



DEEPER
SIGNALS

A Memo on the Scientific Foundations of a Behavioral Digital Twin of the Organization

Research Whitepaper

Last updated: May 4, 2026

Ekho: A Memo on the Scientific Foundations of a Behavioral Digital Twin of the Organization

Ekho is a behavioral digital twin of each individual in an organization, the relationships among them, and the decisions they make together. Nozari and Yordanova (2026) made the case in mathematical terms that organizations need a data-driven, continuously updated workforce model to manage uncertainty. Engineering and operations have used digital twins for decades to test changes before committing to them. There have been efforts to build digital twins for human capital applications, however, they have been mostly impractical and lack fidelity due to technical constraints. Ekho innovates on previous efforts by integrating two new factors: validated psychometric data to make better predictions of future behavior, and large language model (LLM) agents that enact what each person does, to generate emergent organizational phenomena. The product hypothesis is that parameterizing LLM agents with personality characteristics will yield individual and organizational predictions that are useful enough to inform strategic talent management. This memo outlines the scientific rationale behind shaping the development of this new Deeper Signals product.

The scientific foundation of Ekho rests on three propositions, outlined below and substantiated in the sections that follow:

Proposition 1. Contextualized psychometric assessment results predict future behavior leading to leadership and job performance outcomes, and they do so more strongly than decontextualized psychometric measures.

Proposition 2. Agent based simulations powered by LLMs can simulate the behavior of an organization and the people within it, providing a usable digital twin of the firm.

Proposition 3. LLMs can be infused with stable psychometric properties so that an individual agent functions as a digital twin of a real person.

Together these propositions support the integrative claim that a simulation populated by psychometrically parameterized LLM agents (the Ekho architecture) should generate higher-fidelity predictions of how individuals, teams, and leaders actually behave than existing approaches, whether purely mathematical models that lack an agent base, or agent-based simulations whose agents lack empirical psychometric grounding.

Proposition 1: Contextualized Psychometrics Predict Performance

The Predictive Validity of Psychometric Measurement

That personality and ability measures predict performance is among the most replicated findings in industrial-organizational psychology. Schmidt and Hunter (1998) summarized 85 years of selection research and established the operational validity of cognitive and structured methods for predicting job performance. Sackett, Zhang, Berry, and Lievens (2022) corrected systematic errors in earlier range-restriction adjustments and produced a revised hierarchy of operational validities, with structured interviews leading at $\rho = .42$, work samples and assessment centers near $\rho = .33$, cognitive ability at $\rho = .31$, and contextualized personality measures at $\rho = .12$ to $.25$.

Why Contextualization Matters

Tett and Guterman (2000) and Tett and Burnett (2003) proposed Trait Activation Theory, explaining that personality traits express in behavior only when the situation contains cues that activate them. An extraverted person in a role requiring solo analysis may never display the interpersonal energy that Extraversion predicts. A conscientious employee in an environment

without performance feedback may not sustain the goal-directed effort that Conscientiousness predicts. The trait is present, but the situation does not call on it.

Judge and Zapata (2015) tested this directly across many occupations and found that Big Five traits predicted performance more strongly in trait-relevant situations and in weak situations with high discretion. Similarly, Shaffer and Postlethwaite's (2012) meta-analysis of 90 studies found that the validity of personality assessments was approximately double when contextualized ($\rho = .24$). This effect held across all Big Five dimensions and across task, contextual, and overall performance criteria.

These findings have direct implications for Ekho: Conditioning each agent on a contextualized psychometric profile raises the utility and fidelity of the digital twin; and the more deeply Ekho ingests and models the business context, the more faithfully it can reproduce the person-environment interaction that determines behavior.

Beyond the Bright Side

The Big Five captures the bright side of personality however leadership effectiveness depends on more than that. The dark side of personality, the tendencies that surface under pressure, fatigue, or unguarded conditions, predicts derailment in ways the Big Five does not. Spain, Harms, and LeBreton (2014) reviewed the dark-side literature and reported that derailment traits predict counterproductive workplace behavior and managerial failure with effects that are largely independent of bright-side traits. Furnham, Richards, and Paulhus (2013) reviewed evidence that the Dark Triad (narcissism, Machiavellianism, psychopathy) shapes leadership emergence even when those same traits reduce leadership effectiveness.

For Ekho's leadership and succession use cases, we believe creating a digital twin of an individual's full psychological profile will materially improve the validity of the input.

The Stability of Psychometric Profiles

Ekho is intended to support forward-looking decisions: who to hire, how a team will behave six months from now, who should succeed whom. This presumes the psychological profile of an individual is stable enough to project forward. Roberts and DelVecchio (2000), in a meta-analysis of 152 longitudinal studies, found that rank-order consistency of personality traits is moderate in childhood ($r \approx .31$), rises through young adulthood, and reaches $r \approx .74$ by ages 50 to 70. Absolute trait levels change across the lifespan, but a person's standing relative to others is highly stable across the time horizons relevant to organizational decisions. This is the empirical basis for treating an assessed psychometric profile as a meaningful input to an agent that will be simulated months or years into the future.

Relevance of Psychometric Profiles for Digital Twins

The accuracy of any digital twin is bounded by the empirical quality of its inputs. The reviewed research highlights that psychometric data is a predictive input, and its validity is highest when anchored to relevant situations. Parameterizing each Ekho agent with a contextualized psychometric profile therefore equips the digital twin with the most valid representation of an individual that the assessment science currently supports. Without this grounding, an LLM agent describes a generic person in a generic firm, not the actual leader or team member the client is making decisions about.

The reviewed evidence focuses primarily on personality, but the underlying logic of input validity extends further. Following the multi-trait multi-method principle (Campbell & Fiske, 1959), an individual profile becomes more accurate as it draws on convergent evidence from multiple validated sources. We therefore hypothesize that profiling Ekho agents with a richer combination of inputs, including multi-rater feedback, social network analytics, biodata, estimations of expertise and skill proficiency, and performance metrics alongside personality data, will further improve the fidelity of the digital twin.

Proposition 2: LLM Agents Can Simulate the Organization

Six Decades of Trying to Build a Digital Twin of the Firm

Long before the phrase “digital twin” entered the popular vernacular, organizational scholars were trying to build one. Cyert and March (1963) created a running computational model of duopoly competition, treating firms as coalitions of boundedly rational actors. Forrester (1961) developed system dynamics at MIT, modeling organizations as systems of stocks, flows, feedback loops, and delays. Cohen, March, and Olsen (1972) developed the Garbage Can Model, one of the earliest reproducible computational simulations in the social sciences, demonstrating that decision outcomes in organized anarchies can emerge from temporal coincidence rather than rational design.

A more widely used and extensively validated organizational simulation is the Virtual Design Team (VDT), built by Levitt and colleagues at Stanford from 1988. VDT modeled project teams as agent-based discrete-event simulations of boundedly rational actors executing interdependent tasks, predicting project duration, coordination loads, and quality risks. Levitt (2004) documented multi-year validation projects, from intellectual experiments through calibration against real projects at firms such as Lockheed Martin, to prospective intervention.

The VDT program established that, with sustained effort, a computational organizational twin, grounded in rule-based agents, can be brought to a level of accuracy that managers will benefit from and ultimately trust. Other work has extended the same idea into human capital management. Somarathna (2020) modeled employees as autonomous agents in an agent-based simulation and showed that micro-level interactions produced emergent workforce dynamics invisible to other models. More recently, Nozari and Yordanova (2026) made the case in mathematical and optimization terms for a digital twin of human capital under uncertainty.

Why LLMs Improve the Organizational Digital Twin Simulation

LLM agents change the substrate on which the digital twin is built, and address four limitations of traditional agent-based modeling: rigid decision rules, parameter complexity, context updates, and opaque reasoning. LLMs absorb behavioral heterogeneity through prompt-based persona specification and fine-tuning, learn from experience through memory tools and reflection, and produce behaviors whose rationale can be self-inspected in natural language queries to the model.

There are an increasing number of experiments that demonstrate LLM-based multi-agent simulations can model realistic social/organizational phenomena. Park et al. (2023) designed a system that featured 25 LLM-driven characters with persistent memory and distinct personas who planned daily activities, formed relationships, and coordinated group behavior in a simulated town. Human raters judged their behavior more believable than that of agents driven by simpler architectures. Zhu et al. (2024) used LLM agents to simulate holacratic organizational structures, finding that the distribution of agent competence shaped stress and task completion. Chen et al. (2024) simulated hierarchical investment firms and found that authority structures altered agent reasoning and decision quality.

Together this work supports a clear hypothesis: LLM agents can simulate organizations by simulating the people and dynamic contexts within them, and the resulting model can serve as a behavioral digital twin of the firm at usable fidelity. What remains is the question of which are the best parameters to include in the twin.

Proposition 3: LLMs Can Be Infused With Psychometric Properties

Proposition 2 establishes that LLMs interacting with each other and properly prompted can simulate the organization. Proposition 3 asks if LLM powered agents can act as a human would act, consistently over time, and if yes, what would be the most effective data to provide in an effort to effectuate this behavior. The recent evidence on this question is compelling.

LLMs Can Simulate and Predict Human Behavior

Binz et al. (2025) introduced Centaur, a foundation model fine-tuned on more than 10 million human decisions across 160 cognitive experiments. The model produced individual-level predictions that matched held-out human behavior across a wide range of tasks, including tasks the model had never seen. This study demonstrates that a single LLM-based model can capture the regularities of human cognition at scale, and it is evidence that LLM-based systems can act as behavioral predictors of people.

Park et al. (2026) extend this from cognition to whole persons. They recruited 1,052 study participants representative of the U.S. population, conducted a two-hour life-history interview with each, and built a generative agent for every participant by giving an LLM the interview transcript and a structured expert synthesis as memory. The interview-based agents replicated participants' answers on the General Social Survey at 83% of participants' own two-week test-retest reliability, on the Big Five Inventory at 80%, and on behavioral economic games at 66%. This is a direct demonstration that an LLM agent can serve as a behavioral digital

twin of a specific person, and that the resulting twin behaves like its human counterpart on validated psychometric and behavioral measures.

It follows that if a team of LLM agents were prompted to interact with one another under a specified context, the dynamics that emerge from those interactions may represent a probable trajectory of what would occur if their human counterparts interacted under the same conditions. Of course a single run would not on its own provide compelling evidence of an outcome because human social interaction is stochastic, and so is the underlying language model. But if the simulation was run a thousand times under each condition of interest and analyzed the resulting distribution of outcomes via Monte Carlo replication, comparing across conditions, it might be possible to approximate the distribution of likely futures. This is especially the case under conditions of change such as leadership transitions, re-org initiatives, or team conflict, where emergent interactions have an outsized impact on the outcome. It can also be argued that this would do so more accurately than purely computational models or non-LLM agent-based models, because LLM agents can be calibrated to individuals far more richly, mirroring values, predictive personality traits, and past behavioral patterns, and can be continuously updated with new context.

Two further lines of work show that if the input to the LLM-based agent is a psychometric profile, the agent will likely behave consistently in a predictable manner in accordance with that profile. Serapio-García et al. (2023) prompted LLMs with Five-Factor Model profiles and found that the resulting responses showed internal consistency, convergent validity with established personality scales, and the expected pattern of correlations with downstream variables. Jiang et al. (2023) replicated and extended this finding for inducing personality in pretrained models, demonstrating that fine-tuning produces more stable trait

expression than prompting alone. Sorokovikova, Fedorova, Rezagholi, and Yamshchikov (2024) assigned LLM agents distinct Big Five profiles and observed measurable behavioral differences in social dilemma games, with agents high on Agreeableness and Conscientiousness cooperating more reliably, mirroring human findings.

These five lines of work (Centaur on cognition, Park et al. on whole-person twins built from interviews, Serapio-García et al., and Jiang et al. on trait induction from psychometric profiles, and Sorokovikova et al. on behavioral differentiation) establish that an LLM agent can be prompted with the psychometric properties of a specific person and that the resulting agent behaves in ways that systematically reflect those properties. This is the digital twin at the individual level.

Synthesis: A Behavioral Digital Twin of the Organization

Contextualized psychometric assessment is among the strongest available predictors of how individuals behave in defined organizational situations (Proposition 1). LLM agents can simulate organizations and the relationships within them (Proposition 2). LLMs can be infused with psychometric properties so that an individual agent functions as a digital twin of a real person (Proposition 3). Combining these three findings yields the contours of a single integrated simulation architecture: a behavioral digital twin of the organization which itself is populated by digital twins of its actual leaders, teams, and members.

There are three anticipated layers of information in the proposed simulation. The first is the organizational context, the external environment, the internal environment, and business initiatives, structured along the Burke and Litwin (1992) model of organizations, in which the external environment cascades through internal organizational factors and strategy to shape individual behavior. This layer can be held constant or varied across comparative runs, and

supplies the situational conditions that activate individual traits, as established earlier through trait activation theory. The second is each agent's psychometric profile, the validated dispositions that influence behavior within that context. The interplay between individual simulated agents and the external environment determines the emergent behavior (Lewin, 1951). The third is each agent's memory stream: the emergent outcomes of prior interactions, retrieved and folded into the prompt for the next so that prior exchanges shape what comes after.

Three likely constraints to this concept should be highlighted. First, a behavioral digital twin will be most effective for risk identification, not strategic prescription. For example, it is intended to surface where a senior hire may collide with the existing leadership team, where a restructuring may expose fault lines, or where a planned M&A may fracture a team, and afford leadership the time to manage the risk. Second, even validated assessment data carry subgroup differences, and an LLM trained on internet-scale data may add biases of its own. Auditing of simulation outputs is therefore a non-negotiable requirement of any deployment, and is part of our research agenda alongside behavioral validation. Finally, token costs for large-scale LLM-powered agent-based simulation can be prohibitive. This is an active research frontier; recent work on multi-agent KV cache sharing (Bian et al., 2026) and memory-efficient serving for large-scale simulation (Pan et al., 2026) has begun to make large agent populations tractable. Our strategy is to track and contribute to that research while applying the simulation prototypes we have already built to smaller groups — executive teams, boards of directors, and other settings where the agent population is bounded and the value-per-decision is high.

Conclusion

Three areas of research each independently supported by peer-reviewed evidence lead us to a simulated digital organization that can be significantly more powerful than previous

organizational simulation theory: contextualized psychometrics predict performance; LLM agents can simulate organizations; and LLMs can be infused with psychometric properties so that they function as digital twins of real people. Combining the three yields the framework for building a behavioral digital twin of the organization that should outperform an LLM simulation without empirical psychometric grounding. Deeper Signals is actively researching and developing this as a potential tool to help organizational leaders improve their human capital decisions.

References

- Bian, Z., Wu, F., Zhang, C., Dong, H., Liang, Y., & Zhuo, Y. (2026). TokenDance: Scaling Multi-Agent LLM Serving via Collective KV Cache Sharing. *arXiv preprint arXiv:2604.03143*.
- Binz, M., Akata, E., Bethge, M., Brändle, F., Callaway, F., Coda-Forno, J., . . . Schulz, E. (2025). A foundation model to predict and capture human cognition. *Nature*, *644*(8078), 1002-1009.
- Burke, W. W., & Litwin, G. H. (1992). A causal model of organizational performance and change. *Journal of Management*, *18*(3), 523-545.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81-105.
- Chen, C. C., Takamura, H., Kobayashi, I., & Miyao, Y. (2024). Hierarchical organization simulacra in the investment sector. *arXiv preprint arXiv:2410.00354*.
- Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly*, *17*(1), 1-25.
- Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm*. Prentice-Hall.

- Forrester, J. W. (1961). *Industrial dynamics*. MIT Press.
- Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The Dark Triad of personality: A 10 year review. *Social and Personality Psychology Compass*, 7(3), 199-216.
- Jiang, G., Xu, M., Zhu, S. C., Han, W., Zhang, C., & Zhu, Y. (2023). Evaluating and inducing personality in pre-trained language models. *Advances in Neural Information Processing Systems*, 36, 10622-10643.
- Judge, T. A., & Zapata, C. P. (2015). The person-situation debate revisited: Effect of situation strength and trait activation on the validity of the Big Five personality traits in predicting job performance. *Academy of Management Journal*, 58(4), 1149-1179.
- Levitt, R. E. (2004). Computational modeling of organizations comes of age. *Computational & Mathematical Organization Theory*, 10, 127-145.
- Lewin, K. (1951). *Field theory in social science: Selected theoretical papers* (D. Cartwright, Ed.). Harper & Row.
- Nozari, H., & Yordanova, Z. (2026). Digital twin-driven decision support for human capital management under uncertainty. *Green Technologies and Sustainability*, 4, 100353.
- Pan, Z., Shen, Y., Hu, Z., Wang, Z., Manocha, A., Wang, Z., ... & Ding, Y. (2026). ScaleSim: Serving Large-Scale Multi-Agent Simulation with Invocation Distance-Based Memory Management. *arXiv preprint arXiv:2601.21473*.
- Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 1-22.

- Park, J. S., Zou, C. Q., Kamphorst, J., Egan, N., Shaw, A., Hill, B. M., Cai, C., Morris, M. R., Liang, P., Willer, R., & Bernstein, M. S. (2026). LLM agents grounded in self-reports enable general-purpose simulation of individuals. *arXiv preprint arXiv:2411.10109*
- Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126*(1), 3-25.
- Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology, 107*(11), 2040-2068.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262-274.
- Serapio-García, G., Safdari, M., Crepy, C., Sun, L., Fitz, S., Romero, P., ... & Matarić, M. (2023). Personality traits in large language models. *arXiv preprint arXiv:2307.00184*.
- Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology, 65*(3), 445-494.
- Sorokovikova, A., Fedorova, N., Rezagholi, S., & Yamshchikov, I. P. (2024). LLMs simulate big five personality traits: Further evidence. *arXiv preprint arXiv:2402.01765*.
- Somarathna, K. U. S. (2020). An agent-based approach for modeling and simulation of human resource management as a complex system: Management strategy evaluation. *Simulation Modelling Practice and Theory, 104*, 102118.

- Spain, S. M., Harms, P., & LeBreton, J. M. (2014). The dark side of personality at work. *Journal of Organizational Behavior*, 35(S1), S41-S60.
- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, 88(3), 500-517.
- Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality*, 34(4), 397-423.
- Zhu, C., Cheng, Y., Zhang, J., Qiu, Y., Xia, S., & Zhu, H. (2024). Generative organizational behavior simulation using large language model based autonomous agents: A holacracy perspective. *arXiv preprint arXiv:2408.11826*.