

THE COGNITIVE DATA SUPERCYCLE

Key Takeaways - The Cognitive Data Stack (CDS)

Behind the Next Trillion Dollars of Enterprise Value

by Hernan Asorey, Co-Founder, AVC Turing

Companion to the full research paper. Same data. Fewer acronyms.

AVC TURING

March 2026

Investing in the Future of Intelligence

New York | Global

This is a research framework, not investment advice. Company references are illustrative case studies. See Appendix for full disclaimers.

© 2026 Hernan Asorey. All rights reserved.

Published by AVC Turing Ltd.

The Cognitive Data Stack (CDS) framework was introduced by Hernan Asorey, Co-Founder of AVC Turing. Citation and reference to the framework is encouraged with attribution to the author.

No part of this publication may be reproduced or distributed in whole without the prior written permission of the author.

The Big Picture: Why Data Is Eating AI

Everyone's talking about AI models. ChatGPT, Claude, Gemini; the names change every quarter. But here's the thing almost nobody outside the industry understands: the models are becoming interchangeable. They're getting cheaper, faster, and more alike with every release cycle. The marginal difference between the best model and the fifth-best model is shrinking toward zero.

So if the magic isn't in the model, where is it?

It's in the plumbing. The storage systems that feed data to the models. The human experts who teach models right from wrong. The software that turns an AI's prediction into an actual business action. The governance that keeps the whole thing from going sideways. That infrastructure is where the real value is concentrating, and almost nobody is talking about it.

Think of it this way. In the California Gold Rush, the miners who struck it rich were the exception. The people who reliably got rich were the ones selling picks, shovels, and Levi's jeans. We're seeing the same pattern in AI. The "model miners" are in a race to the bottom on price. The infrastructure builders are accumulating durable, compounding value.

We call this infrastructure the Cognitive Data Stack (CDS). This paper maps who's building it, why it matters, and where approximately \$250–\$300 billion in enterprise value is concentrating.

Three Eras of Enterprise Data

Era 1: Big Data (~2010–2020). The philosophy was simple: store everything. Companies hoarded data like squirrels before winter, hoping it would be useful someday. Most of it wasn't.

Era 2: Modern Data Stack (~2020–2024). The cleanup phase. Companies built pipelines to organize all that hoarded data into tidy warehouses and dashboards. The goal: humans reading charts to make decisions.

Era 3: Cognitive Data (2025–present). The paradigm flips. The data isn't for humans reading dashboards anymore; it's for machines making decisions. This changes everything about how data needs to be stored, refined, and delivered. You don't need a pretty chart. You need ground truth delivered at GPU speed.

The Modern Data Stack was built for humans to read dashboards. The Cognitive Data Stack is built for machines to execute decisions.

The Three Layers (Plus a Control Plane)

Imagine a factory. Raw materials come in one end. Finished products come out the other. The CDS works the same way, with three layers:

Layer 1: Physics. This is the loading dock. How do you move mountains of raw data fast enough to feed thousands of hungry GPUs? Think of it as the plumbing and electrical wiring of the factory.

Layer 2: Genesis. This is the factory floor. Raw data gets refined into something useful: human experts teach models right from wrong, doctors' conversations get turned into structured medical records, legal documents get digitized into searchable knowledge bases. The output is "reasoning data"; data explicitly shaped to teach machines how to think.

Layer 3: Activation. This is the shipping department. The AI has made a decision; now what? Activation is the software that takes a prediction and actually does something with it: refunds a customer, routes a drone, triggers a supply chain fix.

And running across all three layers is the Control Plane: the safety systems, governance rules, and audit trails that keep the whole factory from going haywire. Think of it as quality control, security cameras, and fire suppression rolled into one.

This framework isn't a clever marketing construct. It describes four engineering constraints that exist regardless of which model wins: data has to move fast (Physics), models need ground truth (Genesis), predictions have to become actions (Activation), and all of it needs governance (Control Plane). As long as those constraints exist, and they will, these layers are the permanent architecture of AI infrastructure.

Why This Framework Doesn't Break

When we introduced the CDS, the first question from serious investors was: “What kills it?”

The answer is: nothing kills the framework. Individual companies within it will rise, consolidate, and in some cases be displaced. That's normal. But the layers themselves persist because they map to physics and economics, not to any specific vendor's product.

The Modern Data Stack lasted a decade as the industry's organizing taxonomy. When Stitch replaced Fivetran at a specific enterprise, the MDS didn't “fail.” The ELT layer persisted; a company within it shifted. The same logic applies here.

Three structural dynamics protect the architecture from commoditization:

- 1. Data Gravity** (Physics Layer). Once an enterprise moves 50 petabytes of training data into VAST or Databricks, the cost of moving it out is so high that switching becomes economically irrational. You're not locked in by a contract. You're locked in by physics.
- 2. The Reasoning Feedback Loop** (Genesis Layer). Every line of code written in Cursor, every medical note generated by Abridge, every legal query answered by Harvey creates proprietary training data that makes their model smarter. A generic competitor can't bridge this gap by simply deploying a better base model. The data asset compounds.
- 3. The Integration Barrier** (Activation Layer). Peregrine's 5,000 connected law enforcement systems, Anduril's classified defense integrations, Celonis's enterprise process maps; each connection took months or years to build. Replacing them means re-doing all that work. Nobody does that voluntarily.

— — —

The Evidence: Follow the Money

A framework is only as good as the evidence that supports it. Here's what we found when we mapped where private-market capital, engineering talent, and intellectual property are actually flowing.

Across 19 companies and three layers, approximately \$250–\$300 billion in aggregate enterprise value has concentrated in the picks and shovels of the cognitive age. This is not a prediction. It is an observation. The money is already in the ground.

Physics Layer: ~\$183B

Five companies. Dominated by Databricks (\$134B), the de facto operating system for enterprise AI. VAST Data (~\$30B) solved the storage bottleneck. Lambda, Fireworks AI, and WEKA fill out compute cloud, inference serving, and performance acceleration.

Publicly: Snowflake and CoreWeave compete here. So do the hyperscaler infrastructure divisions of AWS, Azure, and GCP.

Genesis Layer: ~\$108B

Eight companies. Split between horizontal foundries (Scale AI at ~\$29B, Surge AI at \$15–\$25B, Snorkel AI at \$1.3B) and vertical digitizers (Cursor at \$29.3B, Harvey at \$8–11B, Glean at \$7.2B, Abridge at \$5.3B, Hippocratic AI at \$3.5B).

The split is itself informative. Roughly equal value in both sub-markets means the industry needs both: universal data manufacturing and industry-specific proprietary datasets.

Activation Layer: ~\$63B

Six companies. Anduril (\$30.5B) anchors kinetic activation (defense and autonomous systems). Applied Intuition (\$15B) validates AI in simulation before deployment. Celonis (\$13B+) translates data into enterprise efficiency. Peregrine (\$2.5B), Hightouch (\$1.2B), and LangChain (\$1.25B) provide integration, write-back rails, and agent orchestration.

The wiring is standardizing fast. MCP (Anthropic) and A2A (Google), now both under the Linux Foundation, are becoming the USB-C and HTTP of agentic AI. This validates the layer's core premise: the scarce resource is not intelligence but the governed connective tissue that lets intelligence act.

Control Plane: The White Space

This is the fascinating part. Unlike every other layer, the Control Plane has no \$1B+ private pure-play company. It's the single biggest white space in the entire AI infrastructure landscape.

Three groups are racing to fill it: cloud giants bundling governance for free, established data governance companies (Collibra, Alation) pivoting toward AI, and AI-native startups (Arize AI) that are purpose-built but pre-scale.

The wild card: regulation. If the EU AI Act and similar frameworks demand independent evaluation, there's a massive market for stand-alone governance platforms. If regulators accept self-assessment, the cloud giants bundle it and the window closes.

— — —

Who Builds All This?

Infrastructure doesn't build itself. The CDS requires specialized human capital across every layer: systems engineers for Physics, domain experts (clinicians, lawyers, linguists) for Genesis, integration specialists for Activation, and governance professionals for the Control Plane.

But the talent layer itself is changing. AI-native building tools (Claude Code, Cursor, Gemini Code Assist, LangChain, CrewAI) are increasingly capable of performing implementation tasks that previously required specialized engineers. The practical implication: enterprises

building the CDS will staff with a combination of human domain experts and AI development tools. The talent platforms and AI builders that can orchestrate both will capture disproportionate value.

Companies like Turing (the talent platform, no affiliation with AVC Turing) represent one approach: matching enterprises with specialized AI talent globally. But they compete alongside enterprise consultancies (Accenture, Deloitte, McKinsey’s QuantumBlack) and the growing ecosystem of agentic coding tools. The Service Layer is adjacent to the CDS rather than embedded within it: it enables all layers without being a layer itself.

— — —

What Could Go Wrong

No thesis is bulletproof. Here’s what would shift value within the framework:

Platform bundling. If AWS, Azure, or Google deliver end-to-end AI infrastructure that’s “good enough” across all layers, standalone companies get squeezed. This risk is most acute in the Control Plane, where hyperscalers are already giving away governance for free.

Synthetic data breakthrough. If AI can generate its own training data that’s as good as human-created data, the Genesis Layer’s human labeling moat erodes. Today, frontier models still need human ground truth. But this could change.

Model providers go vertical. If OpenAI, Anthropic, or Google decide to build their own legal AI, clinical documentation, or defense systems, they could bypass the CDS intermediaries. So far, they’ve focused on general-purpose models. But “so far” is not “forever.”

Open source accelerates. Durable moats must be rooted in proprietary data, not just clever engineering. If open-source tools replicate the software (and they will), the only defense is owning something that can’t be open-sourced: the data itself, the integration depth, the regulatory compliance.

None of these break the framework. They describe value migration within it. The four constraints persist. The companies solving them may change.

— — —

The Bottom Line

The AI market is not one market. It’s at least four: Physics, Genesis, Activation, and the Control Plane. The models are commoditizing. The infrastructure is not.

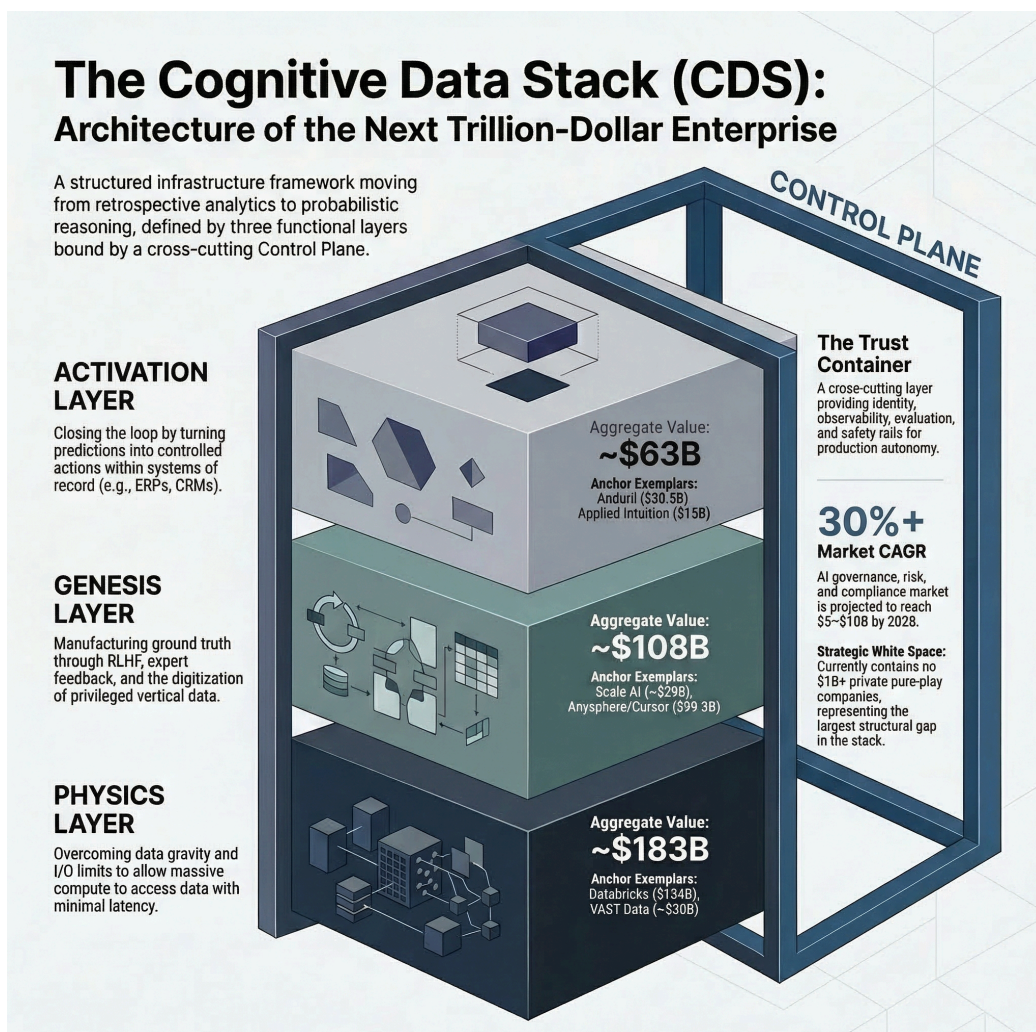
Across 19 companies and three layers, approximately \$250–\$300 billion in aggregate enterprise value has concentrated in the picks and shovels of the cognitive age. The next

generation of \$10B+ companies will be defined by their ability to digitize a specific slice of the physical economy and own the proprietary data asset that results.

The market often mistakes infrastructure for commodity. But the history of every technology cycle tells us the same story: the most durable value accrues to the layers that control the flow of the critical resource. In the AI era, that resource is Reasoning Data.

The Cognitive Data Stack is not a thesis about specific companies. It is a structural description of how data infrastructure must be organized to serve intelligence at scale. The companies profiled here are the evidence that the framework is already operative. As in every technology cycle, individual players will rise, consolidate, and in some cases be displaced. The constraints they solve will persist.

The question is not whether the CDS exists. It is who will build it.



About the Author

Hernán Asorey is Co-Founder and Managing Partner of AVC Turing, a New York-based venture and research firm investing at the intersection of applied AI, critical infrastructure, and global markets. His practitioner credentials span three of the most consequential technology organizations of the past two decades. At Salesforce, he served as the company's first Chief Data Officer, architecting a modern data platform, launching the Einstein AI platform, and embedding real-time intelligence into the enterprise product loop; that work earned recognition from the World Economic Forum at Davos and a dedicated feature in Marc Benioff's book *Trailblazer*. At Microsoft, he partnered with OpenAI to design and deliver the first business-ready generative AI experiences, including enterprise AI copilots and LLM-powered applications across industries, well before the category had a name. At Google, he led data science and engineering globally across the company's most iconic products, shaping growth strategy and overseeing the experimentation and measurement infrastructure that governs decisions at scale.

Asorey serves as Independent Board Member of Contentsquare, Head of the Board of Advisors at Klaviyo (NYSE: KVYO), and holds board and advisory roles at Alation, Syncro, Faros AI, and Trebellar. In 2021, he was named to the HITEC 100 as one of the most influential Hispanic professionals in technology. A Stanford GSB alumnus, he began his career as a university lecturer in Buenos Aires, a teaching instinct that continues to shape how he approaches research: every framework is, at its core, a tool for helping practitioners think more clearly about systems that are changing faster than conventional wisdom can track. The Cognitive Data Stack is that kind of tool.

Appendix: Disclaimers

The materials, content and opportunity described herein are for informational purposes only and for the exclusive use of investors outside the United States of America that are considered sophisticated and/or professional investors in the jurisdictions of their residence and/or organization; this opportunity is not suitable for retail investors.

No U.S. Distribution: *This document is not directed at, nor intended for distribution to, any person or entity in the United States, nor to U.S. persons (as defined under Regulation S of the U.S. Securities Act of 1933), and may not be distributed in any jurisdiction where such distribution would be unlawful.*

Confidentiality: *This document and its contents are confidential and for the exclusive use of the person(s) to whom they are delivered and should not be copied or distributed, in whole or in part, or disclosed by such person(s) to any other person (and by receiving this document the recipient agrees so). This document is an outline of matters for discussion only.*

No Advice: *This document does not constitute an offer to sell or the solicitation of an offer to buy any securities and should not be interpreted as advice, including legal, tax or accounting advice. Potential investors are encouraged to conduct their own evaluation and seek independent advice before making an investment decision. References to specific companies are for illustrative and educational purposes only and do not constitute any recommendation, endorsement, or view on the suitability of any investment in such companies.*

This document has not been registered, reviewed or approved by any regulatory authority and does not constitute an offer of securities. Such an offer may only be made by means of confidential offering materials.

Risk Warning: *Investment in private equity involves a high degree of risk, it is an illiquid investment, speculative in nature and is subject to a risk of loss, including a risk of loss of principal. Past performance is not indicative of future results, and no express or implied guarantees are or can be provided regarding returns or capital recovery.*

AVC Turing Ltd. does not provide investment advice, portfolio management services, or personalized recommendations; all information herein is of a general and educational nature. AVC Turing Ltd. has no affiliation with any of the companies mentioned in this report, and does not have access to non-public, confidential, or insider information relating to any of these companies.

Data Sources & Accuracy: *All financial data, valuations, ratios, and projections presented in this analysis are derived from publicly available sources including research reports, news articles, and market data providers as of Q1 2026. These sources may contain errors, inconsistencies, or incomplete information. Private companies' information and official financial disclosures are limited, and much of the data represents estimates or third-party analyses that cannot be independently verified. No information in this document has been independently verified by AVC Turing Ltd. or its affiliates.*

Private Market Limitations: *Private equity shares are not publicly traded and are subject to transfer restrictions. Valuations are based on limited private market transactions and may not reflect actual realizable value. Liquidity is extremely limited and investment recovery may be difficult or impossible.*

No Reliance: *No person should rely on the information contained in this document for the purposes of making any investment decision. Any investment decision should be based solely on independent due diligence and, where applicable, formal offering documentation.*

© 2026 Hernan Asorey. All rights reserved. Published by AVC Turing Ltd. The Cognitive Data Stack (CDS) framework was introduced by Hernan Asorey, Co-Founder of AVC Turing. Citation and reference is encouraged with attribution. Reproduction or distribution in whole requires prior written permission of the author.

*Investing in the Future of Intelligence
New York | Global*

AVC TURING

*Investing in the Future of Intelligence
New York | Global*