Non-Obvious AI Software & Data Layer Investment Plays

Research Date: February 13, 2026

EXECUTIVE SUMMARY

The AI investment thesis has bifurcated: hardware/infra stocks (SMH index +71% over 24mo) are crushing software (IGV +35%). This creates **opportunity in the software/data layer** where the market is overlooking companies that are essential plumbing for AI adoption. The recent Anthropic Claude legal plugin selloff (Feb 3-4, wiping ~$1T from software stocks) has created especially compelling entry points in data-moat companies.

Key themes: - The "data moat" thesis is being stress-tested in real-time (Anthropic legal plugin panic) - AI training data is a $3B+ market growing 26%+ CAGR - Vector databases are commoditizing (Pinecone looking for a buyer), but the data pipes ARE NOT - Vertical AI is where defensibility lives — workflow + data moats > model moats - AI agents infrastructure is the next wave (post-copilot → agentic)

🏆 TOP PICKS BY CATEGORY

---

1. DATA INFRASTRUCTURE — The Plumbing Nobody Thinks About

Innodata (INOD) — Hidden Gem Score: 9/10

- **Ticker:** INOD (NASDAQ)
- **Market Cap:** ~$1.7B
- **Why Non-Obvious:** This was a sleepy $50M-revenue content services company that has completely reinvented itself as an AI training data powerhouse. It provides data labeling, annotation, and pre-training data to Big Tech AI labs. Most investors still think of it as an old-school BPO.
- **Revenue Growth:** 45%+ organic growth in 2025, AI Services segment growing 60%+. Q3 2025 revenue up 20% YoY with guidance reaffirmed.
- **The Edge:** New Federal Business Unit secured $25M government AI project for 2026. Landing new Big Tech customers regularly. Pre-

training data becoming a major revenue driver.

- **Risk:** Customer concentration with Big Tech labs. If frontier model training slows, so does revenue.
- **Contrarian Take:** This is the closest public-market pure-play on "AI needs data." At $1.7B market cap, it's still small enough to multi-bag.

TaskUs (TASK) — Hidden Gem Score: 8/10

- **Ticker:** TASK (NASDAQ)
- **Market Cap:** ~$1.2-1.5B
- **Why Non-Obvious:** Most people see "outsourced customer service." The reality: TaskUs is quietly becoming a critical AI services provider. Its AI segment ($58.7M in Q3 2025, growing 60.8%) does data labeling, annotation, and red-teaming for foundation model developers.
- **Revenue Growth:** 17-22% YoY overall, but AI Services at 60%+ growth
- **The Edge:** Trust & safety + AI services creates a two-sided moat. As AI generates more content, you need more human review. As AI models improve, you need more RLHF data. TaskUs wins either way.
- **Risk:** Traded down significantly from 2021 highs. Perception as a BPO commodity.
- **Contrarian Take:** At ~1.2x revenue, this is absurdly cheap for a company with a 60%+ growing AI segment.

Confluent (CFLT) — BEING ACQUIRED BY IBM at $31/share

- **Status:** IBM acquiring for ~$31/share as of Dec 2025. Deal pending regulatory approval.
- **Note:** Validates the thesis that data streaming is critical AI infrastructure. IBM paying up because real-time data pipelines are essential for AI workloads.

Informatica (INFA) — ACQUIRED BY SALESFORCE for ~$8B

- **Status:** Completed Nov 2025. Now wholly-owned Salesforce subsidiary.
- **Note:** Another validation — data management/quality is so critical that Salesforce paid $8B for it.

---

2. AI MIDDLEWARE / ORCHESTRATION

Elastic (ESTC) — Hidden Gem Score: 7/10

- **Ticker:** ESTC (NYSE)
- **Market Cap:** ~$6.2B
- **Why Non-Obvious:** Everyone thinks "search engine." But Elastic has quietly become a core AI infrastructure company — its vector search capabilities make it a production-grade alternative to standalone vector databases (which are struggling — Pinecone is looking for a buyer). Elastic combines search, observability, and security in one platform.
- **Revenue Growth:** Mid-teens (~15%) with improving trajectory
- **The Edge:** Unlike pure vector DB startups, Elastic has an existing enterprise customer base and revenue. Vector search is an ADD-ON to existing deployments, not a rip-and-replace.
- **Risk:** Stock beaten down to ~$60 range. Competition from MongoDB's vector capabilities and native cloud offerings.

- **Contrarian Take:** The vector database pure-plays are dying (Pinecone struggling at sub-$1B valuation). The winners will be established data platforms that add vector capabilities. Elastic is winning this.

Dynatrace (DT) — Hidden Gem Score: 7/10

- **Ticker:** DT (NYSE)
- **Market Cap:** ~$15B (estimated)
- **Why Non-Obvious:** Observability sounds boring. But as enterprises deploy AI agents, they need to OBSERVE them. Dynatrace's AI-powered platform (Grail + Davis AI) is uniquely positioned for "AI observing AI." Just announced agentic operations system at Perform 2026 conference.
- **Revenue Growth:** 20% ARR growth, beat-and-raise quarter in Feb 2026. Authorized $1B buyback.
- **The Edge:** Every AI agent deployed needs monitoring. Every LLM call needs observability. Dynatrace's causal AI engine can trace issues across AI-driven systems in ways traditional APM cannot.
- **Risk:** Valuation not cheap. Competes with Datadog, New Relic.
- **Contrarian Take:** The Motley Fool just called it "essential software." As AI agents proliferate, the observability tax on every AI workload accrues to companies like DT.

Vector Database Pure-Plays (Caution)

- **Pinecone** — Private, reportedly looking for buyer as of Sep 2025. Valued near $1B but struggling with commoditization. PostgreSQL + pgvector eating their lunch.
- **Weaviate, Qdrant, Milvus** — All private, open-source-based. Category risk is real.
- **Verdict:** Avoid standalone vector DB companies. The capability is being absorbed by existing platforms (Elastic, MongoDB, PostgreSQL).

---

3. AI SECURITY & GOVERNANCE

PagerDuty (PD) — Hidden Gem Score: 6/10

- **Ticker:** PD (NYSE)
- **Market Cap:** ~$2B
- **Why Non-Obvious:** AI incident response. As companies deploy AI, things WILL break. PagerDuty's AI assistant and automation tools shorten incident response times. They sit at the intersection of DevOps + AI operations.
- **Revenue Growth:** Low-teens, but improving margins
- **The Edge:** Sticky enterprise platform. As AI systems become more complex, incident management becomes MORE critical, not less.
- **Risk:** Growth has been sluggish. Enterprise sales cycles are long.

Clearview AI — Private (Pre-IPO)

- **Status:** Private, exploring IPO pathway
- **Why Notable:** 20B+ image database for facial recognition. Controversial but the data moat is undeniable.
- **Risk:** Massive regulatory/legal headwinds. Fined in multiple jurisdictions.

Broader AI Security Space

- The public market lacks a pure-play AI security/governance company. This is a gap that will likely be filled by IPOs in 2026-2027.
- **Private companies to watch:** Robust Intelligence (AI model security, acquired by Cisco), Arthur AI (model monitoring), Protect AI, CalypsoAI, Lakera
- **Play it through:** CrowdStrike (CRWD) and Palo Alto Networks (PANW) are adding AI security features, but these are NOT hidden gems — they're consensus picks.

---

4. VERTICAL AI SaaS — Where the Real Moats Are

Veeva Systems (VEEV) — Hidden Gem Score: 6/10

- **Ticker:** VEEV (NYSE)
- **Market Cap:** ~$30.7B
- **Why Non-Obvious:** In the Feb 3-4 AI panic selloff, investors lumped Veeva with threatened SaaS companies. But Veeva's life sciences data (clinical trial data, regulatory submissions, drug safety data) is IRREPLACEABLE. AI models can't hallucinate their way through FDA submissions.
- **Revenue Growth:** TTM $3.08B. Expanding AI strategy with Vault CRM transition.
- **The Edge:** Regulatory moat + data moat + switching costs = triple lock. No AI chatbot replaces the system of record for pharmaceutical compliance.
- **Risk:** Already well-known. Valuation reflects quality.

Appian (APPN) — Hidden Gem Score: 7/10

- **Ticker:** APPN (NASDAQ)
- **Market Cap:** ~$2-3B
- **Why Non-Obvious:** Low-code/no-code + AI agents = enterprise workflow automation without the traditional RPA baggage. Appian's process mining + AI capabilities let enterprises build AI-powered workflows without coding.
- **Revenue Growth:** Accelerating from process automation demand
- **The Edge:** Government/regulated industry focus means sticky, high-compliance customers who can't use generic AI tools.

UiPath (PATH) — Hidden Gem Score: 8/10

- **Ticker:** PATH (NYSE)
- **Market Cap:** ~$6-7B
- **Stock Price:** ~$12-13
- **Why Non-Obvious:** UiPath is the poster child for "left for dead." Down massively from IPO highs of $80+. BUT: the company is pivoting from traditional RPA to AI-powered agentic automation. Trading at forward P/S < 4.5x and forward P/E of ~19x — CHEAP for a company that could be the operating system for AI agents in the enterprise.
- **Revenue Growth:** Returning to growth after a difficult 2024-2025 transition
- **The Edge:** 10,000+ enterprise customers, deep process understanding,

massive workflow data. AI agents need to DO things in enterprise systems — UiPath has the connectors, the process knowledge, and the automation infrastructure.

- **Contrarian Take:** The Motley Fool's Jan 2026 analysis says this "could define the next chapter of the tech market." At $12/share, you're paying for the legacy RPA business and getting the AI agent platform for free.

---

5. AI AGENTS INFRASTRUCTURE

UiPath (PATH) — See above (dual-listed in categories 4 & 5)

Best public pure-play on AI agents infrastructure.

Freshworks (FRSH) — Hidden Gem Score: 6/10

- **Ticker:** FRSH (NASDAQ)
- **Market Cap:** ~$4B
- **Why Non-Obvious:** Enterprise AI for the SMB market. Freddy AI assistant makes enterprise-grade automation accessible to small/mid businesses. Most AI agent companies target Fortune 500 — Freshworks goes after the 68,000+ mid-market customers nobody else serves.
- **Revenue Growth:** Mid-teens, improving margins
- **The Edge:** SMB land-and-expand with AI upsell. Once you're the CRM/ITSM platform, adding AI agents is natural.

Amplitude (AMPL) — Hidden Gem Score: 5/10

- **Ticker:** AMPL (NASDAQ)
- **Market Cap:** ~$1B
- **Why Non-Obvious:** Product analytics + AI agents that automatically generate insights and suggest experiments. If AI agents need to understand user behavior, Amplitude's data is the fuel.
- **Revenue Growth:** Low teens
- **Risk:** Heavy competition from Adobe, free alternatives.

---

6. COMPANIES WITH UNIQUE DATA MOATS — THE CONTRARIAN GOLDMINE

Thomson Reuters (TRI) — Hidden Gem Score: 8/10   (POST-SELLOFF OPPORTUNITY)

- **Ticker:** TRI (NYSE/TSX)
- **Market Cap:** ~$60-65B (DOWN from ~$80B pre-selloff)
- **Why Non-Obvious:** Thomson Reuters just got CRUSHED (down 16%+ on Feb 3-4) after Anthropic announced a legal plugin for Claude. The market is panicking that AI will commoditize legal research. **This is WRONG.**
- **The Bull Case:**
  - Westlaw's curated case law database took DECADES to build. Anthropic's plugin doesn't replace it — it needs it.
  - 45% of EBIT comes from legal, but TRI is also tax, compliance, and news.

- TRI already has CoCounsel (AI-powered legal assistant built ON their data).
    - The data providers who get disrupted are those whose data is PUBLICLY AVAILABLE. TRI's data is paywalled, curated, and legally sensitive.
- **Revenue Growth:** Steady high-single-digits with improving margins
- **Contrarian Take:** The ~16% selloff is an overreaction. TRI's data BECOMES the training data for legal AI — they get paid either way. Buy the fear.

RELX (RELX) — Hidden Gem Score: 8/10   (POST-SELLOFF OPPORTUNITY)

- **Ticker:** RELX (LSE/AMS), RELX (NYSE ADR)
- **Market Cap:** ~$70B (down from ~$85B+ pre-selloff)
- **Why Non-Obvious:** Same Anthropic panic selloff as TRI. RELX owns LexisNexis (legal), Elsevier (scientific publishing), and risk analytics. Only 10-13% of EBIT is from legal — the selloff is DISPROPORTIONATE to the actual threat.
- **The Bull Case:**
    - Elsevier's scientific publishing database is the foundation for medical/scientific AI training
    - Risk analytics (insurance underwriting data) has ZERO AI disruption risk — it's the INPUT to AI
    - LexisNexis data moat is as deep as Westlaw's
- **Revenue Growth:** Consistent mid-single-digits with high margins
- **Contrarian Take:** Getting -14% cheaper because of a legal AI tool when only 10% of your business is legal? That's a gift.

Wolters Kluwer (WKL) — Hidden Gem Score: 7/10

- **Ticker:** WKL (AMS: WLSNc), OTC: WTKWY
- **Market Cap:** ~$30B (down from ~$35B)
- **Why Non-Obvious:** Tax, compliance, health, legal workflow software. The data inside Wolters Kluwer's products is the regulatory and compliance knowledge graph that AI models need to access, not replace.
- **Contrarian Take:** Also hit by Anthropic panic. Same thesis as TRI/RELX — these data moats are inputs to AI, not victims of AI.

Verisk Analytics (VRSK) — Hidden Gem Score: 7/10

- **Ticker:** VRSK (NASDAQ)
- **Market Cap:** ~$40B
- **Why Non-Obvious:** Insurance analytics data. Every P&C insurer in the US runs through Verisk's actuarial data. This data becomes MORE valuable in an AI world — it's the training data for insurance AI models.
- **Revenue Growth:** Stable, EPS guidance $6.80-7.00 for 2026
- **The Edge:** Decades of proprietary actuarial data. No AI model can generate insurance loss history — you need the actual data.

Roper Technologies (ROP) — Hidden Gem Score: 6/10

- **Ticker:** ROP (NASDAQ)
- **Market Cap:** ~$37B
- **Why Non-Obvious:** Boring industrial conglomerate that actually owns

a portfolio of vertical software businesses with deep data moats (legal, healthcare, supply chain). AI creating "tens of millions" in new ARR from AI-native products, with acceleration expected in 2026.

- **Revenue Growth:** Organic growth ~4%, but high free cash flow generation
- **The Edge:** Collection of niche vertical software companies, each sitting on proprietary datasets in their verticals.

MSCI (MSCI) — Hidden Gem Score: 7/10

- **Ticker:** MSCI (NYSE)
- **Market Cap:** ~$40B
- **Why Non-Obvious:** Financial indices and ESG data. Also hit in the Anthropic selloff (-7%). But MSCI's index data is ESSENTIAL infrastructure for the financial industry — every ETF, every benchmark, every risk model needs MSCI data.
- **Contrarian Take:** AI doesn't replace financial benchmarks. If anything, quant AI funds need MORE index data.

📊 SUMMARY SCORECARD

Company	Ticker	Mkt Cap	Hidden Gem (1-10)	Category	Key Thesis
Innodata	INOD	$1.7B	9	Data Infra	Pure-play AI training data
TaskUs	TASK	$1.3B	8	Data Infra	AI services at 60% growth, BPO valuation
Thomson Reuters	TRI	$63B	8	Data Moat	Post-selloff, data moat misunderstood
RELX	RELX	$70B	8	Data Moat	Post-selloff, only 10% legal EBIT
UiPath	PATH	$6B	8	AI Agents	Left for dead, AI agent platform
Elastic	ESTC	$6.2B	7	Middleware	Vector search killing pure-play VDBs
Dynatrace	DT	$15B	7	Middleware	AI observability essential
Appian	APPN	$2.5B	7	Vertical AI	Low-code + AI agents for regulated
Wolters Kluwer	WKL	$30B	7	Data Moat	Compliance data AI-proof
					Insurance data

Verisk	VRSK	$40B	7		Data Moat	irreplaceable
MSCI	MSCI	$40B	7		Data Moat	Financial index data + AI
Veeva	VEEV	$31B	6		Vertical AI	Pharma data moat, regulatory lock
Roper Technologies	ROP	$37B	6		Data Moat	Boring conglomerate, vertical data
Freshworks	FRSH	$4B	6		AI Agents	SMB AI automation
PagerDuty	PD	$2B	6		AI Security	AI incident response

TOP 5 ACTIONABLE IDEAS (Highest Conviction)

1. **INOD (Innodata)** — Pure-play AI training data at $1.7B. 45%+ growth. This is THE picks-and-shovels play.
2. **TRI (Thomson Reuters)** — Buy the post-Anthropic-panic dip. Data moat is misunderstood. -16% overreaction.
3. **PATH (UiPath)** — AI agents need automation rails. At $12/share and P/E of 19x, the market is pricing in death, not transformation.
4. **TASK (TaskUs)** — 60% AI revenue growth hidden inside a "BPO" label. 1.2x revenue is value territory.
5. **RELX** — Post-selloff (-14%), only 10-13% legal exposure. Elsevier + risk analytics are AI-proof.

⚠ PRIVATE COMPANIES TO WATCH (Future IPOs)

Company	Category	Status	Valuation	Note
Scale AI	Data Labeling	Private	$14B+	Pivoting to Physical AI/robotics data
Labelbox	Data Labeling	Private	$1B+	Enterprise data labeling platform
Weights & Biases	MLOps	Private	$1.25B	Experiment tracking, model registry
Hugging Face	Model Hub	Private	$4.5B	GitHub of AI models
Anyscale (Ray)	Infra	Private	$1B+	Distributed computing for AI
Arthur AI	Model Monitoring	Private	~$100M	AI governance/monitoring

Protect AI	AI Security	Private	~$100M	ML supply chain security
Cohere	Enterprise AI	Private	$5.5B+	Enterprise-focused LLM provider

MACRO CONTEXT

- **The AI Software Valuation Gap:** SMH (semis) +71% vs IGV (software) +35% over 24 months. Software is the relative value play.
- **Anthropic Panic Selloff (Feb 3-4, 2026):** ~$1T wiped from software/data stocks. Creates buying opportunities in data moat companies.
- **Data Annotation Market:** $3.07B in 2026, growing 32% CAGR to $12.4B by 2031.
- **AI Agent Infrastructure:** The next wave after copilots. Enterprise automation platforms (UiPath, Appian) are the rails.
- **Vector Database Shakeout:** Pure-play vector DBs struggling. Winners are established platforms (Elastic, MongoDB, PostgreSQL) adding vector capabilities.

Disclaimer: This is research analysis, not investment advice. All market data approximate as of Feb 13, 2026.