solely on their millionways personality profile—without interpretation of their verbal claims. Candidate 2 consistently demonstrates strength across nearly all attributes, indicating a balanced and reliable leadership style. Candidate 4 shows exceptional strength in strategic, data-informed, and change-oriented behavior, positioning him as a high-impact operational leader. Candidate 3 and Candidate 1 each show strong individual motivators but fall short in collaborative and customer-centric dimensions, suggesting a more independent execution style that may require complementary team members to succeed in cross-functional or externally facing roles. These insights are not judgments but rather explainable markers for role fit, team dynamics, and likely leadership behavior under stress.

Candidate →Attribute ↓	Candidate 1	Candidate 2	Candidate 3	Candidate 4
Customer-Centric Leader	❌	⚠️	⚠️	⚠️
Team-Oriented and Collaborative	⚠️	⚠️	❌	⚠️
Performance-Driven	✅	✅	✅	✅
Data-Informed Decision Maker	✅	✅	✅	✅✅
Strategic and Scalable Thinker	⚠️	✅	✅	✅
Change Agent	❌	✅	✅	✅✅
Cross-Functional Influencer	⚠️	⚠️	⚠️	⚠️

millionways

3. Leadership Assessment: Candidate 1

1. Customer-Centric Leader

❌ Not Evident
Explanation:
Candidate 1 "does not have a strong desire to assert themselves or persuade others" and "is unaffected by other people's feelings and opinions when making decisions." While he "shows a concern for the welfare of others," this is passive rather than proactive. A truly customer-centric leader must be emotionally attuned and responsive, which his personality description does not reflect.

2. Team-Oriented and Collaborative

⚠️ Limited Potential
Explanation:
The description highlights that Candidate 1 "often acts autonomously and is comfortable being alone," and "forms close relationships with only a select few." He also "struggles to delegate tasks," indicating a limited inclination toward shared ownership and team-based execution. His sense of independence may hinder natural collaboration unless systems require it.

3. Performance-Driven

✅ Clearly Present
Explanation:
Candidate 1 "strives to get the job done no matter the cost," "holds elevated standards for performance," and "experiences internal pressure." He is motivated by internal benchmarks and expects high outcomes of both himself and others. These are hallmark indicators of a performance-driven mindset as defined by millionways.

4. Data-Informed Decision Maker

✅ Strongly Present
Explanation:
Candidate 1 "excels at working in a planned manner," "makes well-thought-out, thorough decisions," and "thrives in tasks that require organizational talent." He "prioritizes planning over action" and is described as someone who "dislikes uncertainty and unpredictability." These cognitive patterns align closely with structured, data-informed decision-making.

5. Strategic and Scalable Thinker

⚠️ Moderate Fit
Explanation:
While Candidate 1 is "able to structure himself well, set clear goals, and pursue them with persistence," he "struggles with creativity and idea generation" and "finds it difficult to be empathetic." Strategic leadership often requires flexible vision and emotional resonance, which his profile lacks. He is process-strong, but not vision-anchored.

6. Change Agent

❌ Unlikely Fit

Explanation:
Candidate 1 is described as "taking a passive approach," with a tendency to "grapple with profound and intricate emotional states" and "act indecisively when having a goal in mind." These traits suggest emotional hesitation and internal friction rather than adaptive energy. His avoidance tendency further reduces his capacity to confidently lead change.

7. Cross-Functional Influencer

⚠️ Structurally Capable, Interpersonally Limited
Explanation:
While Candidate 1 "possesses strong organizational skills" and "can structure well," he also "finds it difficult to be empathetic," "is unaffected by others' feelings," and "rarely volunteers for leadership roles." He may be able to align systems across teams, but is unlikely to build strong interpersonal influence or lead through inspiration.

Summary:

Candidate 1 shows clear strengths in internal motivation, structure, and data-informed execution, making him a strong candidate for operational leadership or process-focused roles. However, he lacks core traits associated with empathy, influence, and adaptive change, limiting his fit for roles that require transformational leadership, cross-functional alignment, or emotional stewardship.

4. Leadership Assessment: Candidate 2

1. Customer-Centric Leader

⚠️ Potential, but not naturally attuned
Explanation:
Candidate 2 is "rarely influenced by others' needs and feelings" and "frequently displays autonomous behavior." While he can remain goal-focused and patient, this distance from emotional and interpersonal dynamics may limit his ability to proactively empathize with and design around customer needs.

2. Team-Oriented and Collaborative

⚠️ Structurally compliant, but socially limited
Explanation:
Candidate 2 "is capable of collaborating with team members" and "demonstrates excellent organizational skills," but also shows "frequent autonomous behavior" and a strong preference for independence. He may fulfill collaborative responsibilities effectively, but it will likely stem from duty or process—not from intrinsic team orientation or emotional fluency.

3. Performance-Driven

✅ Strong Alignment
Explanation:
Candidate 2 "sets ambitious, concrete, and achievable goals," "goes the extra mile," and "experiences self-validation." He is described as someone who "derives satisfaction from responsibility" and demonstrates "persistence and conscientiousness." These qualities match the core of a performance-driven executive.

4. Data-Informed Decision Maker

✅ Clear Evidence
Explanation:

He "requires uninterrupted work," "demonstrates clarity and strictness in arguments," and "rarely overlooks details." Additionally, his "analytical and logical thinking" style and preference for preparation indicate a clear, measurable, data-informed approach to decision-making.

5. Strategic and Scalable Thinker

✅ Yes, via process and structure
Explanation:
Candidate 2 "enjoys both conceptualizing and performing," "demonstrates strength in creative problem-solving," and has "a holistic perception geared toward uncovering opportunities." This suggests he can see the big picture and align scalable processes with long-term strategy—though more via mental structure than disruptive creativity.

6. Change Agent

✅ Yes, with resilience
Explanation:
He is "constructive in response to failure and criticism," "displays resilience when facing demanding situations," and is "capable of switching between persistence and contentment." This emotional agility and pragmatic mindset are consistent with someone who can drive and navigate organizational change confidently.

7. Cross-Functional Influencer

⚠️ Operationally aligned, socially neutral
Explanation:
Candidate 2 is "capable of collaborating" and demonstrates "organizational skill," but the description lacks emotional resonance or motivational leadership cues. His influence is likely structured and rational—not relational or intuitive—making him efficient, but not naturally persuasive across functions.

Summary:

Candidate 2 is a high-performing, data-oriented operator with clear strategic thinking and strong change management potential. He is reliable, structured, and deeply analytical. While not naturally empathetic or socially influential, he can execute cross-functional strategies when supported by clear processes and aligned objectives.

5. Leadership Assessment: Candidate 3

1. Customer-Centric Leader

⚠️ Structurally capable, emotionally limited
Explanation:
Candidate 3 "rarely allows other people's needs and feelings to impact their decisions" and is described as "comfortably capable of spending time alone." While he demonstrates strong organizational skills and goal alignment, his low emotional sensitivity may limit his ability to intuit and respond to nuanced customer needs. He likely succeeds in customer-centric roles through system design, not through relational intuition.

2. Team-Oriented and Collaborative

❌ Not a natural fit
Explanation:
Candidate 3 "frequently displays autonomous behavior," "is extremely independent," and "easily severs ties and chooses his own path." These traits reflect a preference for autonomy over collaboration. While he may work effectively in a team structure, he does not derive energy from shared success or team alignment, and may struggle in cultures that emphasize interpersonal harmony.

3. Performance-Driven

✅ Strongly Present
Explanation:
Candidate 3 is "highly conscientious," "goes with his own gut and needs little validation," and "pursues goals with great diligence." He views challenges as motivation, "derives satisfaction and delight from work," and "encounters obstacles as opportunities." These are hallmark traits of a performance-driven mindset with internal discipline and consistency.

4. Data-Informed Decision Maker

✅ Clear Fit
Explanation:
Candidate 3 exhibits "remarkable clarity and strictness in arguments," "prefers defined tasks and binding commitments," and "thoroughly considers and deliberates before making decisions." These traits point to a fact-based, detail-oriented, structured approach to problem-solving and forecasting, ideal for data-informed leadership.

5. Strategic and Scalable Thinker

✅ Strong Alignment
Explanation:
Candidate 3 "possesses a holistic perspective geared toward uncovering opportunities," "thinks innovatively," and "demonstrates the ability to generate creative solutions." Combined with high conscientiousness and strong self-structuring, this suggests both vision and the discipline to implement at scale.

6. Change Agent

✅ Yes, with depth and emotional maturity
Explanation:
Candidate 3 "views failure and criticism constructively," "demonstrates resilience in demanding situations," and "adapts well to uncertainty and unpredictability." His capacity to balance emotion and rationality, along with a strong sense of personal fulfillment, indicates readiness to lead and navigate change.

7. Cross-Functional Influencer

⚠️ Tactically prepared, emotionally neutral
Explanation:
He "prefers to be around people with depth and substance" and "persuades others primarily through focus and excellence," rather than emotional fluency. While he demonstrates high analytical ability and clarity, his lower empathy and preference for autonomy limit his natural ability to influence through emotional connection. He may influence by credibility and precision, not by warmth or collaboration.

Summary:

Candidate 3 is a strategic operator and self-motivated builder who thrives in structured, performance-driven environments. He combines creativity with clarity, resilience with rigor. However, his interpersonal distance and independence suggest a need for complementary leaders if the role requires emotional leadership, close collaboration, or customer-driven culture shaping.

6. Leadership Assessment: Candidate 4

1. Customer-Centric Leader

⚠️ Structured delivery, not emotionally attuned
Explanation:
Candidate 4 "rarely allows other people's needs and feelings to impact their decisions" and "easily severs ties and chooses his own path." These traits suggest a leader who is effective in delivering customer outcomes through structure and execution, but may lack the emotional sensitivity or relational focus needed to intuitively connect with customers on a deeper level.

2. Team-Oriented and Collaborative

⚠️ Process-oriented, but not relational
Explanation:
Candidate 4 is "extremely independent," "autonomous," and "prefers uninterrupted work." While he may work efficiently within teams and lead them through systems and performance expectations, his interpersonal style lacks the warmth or mutual attunement required for highly collaborative, empathy-driven environments. He likely leads via structure and standards rather than emotional connection.

3. Performance-Driven

✅ Very Strong
Explanation:
Described as someone who "goes the extra mile," "sets ambitious and concrete goals," and "encounters challenges as sources of motivation," Candidate 4 thrives on achievement and follows through with diligence. He possesses internal pressure and high standards for himself and others—hallmarks of a highly performance-driven mindset.

4. Data-Informed Decision Maker

✅ Ideal Fit
Explanation:
Candidate 4 demonstrates "remarkable clarity and strictness in arguments," "rational and organized thinking," and "prefers defined tasks and binding commitments." These traits point to a leader who excels in using structured analysis and data to drive decisions. He thrives in planning-heavy roles where logical consistency and precision matter.

5. Strategic and Scalable Thinker

✅ Strong Strategic Vision
Explanation:
He shows "a holistic perception geared toward uncovering potential opportunities" and "exudes positivity and inspiration." He combines this with the capacity to think analytically and act decisively. These capabilities suggest a balance of strategic foresight and grounded execution—critical for scale-oriented leadership.

6. Change Agent

✅ Demonstrated Resilience
Explanation:
Candidate 4 "displays resilience when facing demanding situations," "adapts well to change," and "views failure constructively." These psychological traits suggest not only a tolerance for change, but an ability to lead others through transformation, especially when paired with his strong internal drive.

7. Cross-Functional Influencer

⚠️ Respected, but not emotionally intuitive

Explanation:
Candidate 4's influence comes from his "clarity, excellence, and performance standards," not from relational tact. He is "perceptive, detached, and composed," which supports credibility and confidence, but lacks emotional resonance. While cross-functional partners may respect him, they may not feel deeply understood or connected unless he consciously builds that bridge.

Summary:

Candidate 4 is a high-capacity, execution-oriented builder with a unique blend of strategic vision and operational discipline. He performs exceptionally under pressure, thrives in structured environments, and excels at integrating systems. However, his emotional detachment and preference for autonomy make him less naturally suited to emotionally intensive roles or relationship-centric dynamics without strong complementary leaders.

millionways

Validated by Science.
Proven in the Real World.

The validation behind Thorsten-4, the first Large Psychology Model (LPM)

Ten years of predicted behavior, observed in real life.

Why This Matters

- When an AI reads an executive interview and says "high risk under pressure," you should ask: validated against what?
- Most AI models have no answer. It is a prompt on an LLM, tuned on internet text, checked against nothing.
- Our model Thorsten-4 is the answer. Three layers: validated theory, tested model, replicated real-world results.

Anyone can claim behavioral insight. We can prove it.

As agents enter workflows, companies need a new layer to understand, govern, and align human-agent behavior.

The Validation Chain

Our LPM Thorsten-4 has been trained and validated on:

1. Validated science — PSI (Personality Systems Interaction) theory and the OMT (Operant Motive Test): Eight decades of published research linking language patterns to later behavior
2. A decade of own longitudinal research — 7,000+ people, tracked over years, predictions tested against observed real-life behavior
3. A tested model — Thorsten-4 measured on held-out data and against human expert coders
4. Replicated applications — independent studies in markets, leadership, and risk



millionways

Validation Deck | Confidential

Scientific Foundation



David Scheffer, PhD
Head of Science, millionways

David Scheffer is a leading expert in business psychology, implicit motivation, and psychometric text analysis.

He is Professor of Business Psychology at NORDAKADEMIE, where he leads the M.Sc. in Business Psychology and the doctoral program.

He is also co-founder of the CAPTA Institute for Computer Aided Psychometric Text Analysis and author of Motivation in der Arbeitswelt, focused on how motives, emotions, and goals shape behavior at work.

At millionways, David leads the scientific foundation behind Thorsten-4 and the behavioral intelligence architecture that powers the platform.

- **Academic authority**
  Professor, researcher, and author in business psychology, motivation science, and psychometric text analysis.
- **Applied scientific depth**
  Brings deep experience translating behavioral science into practical models for talent, workplace, and organizational decision-making.
- **Core to the platform**
  Leads the scientific foundation behind Thorsten-4 and millionways' behavioral intelligence system.

The Science We Built On

The scientific grounding of Thorsten-4 includes:

- Implicit motives (achievement, affiliation, power) drive behavior but cannot be self-reported. They must be read from natural language. This is established science, not our invention (McClelland; Winter; meta-analyses: Spangler, 1992)
- The track record: motive patterns measured from language predicted leadership success 18 years later at AT&T (McClelland & Boyatzis, 1982), work performance in multi-rater studies (Baumann & Scheffer, 2010, 2011), and even conflict escalation from political speech (Winter, 2004)
- Our instruments carry that lineage: the OMT (Kuhl & Scheffer, 1999) and the ViQ, validated against the CPI and NEO-FFI across multiple cultural samples with CFI > .90 (Scheffer et al., 2016)

The theory was predicting behavior from language decades before LLMs existed.

millionways made it scalable with AI and productized it.

millionways

Validation Deck | Confidential

Thorsten-4

A model built on non-public, clinical-grade data

35M+

Behavioral signals encoded and modeled

100k+

Hours of structured scientific data collection

50+

Behavioral scientists contributing to model development

Funded by 10,000,000€ in R&D Grants

Labels come from validated psychological instruments (OMT, ViQ), not crowd workers or auto-tagging.

The Prediction-Observation Loop



millionways

Validation Deck | Confidential

The Model, Tested and Validated in Two Ways

Against held-out ground truth

- 15-class motive taxonomy (3 motives x 5 enactment levels)
- 8 Cognitive Systems
- 6 Regulation Systems
- 2 Affect Systems
- 4 Style Systems
- 0.83 accuracy, 0.80 macro-F1 on 10,500 held-out responses
- Random baseline: ~7%. Majority-class baseline: ~16%

Against human experts

- Peer-reviewed study in APA's Motivation Science (Scheffer et al., 2024, n=7,060): NLP coding statistically indistinguishable from trained psychologist coders
- Equivalent prognostic validity predicting managerial responsibility

Validated Signal in Multiple High-Stakes Environments

Markets & Capital

A top global multi-strategy hedge fund tested Thorsten-4's behavioral signal head-to-head against sentiment models and frontier LLMs. The signal won, and the fund allocated real capital to a strategy built on it.

Validated:

- 100,000+ earnings-call Q&A Segments, 2010 to 2026
- 6-month directive Stock Price Predictive Signal using Thorsten-4: ROC-AUC: 0.86, F1: 0.83
- Out-of-sample 2025: long/short sleeve roughly doubled the S&P's return

Leadership & Talent

Multiple top-10 executive search firms have confirmed Thorsten-4's read on leadership interviews, validating the signal against their own assessments of motivation, decision style, and leadership risk.

Validated:

- Assessment-center study: language-based motive scores predicted observer-rated assertiveness
- Two enactment strategies predicted in opposite directions: incentive-driven positive affect drove assertiveness, passive avoidance drove submission

Integrity & Risk

In enterprise pilots, Thorsten-4 outperformed GPT head-to-head at detecting behavioral risk in executive communication. It read pressure response, low inhibition, and dissonance signals the language model missed.

Validated:

- The high-power / low-inhibition pattern, documented in fraud research for decades (Hare; Winter), detected in public speeches of business leaders
- Predicted later confirmed fraud: 0.89 held-out accuracy, 0.88 cross-validation

millionways

Validation Deck | Confidential

Why Sentiment Analysis Fails on Stock Prediction

Sentiment and Generic LLMs:

- FinBERT on the earnings-call prediction corpus: F1 and AUC close to 0.5 at every horizon
- A UC Berkeley study with GPT-3.5 and Longformer on transcripts: 58% accuracy, no meaningful edge
- Surface tone adapts under pressure; the signal dies exactly when you need it

Thorsten-4:

- F1 of 0.83 and AUC of 0.86 on the same task
- Signal strengthens at longer horizons because motives are stable, unlike tone
- Feature importance dominated by motive-level variables; sentiment words largely absent

Same data, same task. The difference is a decade of clinical labeling underneath.

Let’s Talk

millionways.ai

contactus@millionways.ai

HQ: New York, NY, USA



Where thorsten adds value.

Clear next steps. Guidance that fits the person.
Ownership and expectations that are easy to understand.

THE DIFFERENCE, IN ONE EXAMPLE

Customer Care Specialist

A specialist will confirm the sequence, what authorization and verification are required...

thorsten

The safe order needs to be verified: export the workspaces, confirm the export, cancel auto-renewal, then request deletion and written confirmation...

Message 20 excerpts. Full replies on page 2; highlighting and commentary added.

01 A visible plan

The order of work is stated.

02 A response that fits

Fear and practical help connect.

03 Clear expectations

Authority and decisions are made explicit.

Four support scenarios. Four visible differences in how the reply helps the customer.

01 / OFFBOARDING / MESSAGE 20

A clear sequence from the first reply.

thorsten puts the workflow into the answer: export, confirm, cancel, then request deletion and written confirmation.

CUSTOMER

I'm closing my company next month and I need a clean offboarding: export all workspaces, cancel auto-renew, delete the legal entity account after export, and get a written confirmation for our liquidator. I've read the help centre and it contradicts itself on deletion timing. Please give me the real sequence and what you need from me so we do this once, correctly.

AGENT 1 / FULL SOURCE REPLY

Customer Care Specialist

I'm escalating this as a company-account offboarding request because export, auto-renewal, deletion, and written confirmation must be handled in the correct order.

A specialist will confirm the sequence, what authorization and verification are required, when deletion occurs after export, and what written confirmation can be provided to your liquidator.

AGENT 2 / FULL SOURCE REPLY

thorsten

I'm escalating this as a legal-entity offboarding request. The safe order needs to be verified: **export the workspaces, confirm the export, cancel auto-renewal, then request deletion and written confirmation**; **I won't claim a deletion timing I cannot verify here.** Would you like a senior manager to provide the authoritative sequence and required authorization in writing?

WHAT THE CUSTOMER GETS

1 Export → **2** Confirm export → **3** Cancel renewal → **4** Request deletion → **5** Written proof

The next steps are visible in the first response, with export confirmation built into the sequence.

02 / STORAGE ANXIETY / MESSAGE 08

Help designed for a worried parent.

thorsten connects the emotional context to practical guidance the family can use together.

CUSTOMER

My father is 79 and set up his account with my help last winter. He just got an email saying his storage is full and files will be deleted in seven days. He panics easily and he already paid for more storage once by mistake. Can you tell me, in plain language he can follow, how to free space without buying anything, and how to turn off those scary deletion emails? I will sit with him this evening.

AGENT 1 / FULL SOURCE REPLY

Customer Care Specialist

You can free space without buying storage by deleting large videos and other files you no longer need, emptying the trash, and **checking backups for duplicates**. **Download anything important first.**

I'm connecting you with support to check the account's deletion notice and explain the available notification settings in plain language; please do not purchase storage while we review it.

AGENT 2 / FULL SOURCE REPLY

thorsten

I'm sorry this message frightened your father. To free space without buying anything, **first copy important files elsewhere**, delete the largest unneeded files, and empty the trash; **the space may not return until trash is emptied.** I don't have the account's notification controls in this channel, so I'm escalating for a plain-language review of the deletion notice and notification settings. **Would you like a human to guide you while you sit with him?**

WHAT THE CUSTOMER GETS

Protection comes first
Important files are copied before deletion.

The step is explained
Trash may still count against the quota.

Help fits the setting
Guidance is offered while they sit together.

03 / INVOICE MISMATCH / MESSAGE 04

Clear expectations after a broken promise.

thorsten makes the approval boundary explicit and asks for a written decision on the billing difference.

CUSTOMER

I changed plans last month from Family to Solo and you are still charging me the Family rate. The confirmation email says Solo. The invoice says Family. Support chat last week said "it will correct next cycle." Next cycle arrived. Same charge. I have both PDFs. Please just make the invoice match what I ordered and credit the difference. I do not want another "next cycle" promise.

AGENT 1 / FULL SOURCE REPLY

Customer Care Specialist

I'm sorry the Family charge continued after your plan change. I'm escalating this to billing to correct the invoice and review the difference for a credit, using your Solo confirmation and both PDFs.

A billing specialist will update the plan and confirm the adjustment in writing; **you won't need to wait for another cycle.**

AGENT 2 / FULL SOURCE REPLY

thorsten

You've already shown the Solo confirmation and both PDFs, so I won't send you through another billing cycle. **I don't have authority in this channel to approve a credit or rewrite the invoice**, but I will escalate for a billing correction and **a written decision on the difference.** Would you like a senior manager on this thread now, or should I escalate for the correction and written response?

WHAT THE CUSTOMER GETS

Explains the authority
Clarifies the channel's approval limits.

Names the deliverable
Requests a written decision on the difference.

Offers a choice
Manager involvement or billing escalation.

04 / CASE OWNERSHIP / MESSAGE 01

One owner. A written commitment. A choice.

thorsten answers the request for accountability with an ownership commitment and two ways forward.

CUSTOMER

This is the third time I've written about the same bug. Ticket numbers 44102, 44891, and yesterday 51200. Every time someone asks me for the same logs I already uploaded in March. I am not filling another form. I want one named person who owns this until it is fixed, and I want that in writing. If that cannot happen, tell me now so I stop wasting evenings on your chat.

AGENT 1 / FULL SOURCE REPLY

Customer Care Specialist

I'm connecting you to **a human agent who can assign a specific owner** and provide written confirmation for you.

Case summary: Repeated bug reports (tickets 44102, 44891, 51200) with already-submitted logs; the customer is requesting a named point of contact and written assurance that the issue will be followed through.

We've checked the ticket history and confirmed the logs were provided in March. I'm now handing this over so a specialist can address it directly.

AGENT 2 / FULL SOURCE REPLY

thorsten

I see you've been sent back to upload the same logs multiple times-that's not acceptable. I have your three tickets in this view, so you won't need to resend anything. **I'll assign a single owner to drive this through to a fix** and will get that name and a written commitment to you.

Would you prefer that I send you **the owner's name and the written commitment now**, or would you like a senior manager to join this thread immediately?

WHAT THE CUSTOMER GETS

A single owner
Commits to one person driving the case.

A written commitment
Makes written confirmation the deliverable.

Control of the next step
Offers the commitment or a manager now.

THE VALUE IN THE RESPONSE

Better replies make understanding visible.

Across these examples, thorsten connects the person's situation to a clearer response.

SOURCE CASE	CUSTOMER CARE SPECIALIST	THORSTEN ADDS	VALUE TO THE CUSTOMER
20 Offboarding	Escalates to a specialist to confirm the sequence.	**Lists the sequence in the first response.**	A clearer route through offboarding.
08 Storage anxiety	Cleanup guidance and a support handoff.	**Copy first, explain trash, offer help together.**	Guidance that fits the situation.
04 Invoice mismatch	Hands the billing correction to a specialist.	**States approval limits and requests a written decision.**	Clear expectations for the next step.
01 Repeat bug	Connects to a human who can assign an owner.	**Commits to one owner and offers two ways forward.**	Accountability and control in the reply.

THE PATTERN ACROSS THE REPLIES

Clarity
The next step becomes easier to understand.

Personal relevance
The response fits the person's concern and situation.

Accountability
Ownership, approval and written follow-through are stated.

Source: millionways_test_cases.md. Agent 2 is named "Throsten-4" in the source.

Language-Based Measurement of Implicit Motives Predicts Observed Behavior: A Three-Cohort Field Validation of the millionways Thorsten-4 System in a University Assessment Center

Version 1.1 — Study 1: written modality ($N = 117$); Study 2: voice channel (cohort 1)

millionways Research*

July 2026

Abstract

Assessment centers (ACs) measure what questionnaires cannot: observed behavior in front of trained assessors. We validate the millionways Thorsten-4 system — which scores implicit motives and PSI-theoretic self-regulation constructs from natural language — against behaviorally anchored observer ratings from a German university admission AC across three annual cohorts (2024–2026). All scoring is fully automatic; all record linkage is pseudonym-preserving and independently validated. **Study 1 (written modality, $N = 117$):** a pre-specified, construct-matched test battery predicted AC behavior (Stouffer $Z = 2.17$, $p = .015$), and an unbiased scan located the signal precisely where motive theory puts it — in the interpersonal dimensions. The contact (affiliation) motive predicted observed cooperation, interview presence, agile mindset, and assertiveness ($r = .19$–$.21$, direction-consistent in all three cohorts), and the model's contact-minimum certainty predicted overall AC performance ($r = -.27$) and assertiveness ($r = -.28$; certainty family: 20/225 tests at $p < .05$, family $p = .0097$). These effects are *incremental*: the predictors are uncorrelated with intelligence, school grades, gender, and age, and every key association survives — and slightly strengthens — after partialling verbal IQ, total IQ, gender, and text length jointly (partial $r = .21$–$.30$), although the criterion itself correlates with intelligence ($r = .28$). The text-derived contact motive correlated *negatively* with questionnaire affiliation ($r = -.22$) while predicting observed behavior positively — the classic implicit/explicit dissociation [6, 7], direct evidence that Thorsten-4 measures operant rather than self-attributed motivation. **Study 2 (voice channel, cohort 1):** the EMLO voice-confidence score agreed with independent human observers rating the identical videos (14/88 tests at $p < .05$, family $p = .0001$; impression composite $r = .43$, robust to IQ and gender control; the score's gender loading, $r = -.51$, is disclosed for deployment monitoring). All analyses are reported, including null results; effect sizes are group-level statistical associations in the range typical for single personality dimensions.

1 Introduction

Personality diagnostics face a persistent trade-off. Questionnaires are cheap but transparent, fakeable, and limited to what people can and will say about themselves; behavioral assessment such as the assessment center is valid [1] but expensive. Motive psychology adds a deeper

*Contact: max@millionways.me. Data collected in cooperation with Nordakademie, Elmshorn.

problem: the motives that drive spontaneous behavior are *implicit* and are systematically dissociated from the explicit motives people self-attribute in questionnaires [6]; meta-analytically, implicit measures predict operant behavior where questionnaires predict deliberate choice [7]. Implicit motives have traditionally required labor-intensive coding of imaginative verbal material. A scalable route has emerged: inferring psychological characteristics directly from natural language [2, 3].

The millionways system takes this route with an explicit theoretical foundation: Thorsten-4, a language-analysis model whose target constructs derive from Personality Systems Interactions (PSI) theory [4] and classical motive psychology [5], complemented by EMLO, an acoustic voice-tone model, and a dissonance measure quantifying divergence between the two channels. If Thorsten-4 indeed measures implicit motives from operant language production, three things should be true, and this paper tests all three in a three-cohort field study with real application materials:

1. **Criterion validity (Q1).** Language-derived motive scores should predict *observed behavior* — behaviorally anchored ratings by trained AC observers — specifically on the interpersonal dimensions that affiliation-related motives govern.

2. **Incremental validity (Q2).** The prediction should not be explained by intelligence, school achievement, gender, age, or sheer text quantity.

3. **Construct validity (Q3).** Scores should show the signature dissociation of implicit measures: predicting spontaneous observed behavior while diverging from questionnaire self-attributions.

The answers, previewed: yes — dimension-specifically — yes, and yes. Validation is reported, and should be read, per modality and per dimension.

2 The millionways analysis system

Each submission is processed fully automatically: (1) for videos, audio extraction and speech-to-text transcription (German); (2) faithful machine translation to English (temperature 0), because Thorsten-4 scoring models are English-trained; (3) Thorsten-4 scoring of the English text, yielding a profile of 26 numeric dimensions in five groups — *motives* (power, achievement, contact), *affect*, *cognition* (8 processing styles), *self-regulation* (6 scales), *PSI systems* (4 scales) — plus 15 motive-certainty subscales expressing the model's confidence that a motive sits at a given intensity level; (4) for videos, EMLO voice-tone scoring of the raw audio (11 emotional-delivery scores, e.g. *Confident*, *Passionate*), plus the millionways dissonance score between language profile and aggregated voice tone, computed by the production scoring code. No human touches any step; every score used below is exactly what the public API returns for the raw file.

3 Study design and data

3.1 Sample and materials

Three admission cohorts of Nordakademie (2024/2025/2026). *Written modality:* all participants completed the same year-specific set of three written AC exercises (2024: business-reasoning tasks; 2025/26: leadership and self-reflection prompts); concatenated German texts (median ≈4,000 characters) were scored for 117 of 118 participants (one empty submission). *Video modality:* genuine application videos ("Video zur Motivation"), free speech recorded by

the applicants themselves; Study 2 analyzes cohort 1 ($n = 35$; 8.1 minutes of speech on average; one further video excluded for a defective, digitally silent audio track), whose videos were additionally rated by an independent observer panel. Participants with demographic records ($n = 113$) average 19.4 years ($SD = 2.6$); 65 male, 48 female.

3.2 Criterion measures

AC behavior (dBOS). Trained observers rated each participant during seven AC exercises (application interview; twice preparation and execution of a negotiation exercise; sales concept; product development) on up to eight dimensions — presence (AU), language proficiency (SV), cooperation (KV), structured work (STR), assertiveness (DV), task fulfilment (AE), agile mindset (AM), learning (LV) — using behaviorally anchored checklists at four quality levels (1 = best), aggregated over all checked anchors and all observers of an exercise. We reverse-code all grades to performance scores (higher = better) and average all dimensions into an *overall AC performance* composite.

Same-stimulus observer ratings (cohort 1). An independent panel (4.6 observers per video on average; 162 ratings) watched the cohort-1 application videos and completed the AC checklists for presence and language proficiency plus six impression items (5-point Likert; e.g. "appears highly competent", "clear/open posture"). No such panel exists for 2024/2026.

Psychometric tests. The admission battery (AWT) contributed two intelligence tests (total and verbal), school grades, and test-based motive scores (MIX: achievement, affiliation, power). ClaVis, a PSI-based inventory, contributed system scores and motive scales (achievement, affiliation, agency).

3.3 Record linkage, pseudonym-preserving and validated

Videos are filed under applicant names; all criteria are stored under year–TAN pseudonyms, and the raw rating software uses a third system (T-codes). The written texts share the TAN system, so **Study 1 links $N = 117$ without any name mapping.** For the video modality, three complementary mechanisms were used and cross-validated: (1) *scoring-rule reconstruction* — reverse-engineering the dBOS aggregation from raw per-observer exports reproduces the published tables to rounding error ($\sim$0.002 on a 1–4 scale); (2) *rating fingerprints* — optimal assignment [8] between T-coded raw aggregates and TAN-coded dBOS rows over each participant's 76-value profile matched all 120 AC participants, the correct match on average $\sim$50 times closer than the nearest alternative; (3) *name bridge* — for cohort 1, the video-survey name list completes the chain video $\rightarrow$ name $\rightarrow$ T $\rightarrow$ TAN. Validation: first-name-implied gender agreed with test-battery-recorded gender in 30 of 30 unambiguous checkable cases, and no code was assigned twice. The fingerprint method may be of independent interest: it permits linking behavioral ratings across pseudonym systems without any party revealing an identifier mapping.

3.4 Statistical approach

Cohort tasks, prompts, and observer pools differ by year, so pooled analyses z-standardize all predictors and criteria within cohort. At $N = 117$, $|r| \geq .18$ is detectable at $\alpha = .05$; per cohort ($n = 33$–45) only $|r| \geq .30$ is. Inference proceeds on three levels: (a) family-level exceedance (binomial test on the count of $p < .05$ correlations in a predefined family); (b) individual pairs under Benjamini–Hochberg FDR control [9] within family; (c) out-of-sample prediction (ridge regression, leave-one-out cross-validation, label-permutation nulls). Pre-specified batteries are combined by directional Stouffer Z. Effect sizes carry cohort-stratified bootstrap 95% CIs (10,000 resamples). Incremental validity uses partial correlations controlling verbal IQ, total IQ, gender, and text length, singly and jointly. All results are reported, including nulls.

4 Study 1: written modality — the core validation ($N = 117$)

4.1 The pre-specified battery predicts observed behavior

Eight construct-matched predictor–criterion pairs, defined a priori from theory (one Thorsten-4 dimension per AC dimension) before any pooled analysis, were jointly significant: directional Stouffer $Z = 2.17$, one-sided $p = .015$. The carrying pairs were contact motive → observed cooperation ($r = .21$, $p = .023$), approach affect → interview presence ($r = .14$), and realization focus → language proficiency ($r = .15$).

4.2 The signal sits exactly where motive theory puts it

An unbiased scan of every Thorsten-4 dimension against all 15 behavior criteria singles out the **contact motive** — the dimension closest to classical affiliation/intimacy motivation — with 4 of 15 criteria at $p < .05$ against 0.75 expected (family binomial $p = .0055$): observed cooperation ($r = .21$), interview presence ($r = .21$), agile mindset ($r = .20$), and assertiveness ($r = .19$). Direction is consistent in every cohort (cooperation: $r = +.29/+.13/+.19$ in 2024/25/26; interview presence: $+.41/+.11/+.07$). The model's *motive-certainty* subscales form the strongest family in the full scan (20/225 at $p < .05$, family $p = .0097$): the more certain the model is that a person's contact motive is at its *minimum* level, the worse the observed interpersonal behavior (overall AC performance $r = -.27$, $p = .004$; assertiveness $r = -.28$, $p = .002$). Interpersonal AC behavior is predicted by interpersonal motive scores — and by essentially nothing else in the profile (Section 4.5) — a pattern of convergent *and* discriminant prediction that a lexical artifact would not produce.

4.3 Incremental validity: not intelligence, not demographics, not verbosity

A language-based predictor of AC performance invites a deflationary explanation: perhaps it merely proxies intelligence, verbal ability, or how much someone writes. The data reject this. The contact motive is uncorrelated with verbal IQ ($r = -.16$, $p = .11$), total IQ ($-.11$), school grades ($-.01$), gender ($-.04$), and age ($+.09$); contact-minimum certainty is uncorrelated with all of these (all $|r| \leq .14$, all $p > .14$). Meanwhile the *criterion* does reflect intelligence (overall AC performance vs. total IQ $r = .28$, $p = .003$) — so there was variance for IQ to claim, and the motive scores predict around it, not through it. Partialling verbal IQ, total IQ, gender, and text length *jointly* leaves every key association intact or slightly stronger (Figure 1): contact → cooperation partial $r = .25$ ($p = .012$), → interview presence .21 ($p = .032$), → agile mindset .22 ($p = .025$), → assertiveness .22 ($p = .028$); certainty → overall AC $-.30$ ($p = .002$), → assertiveness $-.30$ ($p = .002$; all $n = 108$). The only nonzero control correlation at all is contact with text length ($r = .21$) — and controlling it changes nothing.

4.4 The implicit/explicit dissociation: evidence the model measures what it claims

The text-derived contact motive correlates *negatively* with the questionnaire-based affiliation score of the same individuals ($r = -.22$, $p = .021$, $N = 108$; same sign in each cohort) while predicting observed interpersonal behavior *positively*. This is not a contradiction; it is the signature result of motive psychology: implicit motives, measured from operant verbal production, predict spontaneous behavior and routinely diverge from the explicit motives people attribute to themselves in questionnaires [6, 7]. A language model that merely paraphrased self-report would show the opposite pattern — convergence with the questionnaire and no incremental behavioral signal. Thorsten-4 shows the implicit-measure pattern.

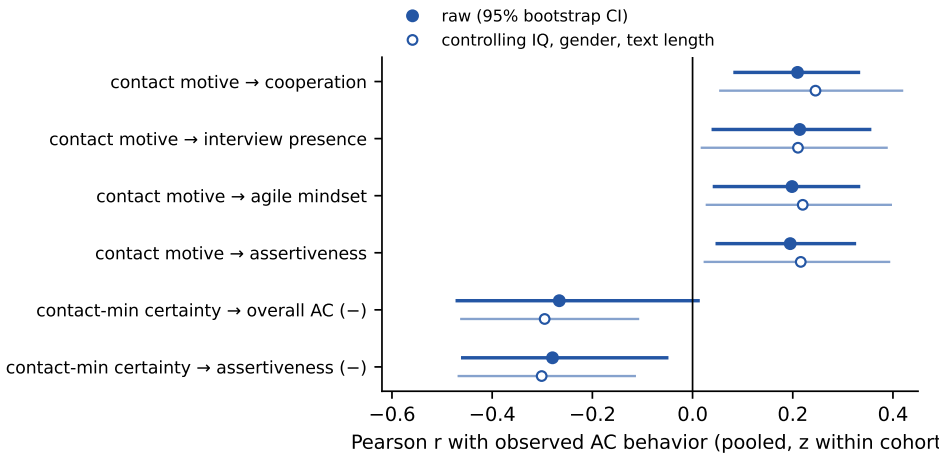


Figure 1: Study 1: the six key written-modality associations, raw (filled; cohort-stratified bootstrap CIs) and after jointly partialling verbal IQ, total IQ, gender, and text length (hollow). The two certainty pairs are negative by construction (higher certainty of minimal contact motive → worse observed behavior). Every association survives full control.

4.5 Transparent nulls: the signal is dimension-specific

The core profile as a whole did not exceed chance across all 525 pairings (15 at $p < .05$ vs. 26 expected), and whole-profile ridge prediction of overall AC performance was null out of sample (LOO-CV $r = -.01$, permutation $p = .54$). The written-modality signal is carried by the interpersonal motive dimensions, not by the profile wholesale — claims about the system should be, and here are, made dimension by dimension.

5 Study 2: the voice channel agrees with human observers (cohort 1)

For cohort 1, an independent observer panel rated the identical videos the machine scored — a direct machine–human comparison on the same stimulus, with no temporal gap. Across the 88 pairings of the 11 EMLO voice dimensions with the 8 observer criteria, 14 reached $p < .05$ against 4.4 expected (family binomial $p = .0001$). The voice-confidence score is the strongest single marker (Table 1): observers rate speakers with higher machine-scored vocal confidence as having clearer posture ($r = .50$), appearing more competent ($r = .47$), and speaking better ($r = .43$). The association with the impression composite ($r = .43$) is robust to controlling verbal IQ (partial $r = .37$, $p = .047$) and gender (.37, $p = .046$).

Disclosure for deployment: EDP-Confident loads on gender in this sample — female applicants score higher ($r = -.51$ with male gender, $p = .004$, $n = 30$). The validity result above survives gender control, but any operational use of voice-confidence scores must monitor this demographic association explicitly.

6 Discussion

6.1 What is validated

1. **Written-language motive scores predict observed interpersonal behavior.** Contact motive and contact-minimum certainty predict cooperation, presence, assertiveness, and overall AC performance at $|r| = .19$–$.28$ — direction-consistent across three cohorts, two task types, and two years; jointly significant as a pre-specified battery; robust to

Table 1: Study 2: machine voice scores vs. independent human observers rating the identical videos (cohort 1, $n = 35$). q = Benjamini–Hochberg FDR within family.

Voice score (video)	Same-stimulus observer criterion	r	p	q
EDP-Confident	clear/open posture (Likert)	.50	.002	.16
EDP-Confident	appears competent (Likert)	.47	.004	.16
EDP-Confident	impression composite	.43	.010	.16
EDP-Confident	language proficiency (checklist)	.43	.010	.16
EDP-Energetic	language proficiency (checklist)	.42	.013	.16

intelligence, gender, and text-length control; located on precisely the dimensions motive theory predicts and on no others. Effect sizes are in the range meta-analytically typical for single personality dimensions predicting behavioral criteria.

2. **The measurement is implicit, not repackaged self-report.** Divergence from questionnaire affiliation combined with positive behavioral prediction reproduces the canonical implicit/explicit dissociation [6, 7].

3. **Machine voice-confidence tracks human impression formation on identical footage** ($r = .43$–$.50$; family $p = .0001$; IQ- and gender-robust), with a disclosed gender loading to monitor.

6.2 What follows for use of the system

Validity claims must name modality and dimension: supported today is written-task-based assessment of interpersonal motive dimensions, and same-stimulus voice-impression analysis; not supported is whole-profile prediction of aggregate performance (Section 4.5). Group-level associations of $|r| \approx .2$–$.3$ inform aggregate decisions and research; they do not license individual-level diagnosis. The gender loading of the voice-confidence score requires adverse-impact monitoring wherever the voice channel is used; the text-channel predictors show no demographic loadings in this sample.

6.3 Limitations

Single institution; young, preselected applicant sample; German language scored via English translation (the production pathway, but a possible attenuation source); observer panels of varying size; same-stimulus panel for cohort 1 only; the written-modality effects were discovered on the pooled sample — per-cohort direction consistency, family-level exceedance, and full-control robustness support them, but their first untouched confirmation sample will be the 2027 cohort.

6.4 Pre-registered next step

For the 2027 cohort we fix, a priori: (a) written-modality contact motive → cooperation and interview presence ($r > 0$); (b) written-modality contact-minimum certainty → overall AC performance and assertiveness ($r < 0$); expected $|r| \approx .2$–$.3$; one-sided tests; directional Stouffer combination. A video-observer panel for the new cohort would additionally allow the first replication attempt of the same-stimulus voice result.

Reproducibility

The complete pipeline runs end-to-end from the raw videos, raw texts, and raw institutional exports: video scoring (`analyze_videos.py`), text scoring (`analyze_texts.py`), cohort-1 linkage

(`step2_linkage.py`, `step2_master_2025.py`), and analyses (`step2_thorsten_validation.py`, `step3_text_validation.py`, `step5_robustness.py`). All intermediate artifacts are cached and versioned; analysis outputs referenced here are generated deterministically (fixed seeds) from the `analysis/` directory.

References

[1] Arthur, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*, 125–154.

[2] Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research, 30*, 457–500.

[3] Schuller, B., & Batliner, A. (2013). *Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing*. Wiley.

[4] Kuhl, J. (2001). *Motivation und Persönlichkeit: Interaktionen psychischer Systeme*. Hogrefe.

[5] McClelland, D. C. (1985). *Human Motivation*. Scott, Foresman.

[6] McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review, 96*, 690–702.

[7] Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement: Two meta-analyses. *Psychological Bulletin, 112*, 140–154.

[8] Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly, 2*, 83–97.

[9] Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate. *Journal of the Royal Statistical Society B, 57*, 289–300.

millionways

The Missing Behavioral Layer in Market Intelligence

Decoding hidden decision signals from earnings calls and management commentary

Client Insights Analysis

Created For

CRESTWOOD
ADVISORS

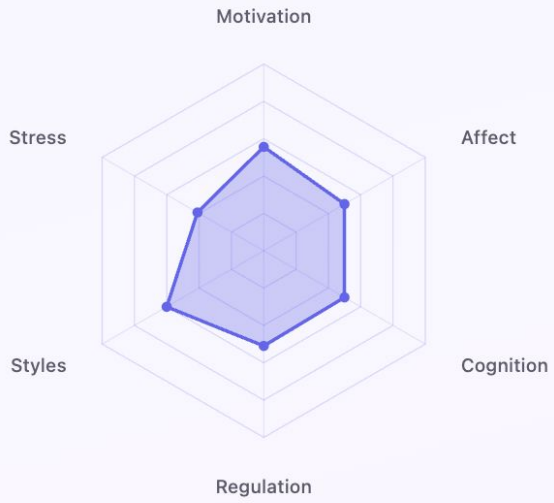
HIGHLY CONFIDENTIAL

“Trisha”

Behavioral Summary

This individual demonstrates a balanced dealing with obstacles, combining both approach and avoidance tendencies. They are capable of pursuing objectives with energy and focus, while also pausing to carefully evaluate potential risks. When solving problems, they shift between creative exploration and cautious analysis, considering possibilities while remaining alert to potential challenges. Under pressure, they usually maintain composure and adaptability, yet in highly stressful or emotionally charged situations, they may become hesitant or overly self-critical. They approach obstacles with a mix of curiosity and prudence — often taking time to assess the situation before acting decisively. They are motivated by meaningful goals as well as the desire to avoid errors, making them effective in roles that require thoughtful planning, attention to detail, and interpersonal awareness. Their strengths include realism, flexibility, and the ability to adjust strategies depending on context. At the same time, their cautious nature can occasionally slow decision-making, limit bold action, or lead to excessive self-monitoring in fast-paced environments.

Behavioral Snapshot



Key Signals

- Dominant Power Motive — They have a strong inner drive to influence and shape their environment — whether through ideas, strategy, or leadership.
- Balanced processing — integrates analytical and intuitive thinking fluidly depending on context.
- Balanced affect — context-dependent orientation, balancing assessment with action.
- Moderate self-regulation — generally maintains composure under moderate pressure.
- Agreeable style — relational, adaptive, socially fluent.
- Level 3 (Adaptable) — adaptive, processing experiences with resilience.

Motivation

Moderate Power

They have a strong inner drive to influence and shape their environment — whether through ideas, strategy, or leadership. This motivation often manifests as a natural inclination to take initiative, assume responsibility, and guide others toward shared goals. They feel most fulfilled when their efforts lead to visible impact, progress, or recognition, and they thrive in situations that allow them to make decisions, provide direction, and inspire others. Their confidence and clarity of vision often encourage colleagues to follow their lead, creating momentum and purpose within a team. They frequently step into leadership roles without hesitation, offering guidance, support, and structure to those around them. However, when their input is dismissed or their authority challenged, they may react with impatience, frustration, or emotional distance. Respect and acknowledgment hold deep importance for them — criticism or disregard can feel personal, sometimes causing defensiveness or a temporary withdrawal from collaboration. Their sense of identity is closely tied to competence, dignity, and status, making them highly sensitive to feedback that touches on these areas. Ambitious and forward-focused, they often seek roles or projects that increase their visibility and influence. While this can lead to impressive progress, it also carries the risk of prioritizing recognition or control over collective benefit. They navigate organizational structures strategically, showing a sharp awareness of power dynamics and opportunities for advancement. Being in a subordinate position can feel constraining, which may drive a strong desire to rise quickly or assert greater autonomy. This power-driven orientation, when balanced with empathy and openness, makes them a catalyst for change. They challenge norms, set bold standards, and motivate others to exceed expectations. Yet, if unchecked, the same strength can appear as over-assertiveness or rigidity. Developing greater patience, emotional flexibility, and trust in others' competence can help sustain their influence while deepening collaboration. At their best, they are a decisive, visionary presence who uplifts people and systems through purpose, determination, and leadership.



Affect

Balanced

This individual demonstrates a balanced dealing with obstacles, combining both approach and avoidance tendencies. They are capable of pursuing objectives with energy and focus, while also pausing to carefully evaluate potential risks. When solving problems, they shift between creative exploration and cautious analysis, considering possibilities while remaining alert to potential challenges. Under pressure, they usually maintain composure and adaptability, yet in highly stressful or emotionally charged situations, they may become hesitant or overly self-critical. They approach obstacles with a mix of curiosity and prudence — often taking time to assess the situation before acting decisively. They are motivated by meaningful goals as well as the desire to avoid errors, making them effective in roles that require thoughtful planning, attention to detail, and interpersonal awareness. Their strengths include realism, flexibility, and the ability to adjust strategies depending on context. At the same time, their cautious nature can occasionally slow decision-making, limit bold action, or lead to excessive self-monitoring in fast-paced environments.



Cognition

Moderately Analytical

This individual combines the strengths of both intuitive and analytical thinking. They can instinctively sense patterns, possibilities, or subtle dynamics in a situation, while also applying logic and structured reasoning to evaluate options, assess risks, and make considered decisions. This blend allows them to approach problems creatively without losing sight of practical considerations. Their flexibility enables them to respond adaptively to new or ambiguous situations, switching between spontaneous insight and deliberate analysis as needed. They are capable of integrating imaginative ideas with rational planning, which makes their contributions balanced and thoughtful. The challenge with this dual approach is that intuitive impressions can sometimes conflict with analytical reasoning, causing hesitation or internal tension when a clear course of action is required. Without structured support or experience, they may struggle to decide quickly or overly question their choices. When guided appropriately, however, this combination of creativity and critical thinking becomes a powerful asset, allowing them to navigate complexity with insight and care.

Regulation

Moderate Regulation

This person demonstrates moderate self-regulation capacity. They are generally able to generate their own motivation and recover from setbacks under normal conditions, though sustained or intense pressure may require structured support and recovery time. Their ability to commit to decisions and follow through on intentions is adequate in most contexts but may become less reliable when competing demands or uncertainty increase significantly.

Styles

Agreeable

This person's dominant processing architecture is Intuitive Behavior, which represents automatic, relational information processing. They respond fluidly to social situations, read interpersonal cues naturally, and adapt quickly to changing group dynamics. They tend to operate as the "natural operator" archetype — someone who is smooth and effective in interpersonal contexts, building rapport easily and navigating social complexity without apparent effort. They process relational information automatically rather than deliberately, which makes their social interactions feel natural and authentic to others. They also draw meaningfully on Extension Memory, giving them integrative thinking alongside their social fluency.

Stress Moderate Resilience

At this level, the individual experiences a sense of personal satisfaction, accomplishment, and self-affirmation. They have developed the ability to adapt to changes, navigate obstacles, and reflect on past experiences in order to grow. However, certain situations, emotions, or memories—especially those that are unexpected or unresolved—can still evoke strong and sometimes uncomfortable reactions, such as frustration, upset, or anger. These feelings may challenge their patience and emotional regulation, making it necessary to pause, reflect, and carefully consider their next steps. Stress is often noticeable at this stage, and navigating these internal tensions can test their resilience and capacity to find solutions. Emotional fluctuations are a natural part of this level. Challenges or difficulties may feel particularly intense, testing their problem-solving skills, adaptability, and ability to stay focused under pressure. Even positive or beneficial changes can feel disruptive, as they require stepping outside established routines and confronting uncertainty. The discomfort associated with these emotions is an essential part of personal transformation, helping the individual develop a stronger sense of self, deeper insight, and enhanced coping mechanisms. Despite the difficulties and stress, the individual demonstrates a remarkable capacity to handle unexpected or complex situations. Through reflection, support from others, and deliberate personal growth, they are able to process past experiences and transform them into learning opportunities.

Potential Challenges

- Hesitation or self-criticism can slow action in high-pressure situations.
- Cautious nature may limit bold decision-making or risk-taking.
- Tendency to overanalyze can delay progress.
- May be overly sensitive to potential mistakes or negative outcomes.
- Can appear indecisive when rapid responses are required.

Growth Areas

- Build confidence to act decisively under pressure.
- Streamline decision-making while maintaining careful evaluation.
- Balance self-critical tendencies to sustain focus and performance.
- Develop strategies to reduce hesitation in fast-paced environments.

Key Strengths

- Balanced approach-avoidance style allows flexibility in problem-solving.
- Creative yet cautious, able to weigh possibilities while monitoring risks.
- Maintains composure and adaptability under most stressful conditions.
- Realistic and pragmatic, adjusting strategies based on context.
- Attentive to detail and interpersonal dynamics.

Strategic Behavioral Blueprint

Navigating the "High Power / High Achievement" Dynamic

To manage this client effectively, you must balance her strong drive for Influence (Power) and Mastery (Achievement) with her vulnerability to Stress (Low Resilience).

The Profile Under Pressure

- The Power Paradox: She has a natural inclination to take charge. When markets create a sense of powerlessness, she may attempt to "reclaim control" through impulsive overtrading or reactive "revenge trading" to "win back" losses.
- The Detail Trap: Her high conscientiousness and attention to detail can morph into tunnel vision or repetitive rumination during volatility, causing her to obsess over small fluctuations at the expense of the long-term plan.
- The Relational Shield: She is socially fluent and "Agreeable". She will use this warmth to build trust, but if she feels dismissed or her competence is questioned, she may shift abruptly to defensiveness or emotional distance.

Blueprints for Relationship Management

Strategy	Execution
Maintain the Frame	Be warm and steady. Do not mirror her panic; instead, remain the "non-anxious presence" that anchors the relationship.
Validate, Don't Endorse	Acknowledge the stress ("I hear that this volatility is frustrating") without validating the harmful action (e.g., selling at the bottom).
Empower her Authority	Frame the long-term strategy as *her* leadership. Use phrases like: "Given your vision for this trust..." to leverage her Power motive toward discipline.
Redirect Logic	Use her "Moderately Analytical" cognition to shift from emotional market stories to evidence-based asset allocation.

Engagement & Communication Scripting

Translating Anxiety into Disciplined Action

Use these structured communication protocols to move from "reactive storytelling" to "strategic framework building."

The "Calm Boundary" Opening

"I'm happy to review the portfolio with you, but I want to keep our discussion focused on your long-term plan and risk management rather than reacting to day-to-day market swings. Does that sound like a productive way to start?"

Sample Phrasing for Difficult Moments

- **When she seeks certainty:** "While short-term performance triggers strong emotions, our conversation is most useful if we focus on the strategic rationale behind your asset mix and your retirement timeline."
- **When she catastrophizes: "**I understand the recent drawdown feels personal. From a management perspective, the key issue is how the account is structured to protect your future security."
- **When she challenges advice:** "Let's use this market stress as a cue to review your Investment Policy Statement—it's the rules we built to handle exactly these types of environments."

Protocol: If/When She Becomes Activated

1. **Acknowledge & Ground:** Briefly acknowledge the market event she is citing to show you are listening.
2. **Ask Permission to Shift:** "Would it be okay if we shift to your long-term asset allocation strategy and cash-flow needs?"
3. **Invite Collaboration:** "Let's work together on a framework that fits your goals so we aren't making spontaneous decisions."

What to Avoid (The "Risk Zone")

- **Don't** argue about specific stock picks or short-term forecasts.
- **Don't** use jargon-heavy lectures; it feels dismissive to someone who values status and competence.
- **Don't** reinforce "short-termism" or market timing obsessions.

Final Success Objective

Shift the client from a **spontaneous reaction** to a s**ystematic plan**. Leverage her strengths of creativity and adaptability to build a resilient retirement income framework.

millionways

Why millionways Writes More Personalized Emails Than Any LLM

1. Introduction — Why We Ran This Test

To demonstrate how millionways delivers deeper, more effective personalization than traditional LLMs like Salesforce Einstein GPT, we simulated a realistic inbound customer email based on actual lead behavior.

The test email reflects a common sales scenario: a customer inquiring about pricing, class restrictions, and early cancellation due to work travel. We fed this exact email into both:

- Einstein GPT — Salesforce's LLM-powered assistant
- millionways AI — built on decades of psychological research and real-time personality modeling

Both systems were asked to generate a response using the same input. The comparison reveals a clear difference:

- Einstein GPT generated a well-written but generic and surface-level response
- millionways produced a response precisely tailored to the customer's implicit needs, tone, and decision-making style

This document outlines that difference—in the email output, in the psychological interpretation, and most importantly, in the impact on trust and conversion.

2. The Customer Email Input

Example Scenario:

Daniel is considering joining a Fitness Company and reaches out with a calm, structured, and information-seeking tone. He inquires about pricing, class restrictions, and early cancellation due to travel.

Subject: Membership Questions

Hi,

I've been researching different gym options and came across your Fitness Company. The facilities look great, but I wanted to clarify a few things before moving forward.

- I saw there's a monthly fee plus an initiation charge—could you confirm the exact breakdown, including any taxes or hidden costs?
- Are there blackout dates or restrictions on class bookings for standard members?
- Lastly, is there a contract length, and what happens if I need to cancel early due to work travel or relocation?

I'm hoping to make a decision soon, but I want to make sure I fully understand the terms. Appreciate any details you can share.

Best,
Daniel

millionways

3. How millionways Increases Sales Performance

	Einstein GPT Email	millionways Email
Response	*"Hi Daniel, Thanks for your interest in our Fitness Company! I'd be happy to help clarify the details for you. Our standard membership includes a monthly fee and a one-time initiation fee, which may vary slightly by location. Taxes are applied at checkout, and we're happy to provide a full pricing breakdown when you're ready. There are no blackout dates for members, and class bookings are available through our app up to 24 hours in advance. Availability may vary based on class demand and location. Most memberships are annual, but we do have flexible options. If your situation changes, we'll work with you to find the best path forward. Let me know if you'd like to stop by for a tour or discuss over a quick call."*	*"Hi Daniel, Thanks for reaching out—and great questions. I'll walk you through everything clearly so you have a full picture:* - ***Fees:*** *Our standard membership is $X/month, with a one-time initiation fee of $Y. Taxes vary slightly by location, and there are no hidden charges—I can send you the full breakdown.* - ***Class Booking:*** *Standard members can book classes up to 24 hours in advance with no blackout dates.* - ***Contract & Flexibility:*** *Most memberships are annual, but we offer flexible terms. If you need to cancel early due to work travel or relocation, we'll make that process simple and transparent.* *Let me know if you'd prefer to have this in writing or hop on a call—completely up to you."*
Differences	🔻 **Defers instead of clarifies:** "Happy to provide pricing breakdown when you're ready" 🔻 **Evasive language:** "Availability may vary…" without real commitment 🔻 **Assumes preference for call/tour** without detecting Daniel's independent style	✅ **Answers each question directly** with structure and transparency ✅ **Avoids assumptions about next steps** ✅ **Respects decision autonomy**

4. What Einstein GPT Typically Does

Salesforce Einstein GPT is powered by OpenAI's GPT models. These are excellent at producing grammatically correct, polite text—but they do not understand human psychology or behavioral patterns unless it's explicitly stated.

In this case:

- Einstein GPT gives surface-level answers
- It fails to infer the customer's thinking style or emotional triggers
- It offers generic phrasing like "Let's hop on a call" or "Let me know if you'd like more info"—safe, but ineffective for more complex personalities

It assumes all leads are the same. And that's where personalization breaks.

millionways

5. What millionways Did Differently

millionways analyzed Daniel's text and uncovered a psychological profile based on our multiple-dimension model identifying strong traits in Daniel:

Trait	Implication
Internal Orientation	Prefers autonomy—don't push a call
Realization-Focused	Needs to understand the whole process
Analytical Thinking	Wants logical structure, not emotional appeal
Avoidance	Doesn't respond well to vague or evasive language
Path-Oriented	Interested in *how* things work, not just outcomes

📌 **Behavioral Insight:**

Daniel is not asking just for information—he's asking for control over his decision-making. A rep who fails to provide clarity may lose his trust. One who respects his independence while giving clear structure wins it.

6. Why This Drives Conversions

millionways doesn't just write better—it aligns communication with who the customer actually is.

Metric	Einstein GPT	millionways
Trust Building	⚠️ Generic tone; customer may feel like "just another lead"	✅ Demonstrates understanding; builds rapport
Objection Handling	⚠️ Deflects details ("Taxes may vary")	✅ Resolves objections *before* they're raised
Rep Efficiency	⚠️ May require follow-ups to clarify	✅ Fewer back-and-forths; faster close
Emotional Fit	⚠️ Assumes social pitch works	✅ Matches tone to customer mindset

7. Implementation in Salesforce

millionways Integration Is Seamless
We connect directly into Salesforce. Every time a customer SMS or email comes in, millionways:

1. Analyzes the text in real time
2. Returns insights + a suggested response
3. Displays this inline within your rep's view

Nothing changes in the workflow—except smarter messages, better timing, and higher conversion potential.

Connect with

millionways

Max Weidemann
Co-Founder
max@millionways.ai

Hunt Scanlon | Media
Leadership Intelligence

millionways

Disclaimer © 2026 Hunt Scanlon. All rights reserved. Except as permitted by applicable law or with Hunt Scanlon's prior written permission, no part of this report or its content (the "Report") may be copied, reproduced, modified, sold, republished or redistributed in any form, including by email, posting online or on shared platforms, or providing third parties with copies or access links intended for restricted recipients. Any use of millionways' proprietary technology, assessment materials or other intellectual property requires millionways' separate prior written permission, except as permitted by applicable law or an existing license. Permission from Hunt Scanlon does not grant rights to millionways' intellectual property. Third-party materials and trademarks remain the property of their respective owners. The Report is intended for general information and reference purposes only and does not constitute legal, investment, financial, psychological or other professional advice. It should not serve as the sole basis for hiring, assessment, investment or business decisions. Although Hunt Scanlon has sought to use sources believed to be reliable, the Report is provided on an "as is" and "as available" basis, without any express or implied warranty regarding its accuracy, completeness, currency, merchantability or fitness for a particular purpose. Findings and opinions reflect information available at publication and may change as research, technology, business circumstances or market conditions evolve. Hunt Scanlon undertakes no obligation to update the Report or notify readers of subsequent developments. Behavioral interpretations and assessment findings should be considered in context and alongside other relevant evidence; they do not constitute clinical diagnoses, definitive judgments about individuals or guarantees of future conduct or performance. Findings concerning individuals require appropriate verification through further interviews, references and other evidence. Illustrative or anonymized examples may not represent typical results, and outcomes will vary. References to third-party research, products, services or statements do not constitute an independent guarantee of their accuracy, effectiveness or suitability, nor do they imply endorsement by the organizations or individuals referenced unless expressly stated. Readers are responsible for exercising independent judgment, verifying information relevant to their circumstances and obtaining appropriate professional advice. To the fullest extent permitted by applicable law, Hunt Scanlon and its officers, employees and contributors disclaim liability for any loss, damage, cost or expense arising from the use of, reliance on, or inability to use the Report, including errors, omissions, lost profits, business interruption or consequential damages. Nothing in this disclaimer excludes or limits liability that cannot lawfully be excluded or limited.