theCUBE research
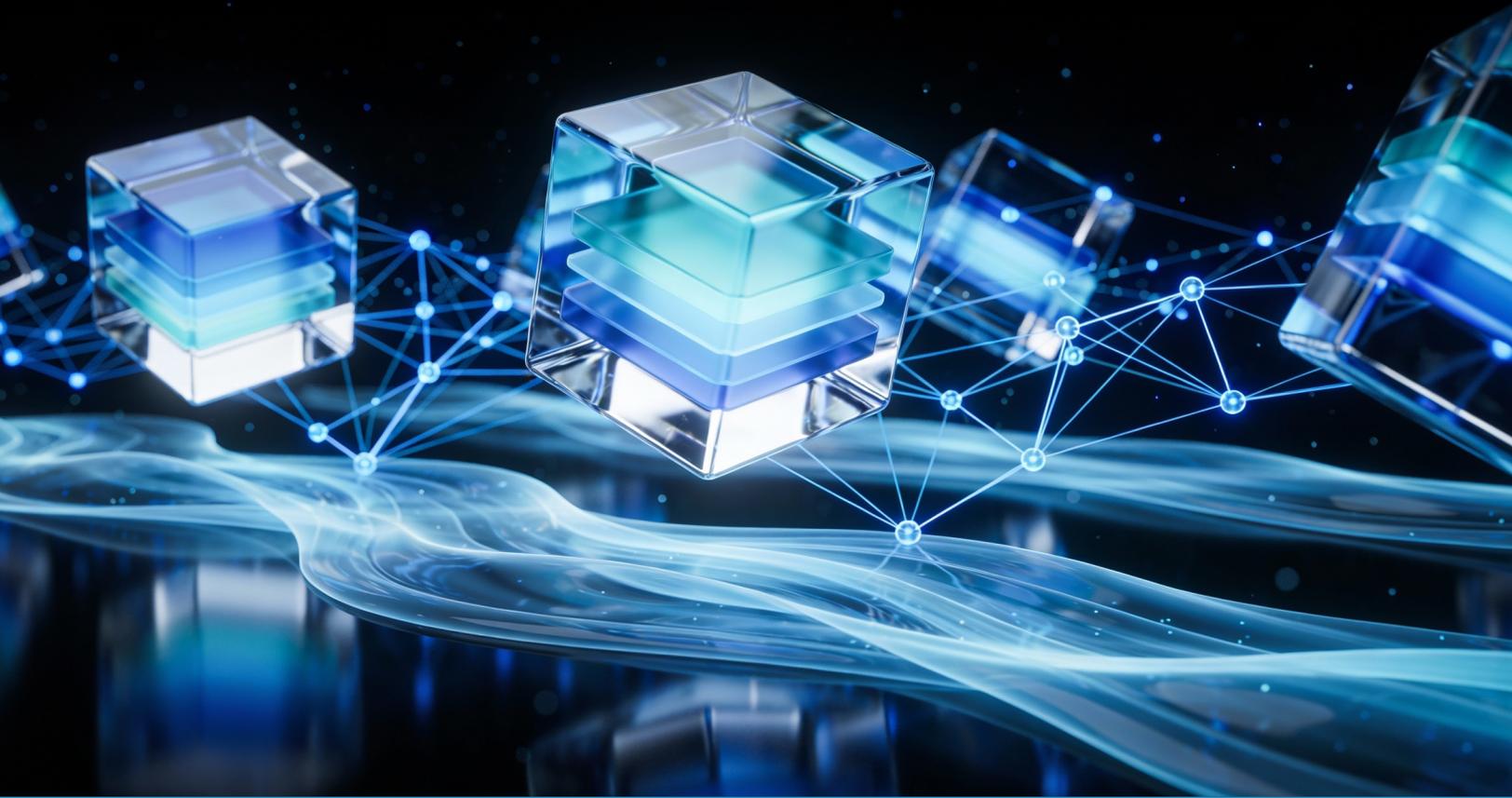
How AI Teams Reclaim Time, Velocity, and Budget with Union.ai

Paul Nashawaty
Principal Analyst | theCUBE Research

February 2026

Executive Summary

Union.ai is an AI development infrastructure platform that empowers organizations to accelerate AI/ML innovation by eliminating operational complexity and optimizing developer productivity. This economic validation, led by theCUBE Research's analyst team, quantifies the real-world financial and operational impact of Union.ai for mid-market and enterprise technology teams. The findings draw from a combination of survey data, customer interviews, and conservative financial modeling to ensure credibility and practical relevance.

The composite organization used in this study represents a 2,500-employee company with a 15-person AI/ML team, reflecting a mix of practitioners and managers across technology, healthcare, and climate sectors; while modeled at enterprise scale, the findings are broadly applicable to start-ups and smaller organizations facing similar AI adoption and operational challenges. Prior to adopting Union.ai, these teams faced challenges such as excessive time spent on infrastructure maintenance, slow deployment cycles, and the mounting cost of technical debt. Practitioners reported acute pain from operational complexity and reliability challenges, while managers struggled to balance innovation goals with resource constraints, reliability expectations, and compliance demands.

With the adoption of Union.ai, organizations experienced a measurable and meaningful operational shift. Developer maintenance time decreased by 60–65%, time-to-production accelerated by more than a week per deployment, and infrastructure cost avoidance reached $200K–$400K annually. Over a three-year period, these benefits generated a 982% ROI, $4.8M in net present value, and a 2.4-month payback period[1]. These outcomes were reinforced by customer feedback (e.g., Wherobots achieved a 50% reduction in engineering time, Pachama saw a 10x acceleration in iteration speed) and by broad survey results showing consistently high satisfaction scores (4–5/5).

Union.ai delivers a step-change in workflow automation, developer experience, and enterprise-grade reliability. For organizations seeking scalable, efficient, and compliant AI/ML operations, Union.ai is a proven catalyst for operational excellence and accelerated innovation. Ultimately, Union.ai helps organizations streamline innovation, optimize operational costs, and generate measurable performance gains.

Methodology

Research Approach

This economic validation employs theCUBE Research's analyst–led economic innovation index framework, which is designed to bring independent, technical rigor to financial impact analysis. The framework integrates both quantitative and qualitative data sources to ensure that findings are robust, defensible, and actionable. Quantitative inputs are drawn from a large–scale survey of AI/ML practitioners and managers, while qualitative depth comes from targeted customer interviews. A conservative risk–adjustment methodology is applied across the analysis, with benefits discounted and assumptions rigorously tested to preserve accuracy and credibility.

Data Collection

The quantitative foundation for this analysis is a proprietary survey fielded in October 2025, with 489 respondents representing a spectrum of company sizes (33.5% from 100–999 employees, 53.4% from 1,000–4,999 employees, and 13.1% from 5,000+ employees), industries (6.8% manufacturing, 5.9% telecommunications, 4.1% financial services, 3.3% retail, and more), and roles (15.1% individual contributors, 33.3% engineering managers or technical leads, 21.3% vice president or director, 18.2% C–level executive, and more). Survey questions covered workflow orchestration challenges, deployment frequency, reliability, cost structure, satisfaction, and business outcomes. To ensure the survey data reflected real–world Union.ai impact, respondent selection intentionally included organizations at different stages of workflow orchestration maturity.

The qualitative component consists of in–depth interviews with three Union.ai customers: Wherobots, Pachama, and HederaDx. These companies were selected to cover a range of use cases, technical environments, and business scales. Interview protocols focused on before–and–after operational metrics, pain points, productivity changes, and business value realization. The goal was to supplement statistical findings with concrete, narrative–driven evidence from the field.

Composite Organization Model

To anchor the financial analysis, we constructed a composite organization that mirrors the median survey respondent, validated by interview findings. This composite consists of 2,500 employees and a 15–person AI/ML team, with a representative mix of individual contributors and managers. Prior to Union. ai adoption, the composite operated self–managed Airflow for workflow orchestration, deploying new models or features on a weekly basis. These current state assumptions reflect common patterns observed among Union.ai's ideal customer profile (ICP), including mid–market and enterprise organizations with cloud– first architectures and significant scaling challenges, as well as AI–native startups that rely on training and deploying models to deliver their core product value, all facing growing compliance and operational efficiency requirements.

This composite model provides a realistic, defensible baseline for quantifying the incremental benefits and costs of Union.ai adoption, ensuring that results are both relevant for decision–makers and grounded in industry reality.

Current State Analysis

The AI/ML Orchestration Challenge

The industry landscape for AI/ML is increasingly shaped by operational friction rather than a lack of experimentation. While many organizations are actively building and deploying models, progress is frequently constrained by gaps in AI development infrastructure and operational readiness. Survey data supports this tension: 82% of respondents report skill gaps in AI/ML operations within their teams, with 31.3% describing these gaps as extremely prevalent and another 21.9% as significantly prevalent. These challenges are reflected in overall infrastructure sentiment, as only 16.5% of participants report being extremely satisfied with their current AI/ML software stack. The top pain points cited include reliability (33.3% of respondents), operational complexity (30.9%), compliance (15.7%), and escalating costs (7.8%).

Nearly half of respondents (49.3%) say compliance and data governance are a high priority when developing AI/ML systems, including 24% who rank it as a top priority

and 18.4% who consider it a moderate priority. As regulatory scrutiny over data governance and privacy inevitably grows, auditability and compliance of AI/ML workflows becomes a primary lever for improvement. This shows the importance of platforms that streamline orchestration, lower cognitive overhead, and enable teams to redirect effort from operational maintenance toward delivering new models and business value.

Current State Costs

Maintaining and operating AI software infrastructure has proven costly in both time and resources. Rather than being a rounding error, infrastructure overhead is showing up as real lost delivery time. In our survey, compute efficiency work alone accounts for weeks of engineering time per year for most organizations, driven by efforts to tune utilization, right-size resources, and troubleshoot performance issues. These efforts are further compounded by tool sprawl. With 75% of respondents relying on six to fifteen orchestration or

monitoring tools, engineers often lack a unified view of workload behavior, making compute optimization slower, more manual, and more error-prone. As a result, integration overhead and cognitive load directly extend the time required to address efficiency issues. Survey data highlights the scale of this drag: 43.8% of participants lose one to two weeks per project in a year addressing compute efficiency challenges, 28.4% lose two to four weeks, and a significant share experience even longer delays—11.3% report losing one to two months, and 6.1% more than two months annually.

Platform engineering capacity is emerging as a structural constraint in the current state. As AI initiatives grow more sophisticated and demanding, the operational burden placed on platform engineering teams is growing faster than headcount or tooling maturity. Interview data suggests that, if current patterns of tooling fragmentation and bespoke integration persist, platform engineering effort required to manage AI infrastructure and reliability will increase substantially over the coming years, placing significant strain on existing teams or necessitating additional hires. This trajectory creates increasing strain on existing platform engineers and forces organizations into difficult trade-offs: stretching teams beyond sustainable limits, slowing AI delivery, or hiring additional specialized staff to manage orchestration, reliability, and compliance. In many organizations, platform engineering is becoming a bottleneck not because of lack of skill, but because the scope of responsibility is expanding faster than the capacity to absorb it.

As a result, enterprise expectations around platform value are shifting. Organizations are moving away from long integration and payoff cycles in favor of solutions that deliver measurable outcomes within quarters, not years. In the survey, 11% of participants expect to see return on investment from a managed AI development platform immediately, while 71% expect ROI within three months to six months. This compressed time horizon reflects mounting pressure to translate AI investments into operational and financial results more quickly.

Practitioner vs. Manager Comparison

Survey data and interviews reveal a clear role-based perception gap across the AI development lifecycle, particularly around operational complexity and execution reliability. While both practitioners and managers acknowledge that AI systems are becoming harder to operate, practitioners are substantially more likely to identify operational complexity as their primary challenge, whereas managers place greater emphasis on reliability outcomes and governance concerns.

When asked to identify their top challenge in operating AI systems, 45% of practitioners cited the operational complexity of data, tools, and teams as their primary concern. By comparison, only 31.6% of managers selected operational complexity as their top issue. Managers were more likely to prioritize reliability of training, inference, and production workflows (36.3%) than practitioners (25%), suggesting that leadership tends to observe the downstream symptoms of complexity rather than the day-to-day strain it creates.

This divergence is further reflected in how frequently teams must intervene operationally. Practitioners report significantly higher retraining intensity: 28% of practitioners indicate that production AI models require daily retraining, compared to approximately 14% of managers. More broadly, over 80% of practitioners report retraining models on a quarterly or more frequent basis, versus roughly 60% of managers. For practitioners, this frequency reinforces the reality that instability is a continuous operating condition rather than an episodic event.

Reliability emerges as a shared concern across roles, but it is experienced differently. Practitioners experience reliability as persistent operational pressure, manifested through repeated interventions, debugging cycles, and incident response. Managers, in contrast, are more likely to frame reliability as an outcome-level risk to be managed through planning, tooling investments, or

governance controls. This framing gap obscures the cumulative operational cost borne by execution teams.

Tooling complexity amplifies this visibility gap. Practitioners are more likely to operate across multiple orchestration and monitoring tools simultaneously, absorbing the cognitive load and integration overhead required to keep systems running. While managers recognize orchestration and observability as important capabilities, survey responses suggest they underestimate how fragmented tooling translates into sustained execution drag. As a result, operational strain is often interpreted at the management level as a temporary inefficiency rather than a structural constraint on AI delivery.

This divergence exposes a growing visibility gap across AI organizations. Practitioners absorb the immediate cost of complexity through lost time, rework, and firefighting, while managers primarily observe lagging indicators such as slower deployment cycles, rising infrastructure costs, and missed timelines. Without shared metrics that explicitly connect practitioner–level operational friction to business outcomes, organizations risk normalizing inefficiency across the AI stack.

As AI initiatives face increasing scrutiny from executives, boards, and regulators, this misalignment carries tangible consequences. When managers underestimate the operational strain experienced by their teams, platform decisions may optimize for surface–level efficiency while compounding execution risk beneath the surface.

Bridging this gap requires not only improved tooling, but clearer operational signals that translate practitioner pain into metrics leadership can act on.

Quantified Pain Points

- **Developer productivity drain:** Survey and interviews converge on an estimate of 7,800 hours/year lost to maintenance and firefighting for a 15–person AI/ML team.

- **Slow deployment cycles:** Teams deploy new models or features weekly at best, with delays common due to infrastructure issues.

- **Infrastructure costs:** The composite organization spends approximately $800K annually on cloud and platform resources, with 15–20% attributed to inefficiencies and rework.

- **Staffing pressure:** There is a critical need for dedicated platform engineers; over 30% of surveyed organizations have had to hire or reassign staff solely to manage orchestration and reliability.

In summary, the current state is defined by high operational overhead, fragmented tooling, and significant opportunity costs, setting the stage for transformative improvement with a unified AI development infrastructure platform like Union.ai.

Union.ai Solution Overview

Platform Capabilities

Union.ai delivers an end-to-end platform for automation of AI/ML workflows, including workflow orchestration, training and fine-tuning, inference, and observability. It extends the functionality of Flyte, its open-source AI/ML solution, to offer a managed, enterprise-grade platform. The platform is built for cloud-native scalability, enabling seamless expansion across data, compute, and storage resources. A core focus on developer experience ensures intuitive interfaces, streamlined onboarding, and reduced friction for both individual contributors and engineering managers. Key reliability features include self-healing workflows (e.g., automatic retries with more resources) and robust error handling, minimizing disruptions in production environments. Union.ai also offers enterprise-grade compliance capabilities, such as reproducibility, data lineage, role-based access control (RBAC), and advanced observability, supporting highly regulated industries and audit requirements.

Key Differentiators

Union.ai stands out through measurable business impact and clear technical differentiation, driven by reductions in workflow friction and operational overhead. Survey results indicate that among Union.ai/Flyte users, reductions in workflow friction are most frequently cited as the factor contributing the greatest time and cost savings, reinforcing the platform's role in streamlining day-to-day AI/ML operations.

These benefits extend directly to developer productivity. Users report meaningful time savings each week (48% cite 5-10 hours saved per week), pointing to a consistent reduction in operational overhead and less time spent managing infrastructure and compute inefficiencies. This reclaimed capacity allows teams to focus more on building models, improving data pipelines, and accelerating time-to-value. Ease of use and scalability further distinguish the platform. 69.2% of respondents rate Union.ai and Flyte's ease of use and scalability as better than other workflow orchestration tools they have

used, with a significant portion (29.1%) describing it as much better. This combination supports adoption across both individual contributors and engineering leaders, lowering friction without compromising the ability to support complex, production-grade workflows.

Union.ai's approach is differentiated not only by its managed capabilities, but by the flexibility of its deployment options across fully managed BYOC environments, hybrid-managed models that allow organizations to retain greater control over data-sensitive components, and fully self-hosted deployments for the most regulated enterprise use cases. Rather than requiring teams to shoulder the full burden of ongoing platform maintenance, upgrades, and operational support in every scenario, Union.ai enables organizations to offload day-to-day operational complexity where appropriate while maintaining control where required. Across these deployment models, Union.ai integrates easily with existing ML stacks, including tools for data processing, model training, and deployment. Reliability is reflected not only in the platform's architecture, which supports long-running, complex AI workflows, but also in customer experience. Interview feedback consistently highlights responsive support, centralized execution, and reduced troubleshooting complexity as important contributors to confidence in production operations.

Implementation Approach

Most organizations complete migration to Union.ai within 6 weeks, typically involving a core team of four engineers. The onboarding process includes comprehensive training and hands-on support, enabling teams to deploy their first production workflow in less than one month. Union.ai's implementation methodology minimizes disruption, accelerates value realization, and ensures teams are equipped to leverage the platform's full capabilities from day one.

Economic Benefits Analysis

Developer Productivity Gains

Following adoption of Union.ai, customers reported material improvements in development velocity and focus. At Pachama, a core production algorithm achieved approximately 10× faster iteration, enabled by standardized workflows and parallel experimentation. At Wherobots, Union.ai reduced the effort required to deliver a new product capability, with the team estimating roughly 50% savings in time-to-market and build effort compared to implementing equivalent orchestration and execution internally. At HederaDx, leadership estimated 20-50% faster development velocity, alongside a shift toward a more stable operational posture, with maintenance activities comprising closer to 20% of engineering effort with Union.ai, compared to approximately 30% prior to adoption. For a 15-person AI/ML team, this equates to 4,680–5,070 hours saved annually. The financial value of these reclaimed hours is estimated at $561K–$608K per year, based on average fully loaded compensation rates. These gains are validated by both customer interviews and broad survey responses, with practitioners reporting more time for innovation and less time "firefighting."

Faster Time-to-Production

The typical time-to-production for new workflows in the baseline state was 2–3 weeks per deployment. After adopting Union.ai, teams report significantly faster deployment and development cycles, driven by reduced orchestration friction and more reliable paths from experimentation to production. As organizations mature on the platform, deployment velocity increases from 24 to 36 to 48 deployments per year. The financial impact of this acceleration is significant, with modeled value rising from $518K in Year 1 to $778K in Year 2 and $1.04M in Year 3, driven by faster delivery of business value and improved responsiveness to market opportunities.

Infrastructure Cost Avoidance

Union.ai enables infrastructure cost avoidance by extending the effective capacity of platform engineering teams, rather than replacing them. In an environment where the operational demands on platform engineers are accelerating, Union.ai reduces the incremental labor required to manage orchestration, upgrades, reliability, and compliance across the AI stack. Interview

participants estimate the fully burdened cost of a platform engineering role at $150K–$200K per year, factoring in salary and overhead.

By consolidating orchestration, monitoring, and governance functionality into a unified platform, Union.ai allows existing platform engineers to support growing AI workloads without a proportional increase in headcount. Organizations consistently describe this impact as equivalent to avoiding the need to hire one to two additional platform engineers, or materially reducing the overtime and operational load that would otherwise be required to keep pace with AI-driven complexity. Additional cost avoidance is realized through reduced tooling sprawl and lower licensing overhead, further amplifying the productivity of existing teams.

Infrastructure Efficiency

Union.ai's orchestration capabilities drive better resource utilization, auto-scaling, and optimization of cloud workloads. Survey data and interviews indicate a ~20% average reduction in infrastructure costs, primarily through reduced cloud waste and more efficient scheduling of compute and storage. For the composite

organization, this translates to $120K–$160K in annual savings on an $800K infrastructure budget.

Innovation Acceleration

By freeing up engineering capacity, Union.ai enables teams to redirect effort toward higher-value, revenue-generating initiatives. Interview participants report faster experimentation, shorter iteration cycles, and improved time-to-market for AI-driven features. Within the composite organization, this translated into the delivery of two to four additional models or features per year, with an estimated incremental revenue impact of $300K–$600K annually. These gains compound over time as teams build on early productivity improvements and stabilize their AI infrastructure.

Notably, this estimate is intentionally conservative and does not capture the additional economic value associated with outpacing competitors in markets where organizations are racing to implement transformative AI initiatives. In highly competitive environments, even modest acceleration in AI delivery can translate into disproportionate strategic and revenue advantages.

Cost Analysis

Year 1 Costs

The initial year of Union.ai adoption reflects an upfront guaranteed payment, ongoing usage–based platform fees, and internal resources allocated to implementation. The composite organization incurs:

Union.ai fee estimate:	$150,000
Migration engineering time:	$115,000 (based on 4 engineers over 6 weeks)
Training and onboarding	$25,000
Training and onboarding	**$290,000**

Years 2–3 Costs

Ongoing costs in subsequent years are streamlined, reflecting only recurring platform fees and minimal incremental support:

Union.ai fee estimate:	$150,000
Ongoing support/training:	Minimal, absorbed into regular team operations and supplemented by included white-glove support and FDE resources
Training and onboarding	**$150,000 per year**

Cost Comparison

Union.ai delivers a favorable total cost of ownership (TCO) compared to DIY and alternative managed platforms. DIY solutions incur a significant added burden to platform engineers, tooling sprawl, and higher maintenance overhead. In contrast, Union.ai consolidates orchestration, monitoring, and compliance on an open-source foundation, reducing both direct and indirect expenses. Compared to alternative managed platforms, Union.ai offers more predictable pricing and lower cost per deployment, particularly at higher deployment frequencies. For the composite organization, Union.ai's cost per deployment declines over time as operational efficiency improves, thus amplifying ROI.

Financial Summary

3-Year Financial Model

The financial model for the composite organization quantifies the annual benefits and costs of Union. ai adoption over a three-year horizon. In Year 1, total benefits reach $1.4M against costs of $290K, yielding a net benefit of $1.2M. By Year 3, cumulative benefits

exceed $6.3M while total costs remain under $600K, resulting in a cumulative net benefit of $5.71M[1]. The model is based on actual survey data, validated by customer interviews, and all assumptions are transparently documented.

Summary Table:

Year	Total Benefits	Total Costs	Net Benefit
1	$1.4M	$290K	$1.2M
2	$2M	$150K	$1.9M
3	$2.9M	$150K	$2.8M
Total	$6.3M	$590K	$5.71M

Cumulative net benefit and payback period are visualized in the accompanying chart (see full report for details).

Key Metrics

ROI:	982% (3-year)
Net Present Value (NPV)	$4.8M (10% discount rate)
Payback Period:	2.4 months
Benefit-to-Cost Ratio:	10.8:1

Risk-Adjusted Scenarios

To ensure conservative, defensible results, the economic model was stress-tested under multiple adverse conditions:

- **Built-in conservatism:** A 20% reduction was applied to all productivity-related benefits.

- **Downside sensitivity:** Scenarios assume either 50% lower realized benefits or 50% higher total costs.

- **Outcome resilience:** Even under these stressed assumptions, Union.ai adoption continues to deliver a positive net present value (NPV) and achieves payback within the first year.

Together, these results indicate that the economic case for Union.ai remains resilient under unfavorable conditions, giving decision-makers a high degree of confidence in the durability of the projected financial outcomes.

Customer Evidence

Customer Evidence: Case Study 1 – Wherobots

Company Profile

Wherobots is an AI-powered data processing startup specializing in advanced analytics and workflow automation for enterprise clients. The company's engineering team is tasked with rapidly developing and deploying new features to maintain a competitive edge in a fast-moving market.

Challenge

Wherobots faced increasing pressure to reduce time-to-market for new features while maintaining high reliability and compliance standards. The existing DIY orchestration stack required consistent maintenance and troubleshooting, consuming valuable developer time and delaying innovation. The team also anticipated the need to hire additional infrastructure engineers to keep pace with growth.

Solution

By adopting Union.ai's managed workflow orchestration platform, Wherobots was able to automate complex pipelines, streamline deployment processes, and offload infrastructure maintenance. The transition enabled the engineering team to focus on delivering value rather than maintaining tools.

Results

Wherobots achieved a 50% reduction in time-to-market for new features, enabling faster business impact and improved customer responsiveness. The organization also avoided the need to hire an additional infrastructure engineer, realizing $150,000–$200,000 in annual cost avoidance. Beyond the financial gains, teams reported higher developer satisfaction and a noticeable reduction in operational complexity.

Key Quote

"Union.ai enabled us to move twice as fast, without doubling our headcount. We're now focused on delivering new features, not fighting with infrastructure."

— Maxime Petazzoni, Wherobots

Customer Evidence:
Case Study 2 – Pachama

Company Profile

Pachama is a climate technology company focused on carbon credit verification and forest conservation, leveraging advanced AI/ML workflows to analyze satellite and environmental data. Rapid innovation and scalable, reliable infrastructure are mission-critical for their impact-driven engineering team.

Challenge

Pachama's growth was constrained by a self-managed Dataflow stack that struggled to support scaling requirements and complex, reproducible workflows. Developers faced bottlenecks from infrastructure toil, slow iteration cycles, and limited automation, which hindered both speed and scientific rigor.

Solution

Pachama migrated to Union.ai's AI development platform, enabling robust orchestration, reproducibility, and cloud-native scalability. The transition eliminated manual infrastructure management and streamlined the deployment of new data pipelines and models.

Results

Pachama achieved a 10x increase in iteration speed, enabling faster experimentation and quicker delivery of new features. The organization also significantly reduced infrastructure toil, allowing engineers to focus more on core climate science and product development. These improvements, combined with stronger reproducibility and compliance, enhanced Pachama's operational credibility within the carbon credit market.

Key Quote

"Union.ai took infrastructure off our plate and made it possible to iterate ten times faster. We can now focus on building models that matter."

— Marie Hoeger, Pachama

Customer Evidence:
Case Study 3 – HederaDx

Company Profile

HederaDx is a bioinformatics company specializing in liquid biopsy diagnostics, utilizing sophisticated AI/ML workflows to analyze large-scale genomic and clinical data. The team's mission is to deliver fast and accurate diagnostic insights for healthcare providers and patients.

Challenge

HederaDx faced the challenge of managing complex, data-intensive workflows without in-house Kubernetes expertise. The need to ensure reliability, reproducibility, and compliance placed significant strain on limited engineering resources. Manual error tracking and infrastructure maintenance created bottlenecks and increased operational risk.

Solution

By adopting Union.ai's managed orchestration platform, HederaDx centralized error management, automated

pipeline operations, and eliminated the need for dedicated platform engineering hires. The platform's built-in compliance and observability features enabled the team to meet stringent healthcare and data security requirements without additional tooling overhead.

Results

HederaDx avoided $150,000–$200,000 per resource in annual costs by eliminating the need for additional platform engineering hires. By centralizing error management, the team reduced downtime and accelerated troubleshooting, while improvements in developer productivity and workflow reliability enabled faster delivery of diagnostic insights.

Key Quote

"Union.ai allowed us to operate complex bioinformatics pipelines without building out a specialized infrastructure team. We're delivering results faster and with greater confidence."

— Thomas Vetterli, HederaDx

Industry Context & Benchmarking

Market Landscape

AI/ML Orchestration Market Overview

The AI/ML orchestration market is experiencing significant expansion as organizations across industries invest in automation, data-driven decision-making, and scalable machine learning workflows. According to theCUBE Research's proprietary research and survey data, over 80% of mid-market and enterprise organizations, as well as a growing share of AI-native startups and small businesses, have launched or plan to launch AI/ML initiatives in the next 12–18 months. As these initiatives mature, the need for robust, reliable orchestration platforms becomes critical to managing complexity and accelerating value realization.

Growth Drivers

Key growth drivers include:

- **Widespread AI adoption:** Survey results show a year-over-year increase in AI/ML project launches, with 62% of respondents citing AI as a strategic priority.

- **Workflow complexity:** As organizations scale, pipelines become more intricate, spanning multiple data sources, models, and deployment environments. 54% of survey respondents report significant challenges managing cross-team and cross-environment orchestration.

- **Scalability requirements:** The shift to cloud-native architectures and distributed teams demands platforms that can scale seamlessly and support rapid iteration.

Competitive Landscape: DIY vs. Managed Solutions

The AI orchestration market is increasingly defined by a split between DIY frameworks and managed AI development platforms, with most enterprises now operating somewhere in between. Survey data shows that organizations are no longer debating whether to operationalize AI, but how to do so without absorbing unsustainable levels of complexity, cost volatility, and execution risk.

DIY orchestration approaches remain prevalent, particularly among organizations with established platform engineering teams. Tools such as Apache Airflow, Sagemaker, and Temporal are widely used to coordinate training and data workflows. However, the survey highlights the operational burden that accompanies these approaches as AI systems scale. When asked about the primary challenge in moving AI from experimentation to production, 31.9% of respondents cited operational complexity across data, tools, and teams, nearly matching 33.3% who cited reliability of training, inference, and production workflows. This parity suggests that complexity itself is now as constraining as reliability failures.

The cost of this complexity is visible in time lost to infrastructure management. A majority of respondents report spending significant engineering time addressing compute efficiency issues, with 43.8% losing one to two weeks per project, 28.4% losing two to four weeks per project, and over 17% losing one month or more per project annually to compute optimization and infrastructure tuning. As retraining intensity increases, these overheads compound and directly slow iteration cycles.

Managed AI orchestration platforms are increasingly evaluated as a way to reduce this operational drag, but the data does not indicate a wholesale shift away from DIY tooling. Instead, enterprises are converging on hybrid orchestration strategies. The most common future-state approach reported is a hybrid mix of DIY and managed platforms (41.8%), followed by adoption of a managed AI development platform at approximately 28%. Pure DIY and fully outsourced strategies remain minority positions. This reflects a pragmatic market response: organizations are selectively offloading the most failure-prone and labor-intensive aspects of AI orchestration while preserving flexibility where it still delivers value.

The decision to adopt managed platforms is increasingly driven by operational outcomes rather than experimentation speed. When asked what capabilities most influence a shift toward managed platforms, respondents ranked cost optimization and predictability above faster onboarding and usability, signaling a maturation in buying criteria. Time-to-value expectations further reinforce this trend, with three to six months emerging as the most common window for realizing ROI from managed platforms, compared to longer and less predictable payoff periods for internally built solutions.

Taken together, the data suggests that the DIY vs. managed decision is no longer ideological. As AI workloads move deeper into production and face increasing scrutiny around reliability, cost control, and governance, organizations are prioritizing orchestration models that reduce recurring operational effort. Managed platforms are not replacing DIY frameworks outright, but they are increasingly used to absorb complexity that would otherwise scale linearly with model count, retraining frequency, and regulatory pressure.

Competitive Comparison: General-Purpose Orchestration vs. AI-Native Managed Platforms

As AI systems scale into production, orchestration platforms increasingly differentiate based on where operational complexity is absorbed. Survey data shows that reliability, retraining frequency, cost control, and governance are now persistent operational conditions, not edge cases. Against that backdrop, general-purpose orchestration tools and AI-native managed platforms address fundamentally different parts of the problem.

Apache Airflow remains widely adopted due to ecosystem maturity and familiarity, particularly for batch-oriented data pipelines. However, extending Airflow into production AI workflows typically requires additional tooling for model versioning, retraining loops, observability, and governance. This aligns with survey findings that 31.9% of respondents cite operational complexity as a primary challenge, nearly equal to reliability concerns at 33.3%, indicating that stitching together AI pipelines across multiple systems has become a first-order constraint.

Temporal provides strong guarantees for durable, long-running workflows and is well-suited for complex coordination logic. While this addresses certain reliability concerns, the survey data indicates that reliability alone does not eliminate execution drag. More than 70% of respondents report losing weeks or months per year

to compute efficiency and infrastructure optimization, reflecting ongoing effort that general-purpose workflow engines do not abstract away. For AI teams, Temporal still requires significant domain-specific engineering to manage training, retraining, lineage, and compliance workflows.

SageMaker offers a more integrated, managed approach to AI development within AWS, reducing initial setup and centralizing lifecycle components. However, the survey reflects a market dominated by hybrid strategies, with 41.8% of organizations planning to mix DIY and managed platforms, underscoring the need for portability and interoperability across environments. As cost predictability emerges as a top driver for adopting managed platforms, ranked ahead of onboarding speed, platform lock-in and opaque cost dynamics become more salient at scale.

AI-native managed orchestration platforms, such as Union.ai, are designed explicitly around the operational patterns highlighted in the survey: frequent retraining, continuous iteration, cost variability, and expanding governance requirements. Rather than acting as generic workflow engines, these platforms standardize core AI lifecycle operations and embed observability, reproducibility, and policy controls directly into orchestration. This approach aligns with survey findings that organizations expect measurable ROI within three to six months from managed platforms and are increasingly using them to absorb recurring operational complexity rather than replace existing tools outright.

Benchmark Comparisons

Union.ai Customers vs. Industry Median

Union.ai customers consistently outperform the industry median across key operational and business metrics, as validated by both survey data and direct customer interviews.

Metric	Union.ai Customers	Industry Median	Observed Impact
Time-to-Production	< 1 week for new workflows	2–3 weeks	Faster delivery of AI value and greater business agility
Customer Satisfactionv	4–5 out of 5	3–4 out of 5	Stronger developer experience and support responsiveness
Operational & Infrastructure Cost Efficiency	50%+ reduction	Baseline	Lower maintenance burden, reduced tool sprawl, and avoided platform engineering hires
Operational & Infrastructure Cost Efficiency	50%+ reduction	Baseline	Greater confidence in scaling advanced AI/ML workflows

These benchmark comparisons show that Union. ai customers consistently outperform the industry median across speed, satisfaction, cost efficiency, and operational confidence, which reinforces Union.ai's role as a catalyst for both operational excellence and sustained AI-driven innovation.

theCUBE Research Analyst Perspective

Trends in Day 0/1/2 Lifecycle Management

theCUBE Research's syndicated research highlights a growing emphasis on holistic lifecycle management in AI/ML, from initial environment setup (Day 0) through deployment (Day 1) and ongoing operations (Day 2). Organizations increasingly recognize that success in AI/ML hinges not just on model development, but on streamlining the entire lifecycle, including compliance, monitoring, and continuous improvement. Survey data shows that 68% of decision-makers cite end-to-end orchestration as a top future investment priority.

The Shift to Managed Platforms

There is a clear market shift from DIY and self-managed solutions to managed platforms that provide reliability, compliance, and operational efficiency out of the box. This transition is driven by the need to reduce technical debt, accelerate innovation, and ensure enterprise-grade support. Managed orchestration platforms like Union.ai are now seen as critical enablers of AI/ML scale, freeing teams to focus on core business objectives rather than infrastructure management.

Why Orchestration is Critical for AI/ML Success

Effective workflow orchestration ensures workflows are scalable, reproducible, and resilient. As AI/ML initiatives become more central to business strategy, orchestration platforms serve as the backbone for innovation, compliance, and operational excellence. Without robust orchestration, organizations face bottlenecks, increased risk, and limited ROI from their AI/ML investments.

Union.ai Positioning in the Market

Union.ai is well-situated at the intersection of technical rigor and operational simplicity. Union.ai and Flyte deliver managed workflow orchestration, strong compliance features, and a developer-centric experience that address the most pressing needs identified in both the survey research and customer interviews. theCUBE Research's analyst perspective reinforces Union.ai's differentiation: it delivers the reliability and agility that modern AI/ML teams require, while reducing the burden of infrastructure management and enabling faster, more confident innovation.

Recommendations & Next Steps

Ideal Customer Profile

Union.ai aligns most closely with organizations across startup, mid-market, and enterprise segments who operate complex, rapidly evolving AI/ML workflows and seek to reduce execution-layer operational burden as those systems scale. Teams that are constrained by infrastructure maintenance, facing bottlenecks in deployment cycles, or struggling to meet compliance and scalability requirements will realize the greatest value. The platform is especially impactful for organizations seeking to accelerate innovation while maintaining operational rigor and cost efficiency.

Implementation Roadmap

A typical Union.ai implementation begins with a foc migration plan, involving a core engineering team over a 6-week timeline. The process includes comprehensive training, platform configuration, and support for initial production deployments. Most organizations see their first workflows running in production within the first month, with full migration and optimization completed by week six. Ongoing support and best practices ensure teams continue to maximize value post-launch.

ROI Realization Timeline

Union.ai delivers measurable financial and operational benefits within the first quarter of adoption:

- **Time-to-value:** Most teams achieve a payback period of 2.4 months, with ROI surpassing 980% over three years.

- **Operational transformation:** Developer productivity gains, faster deployment cycles, and infrastructure cost avoidance are realized almost immediately, building a strong foundation for continuous innovation.

- **Strategic impact:** By unifying orchestration, training & fine-tuning, inference, and compliance, Union. ai positions organizations to capitalize on AI/ML opportunities, reduce risk, and achieve sustained competitive advantage.

In summary, Union.ai is a catalyst for operational excellence, rapid innovation, and measurable business value. The platform consistently delivers on three core promises highlighted throughout this report: transformative improvements in developer productivity, significantly faster time-to-production, and durable cost efficiency at scale. Together, these advantages enable organizations to operationalize AI with greater confidence, speed, and financial clarity.
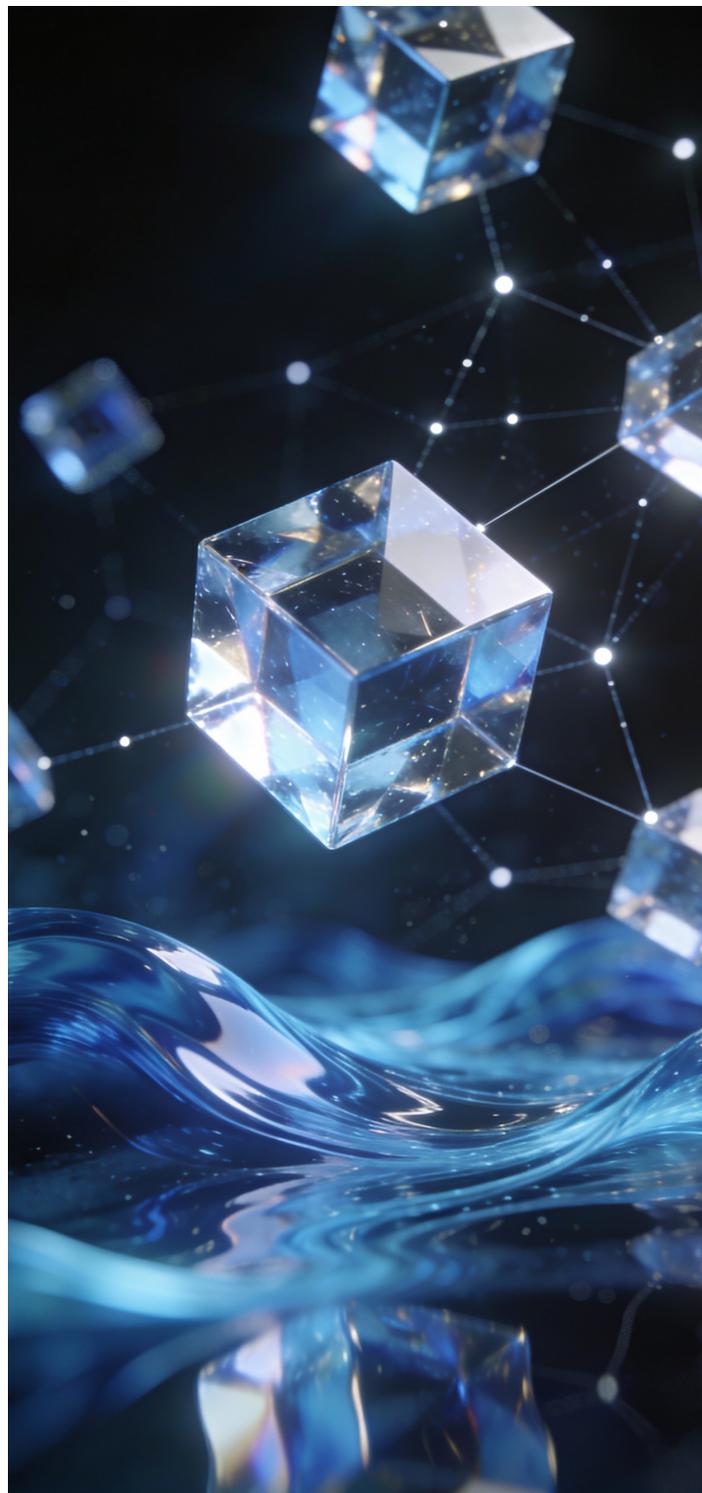
Appendix

Financial Model – 3–Year Analysis ROI

Union.ai Economic Validation – Financial Model

Composite Organization Profile

Company Size:
Mid-market technology company

Employees:
2,500

AI/ML Team Size:
15 engineers
(data scientists, ML engineers, platform engineers)

Annual Team Fully–Loaded Cost:
$3,750,000 ($250k average per engineer)

Current State:
Self-managed orchestration (Airflow + custom tooling)

Deployment Frequency:
Weekly to multiple times per week

Annual Infrastructure Spend (AI/ML):
$800,000

Key Assumptions (Data Sources)

Survey Data Insights (n=~500)

- **Budget Allocation:** 10–25% of IT budget to data platforms (most common range)

- **Pipeline Automation:** 26–50% automated (most common)

- **Top Challenges:** Scaling for AI, cost, quality issues, skills shortage

- **Deployment Frequency:** Daily to weekly for most organizations

- **Satisfaction with Managed Platforms:** 4–5/5 rating (extremely satisfied)

- **Time to Production:** Less than 1 week for managed platforms

- **Cost Reduction:** Greater than 50% reported by managed platform users

Customer Interview Insights

- **Wherobots:** 50% time-to-market savings, avoided $150–200k infrastructure hire

- **HederaDx:** $150–200k per resource saved, 4/5 recommendation

- **Pachama:** 10x iteration speed increase, dramatic reduction in infrastructure toil

Financial Model – 3–Year Analysis

Year 1: Implementation & Ramp

Benefits

1. Developer Productivity Gains –
Baseline: 15 engineers × 40 hours/week × 52 weeks = 31,200 hours/year – Time spent on infrastructure/orchestration maintenance: 25% = 7,800 hours – Union.ai reduction: 60% of maintenance time = 4,680 hours saved – Value: 4,680 hours × $120/hour = **$561,600**

2. Faster Time–to–Production –
urrent average: 2–3 weeks per model deployment – With Union: Less than 1 week (survey data) – Time savings: 1.5 weeks average × 24 deployments/year = 36 weeks saved – Team capacity freed: 36 weeks × 3 engineers average = 108 engineer–weeks – Value: 108 weeks × 40 hours × $120/hour = **$518,400**

3. Infrastructure Cost Avoidance –
Avoided dedicated platform engineer hire: **$200,000** (fully loaded) – Reduced infrastructure complexity/tooling: **$50,000** – Total: **$250,000**

4. Infrastructure Efficiency –
Improved resource utilization through better orchestration:
15% reduction – Current spend: $800,000 – Savings: **$120,000**

Year 1 Total Benefits: $1,450,000

Costs

1. Union.ai Platform Fees –
Enterprise annual subscription: **$150,000**

2. Migration & Implementation –
Engineering time: 4 engineers × 6 weeks × 40 hours
× $120/hour = $115,200 – Training & onboarding: **$25,000** –
Total: **$140,200**

Year 1 Total Costs: $290,200

Year 1 Net Benefit: $1,159,800

Year 2: Full Production

Benefits

1. Developer Productivity Gains – Full year at 60%
maintenance reduction: **$561,600**

2. Faster Time-to-Production – Increased deployment
frequency: 36 deployments/year (50% increase) – Time
savings: 1.5 weeks × 36 = 54 weeks saved – Value: 54 weeks ×
40 hours × 3 engineers × $120/hour = **$777,600**

3. Infrastructure Cost Avoidance – Avoided platform
engineer: **$200,000** – Avoided additional tooling/complexity:
$75,000 – Total: **$275,000**

4. Infrastructure Efficiency – Sustained 15%
reduction: **$120,000**

5. Innovation Acceleration – Freed capacity enables 2
new revenue-generating models – Conservative revenue
impact: **$300,000**

Year 2 Total Benefits: $2,034,200

Costs

1. Union.ai Platform Fees: $150,000

Year 2 Total Costs: $150,000

Year 2 Net Benefit: $1,884,200

Year 3: Optimization & Scale

Benefits

1. Developer Productivity Gains – Full optimization at 65%
maintenance reduction: $608,400

2. Faster Time-to-Production – 48 deployments/year
(doubled from baseline) – Time savings: 1.5 weeks × 48 = 72
weeks – Value: 72 weeks × 40 hours × 3 engineers ×
$120/hour = **$1,036,800**

3. Infrastructure Cost Avoidance – Avoided 2 platform
engineers: **$400,000** – Avoided tooling complexity: **$100,000**
– Total: **$500,000**

4. Infrastructure Efficiency – 20% reduction through
maturity: **$160,000**

5. Innovation Acceleration – 4 new revenue-generating
models: **$600,000**

Year 3 Total Benefits: $2,905,200

Costs

1. Union.ai Platform Fees: $150,000

Year 3 Total Costs: $150,000

Year 3 Net Benefit: $2,755,200

3-Year Summary

Metric	Year 1	Year 2	Year 3	Total
Total Benefits	$1,450,000	$2,034,200		
$2,905,200	$6,389,400			
Total Costs	$290,200	$150,000	$150,000	
$590,200				
Net Benefit	$1,159,800	$1,884,200	$2,755,200	
$5,799,200				

Key Financial Metrics

ROI: 982% (3-year)

NPV (10% discount rate): $4,847,000

Payback Period: 2.4 months

Benefit-to-Cost Ratio: 10.8:1

Risk Adjustments

Conservative Adjustments Applied: – Developer productivity: Reduced by 20% from interview findings – Time-to-production: Conservative 1.5 weeks vs. customer reports of 50%+ savings – Infrastructure efficiency: 15–20% vs. survey reports of 50%+ reduction – Innovation revenue: Highly conservative estimates

Sensitivity Analysis: – If benefits are 50% lower: ROI = 491%, Payback = 4.8 months – If costs are 50% higher: ROI = 656%, Payback = 3.6 months – Model remains highly favorable under stress scenarios
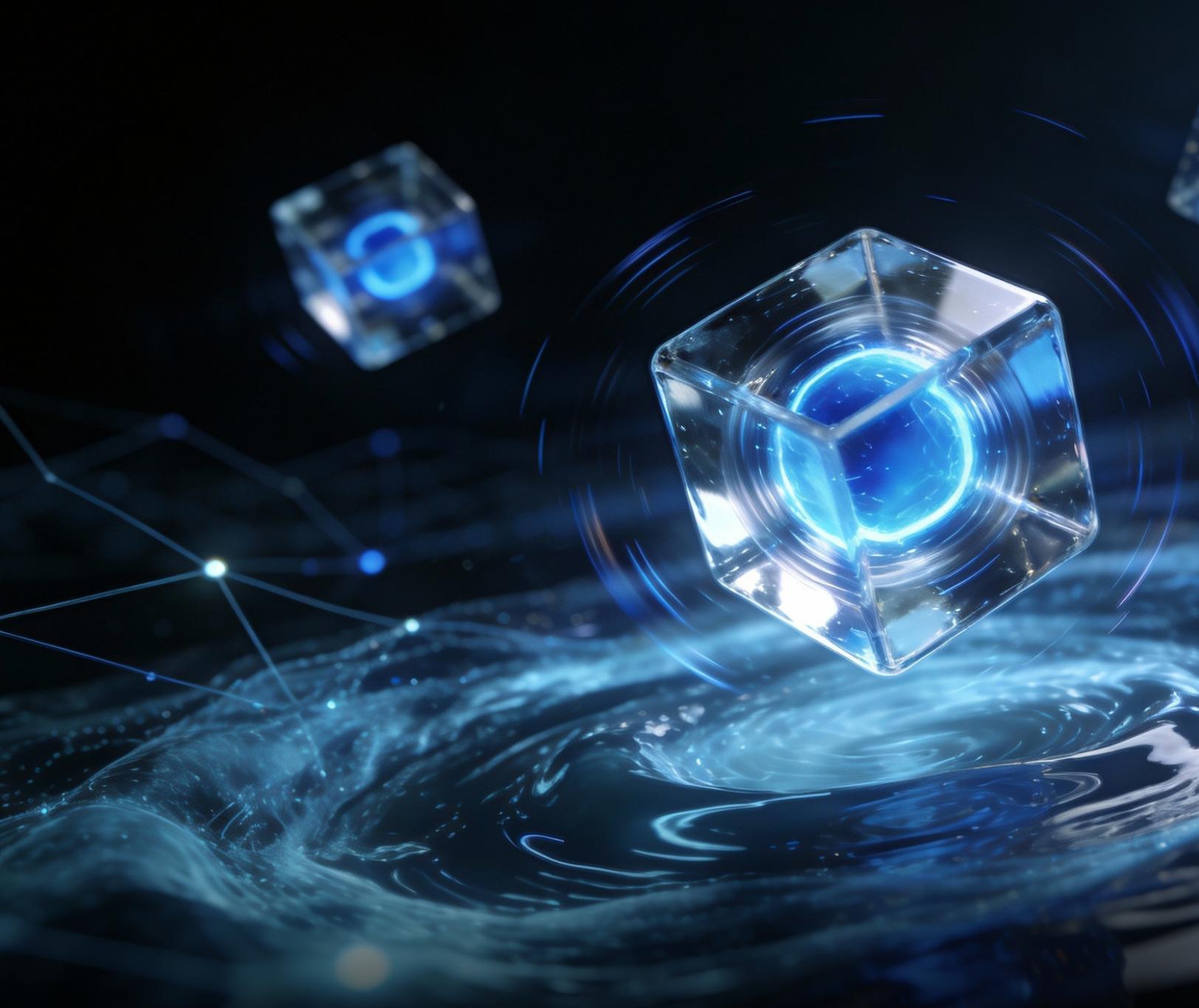
Data Source Attribution

Survey Data: Union.ai Market Research (Oct 2025, n=~500)

Customer Interviews: Wherobots, Pachama, HederaDx

Industry Benchmarks: ECI proprietary research

Financial Assumptions: Conservtative estimates based on composite organization model

Disclaimer

All trademark names are the property of their respective companies. Information contained in this publication has been obtained by sources theCUBE Research, a SiliconANGLE Media company, considers to be reliable but is not warranted by theCUBE Research. The publication may contain opinions of theCUBE Research, which are subject to change. This publication is copyrighted by theCUBE Research, a SiliconANGLE Media company.

Contact

Silicon Valley
989 Commercial Street
Palo Alto, CA 94303

Boston Metro
95 Mount Royal Avenue
Marlborough, MA 01752

David Butler
david.butler@siliconangle.com
774-463-3400

theCUBE research