trebllе

Presenting Global API Trends

Anatomy of an API

2025 EDITION

Enterprise Technology Insights
from 1 Billion API Requests

Anatomy of an API: 2025 Edition

Enterprise Technology Insights from 1 Billion API Requests

Anatomy of an API 2025

Author: Vedran Cindrić
Research & Data Analysis: Vedran Cindrić, Tea Cindrić
Design: Mateo Galić

Special Thanks

Treblle team that seamlessly handles 3 billion API requests and 9 Terabytes of data each month: Akhil Jayaraj, Bhushan Gaykawad. Dubravko Antunović, Filip Leskovec, Harsha Chelle, Ivan Kulić, Ivan Mikodanić, Iva Bačani, Iva Ivanković, Lucija Frljak, Mateo Galić, Rahul Khinchi, Savan Kharod, Scott Ehrlich, Tea Cindrić.

API experts and professionals who contributed to this report: Mark Boyd, Erik Wilde, Emmanuel Paraskakis, James Higginbotham, Mark O'Neill

Table of Contents

Executive Summary

This year, Treblle analyzed over **1 billion API requests** across the global digital economy. The data reveals a landscape defined by a sharp paradox: organizations have successfully engineered for speed, efficiency, and AI capability, but they have dangerously neglected the foundations of security and governance. **We are driving faster than ever, but largely without seatbelts.**

For the modern enterprise leader, this report is not merely a collection of metrics; it is a warning signal. The data suggests that while the "AI Revolution" has normalized, a *"Complexity Crisis"* is taking its place. We are building larger, faster, and more capable APIs, but we are leaving them unencrypted, unauthenticated, and unmanaged at alarming rates.

Here are the five critical signals defining the API landscape in 2025.

1. AI's Gold Rush Phase Is Over

The frantic "try everything" era has ended. Following an explosive 807% growth in AI-related API traffic in 2024, 2025 saw growth stabilize to a "normalized" 42%.

This does not mean AI is fading; it means it is maturing. Organizations have moved past the experimentation phase of 2024 and are now focusing on integrating AI into sustainable, production-grade workflows.

2. The Need for Speed: APIs Are Twice as Fast

Performance optimization has become the primary engineering KPI. The average global API load time dropped from 695ms in 2024 to just 322ms in 2025.

This massive 53% reduction in latency is not accidental. In an economy where AI agents demand near-instant responses to function, latency is no longer just a user experience metric; it is an operational necessity.

3. The Great API Security Illusion

We are moving fast, but we are exposed. There is a widening gap between how secure we think we are and what the traffic actually shows:
42% of traffic is still unencrypted HTTP.
47% of APIs process requests without Authentication.
46% of APIs lack any formal versioning strategy.

This is the "Governance Paradox." While we invest in advanced firewalls and AI defense, basic hygiene, such as SSL and Auth, is being skipped, likely due to the assumption that "internal" traffic is safe traffic. In a Zero Trust world, this assumption is a liability.

4. APIs Are Now Doing the Heavy Lifting

APIs have evolved from "Reading" data to "Doing" work. POST requests, used to store data and trigger actions, have nearly doubled, now accounting for 43% of all traffic.

Historically, APIs were read-heavy (GET requests), functioning as digital libraries. Today, they are digital factories. This shift indicates that APIs are becoming the engines of core business operations and AI agent actions, moving money, processing orders, and generating content rather than just displaying it.

5. The Rise of the Mega-API

Microservices are consolidating into massive platforms. The number of APIs with 100+ endpoints grew tenfold, from just 4% in 2024 to 38% in 2025.

The era of small, single-purpose microservices is waning. To support complex AI contexts (MCP) and reduce network chatter, teams are building massive, feature-rich "Mega-APIs." This consolidation simplifies connectivity but explodes complexity, creating "too big to fail" infrastructure that is increasingly difficult to document and maintain.

Letter from the CEO

Three months ago, my life changed in a way that no amount of data analysis could predict: I became a dad for the first time.

Those early days are a blur of joy, exhaustion, and confusion. But as I was holding my son, trying to decipher his various cries at 3 AM, my engineer brain couldn't help but make a connection. He reminded me of an API - in a sense. I know, weird, but hear me out.

Why? Well, just like many of the APIs we analyze in this report, he didn't come with documentation. Or, if he did, it was severely outdated. I was sending requests like food, sleep, rocking, and getting wildly inconsistent responses. It took time for him to get used to us, and even longer for us to understand his endpoints.

A Year of Velocity

While my personal life was slowing down to the rhythm of a newborn, life at Treblle was moving at breakneck speed. 2025 has been the most intense year of growth in our history.

We made a promise to ourselves to push the boundaries of what an API Intelligence platform could be, and we delivered. We shipped major product updates every single month. We launched entirely new products for **API Security, Discovery, Governance,** and **Agentic AI,** transforming Treblle from an observability tool into a comprehensive enterprise intelligence platform for the API economy.

Also, this year we welcomed some of the world's largest banks, insurance giants, and financial services firms to the Treblle platform. Getting mentioned in the **Gartner Magic Quadrant** for API Management was not just a badge of honor; it was validation that our approach, focusing on real-time, run-time data, is exactly what the enterprises need.

The Evolution of our Report

This report you are reading is very special to me. It started back in 2022 as a humble, single-page visualization, a snapshot of what we were seeing. Today, it has grown into a comprehensive industry benchmark.

What makes the Anatomy of an API unique is its source. We are not asking API professionals what they think they are doing in a survey. We are looking at what they are actually doing in production. This is run-time data. It doesn't lie, it doesn't have a recency bias, and it doesn't sugarcoat the truth.

Because of this, we can see the reality of the industry with perfect clarity. And honestly? I am incredibly optimistic.

For years, APIs were treated as plumbing - necessary, but invisible. The data in this report shows a shift. I am happy to see companies finally taking APIs seriously.

However, the data also reveals the growing pains. Governance, Security, and Discovery are still lagging behind raw speed, which comes at the cost of low or now security. Getting security up to speed is the next great challenge we must solve. But the fact that these conversations are now happening at the C-level, rather than just in the engineering slack channels, is massive progress.

As I look at my son, and then at the industry we are building, I am excited. We are entering an era where AI is making software more capable, more autonomous, and more interconnected than ever before. And the glue holding it all together?
It's APIs.

Today, every business is a digital business, and every digital business runs on APIs. Whether you are a bank, a retailer, or a startup, your APIs are your products.

Vedran Cindrić
CEO @ Treblle

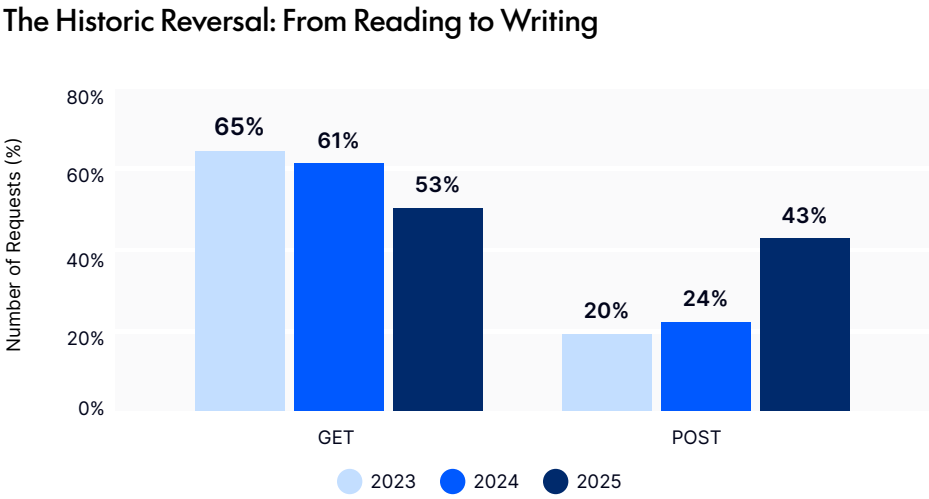
Vedran Cindrić

The Fundamental Shift

The data from 1 billion API requests reveals a fundamental transformation in how the world uses and thinks about APIs. No longer relegated to backend technical tools, APIs have emerged as the primary interface for business operations, customer experiences, and strategic innovation. This shift isn't gradual; it's a dramatic acceleration that redefines the role of APIs in the enterprise technology stack.

The Historic Reversal: From Reading to Writing

For the first time in API history, we're witnessing a fundamental behavioral shift that challenges decades of conventional wisdom about how APIs function in enterprise environments.



HTTP Method change YoY, *source: Treblle; sample data 2025*

The numbers tell a compelling story: GET requests, which have dominated API traffic since the inception of REST architecture, dropped from 61% to 53% of all traffic. Meanwhile, POST requests surged by an unprecedented 80%, growing from 24% to 43% of all API calls. This isn't just a statistical anomaly; it represents a fundamental reimagining of what APIs do in the enterprise.

This shift has profound implications for enterprise architects and IT leaders. APIs are no longer primarily data retrieval mechanisms; they've become transactional platforms where value is created, not just accessed. The surge in POST requests indicates that APIs are increasingly used for:

- **AI Model Interactions:** Every prompt sent to an LLM, every image generated, every prediction requested - these all require POST requests carrying complex payloads
- **Real-time Collaboration:** Document editing, messaging, and collaborative workflows depend on constant write operations
- **IoT Data Ingestion:** Billions of sensors and devices reporting status, metrics, and events
- **Transaction Processing:** Financial transactions, order processing, and supply chain updates

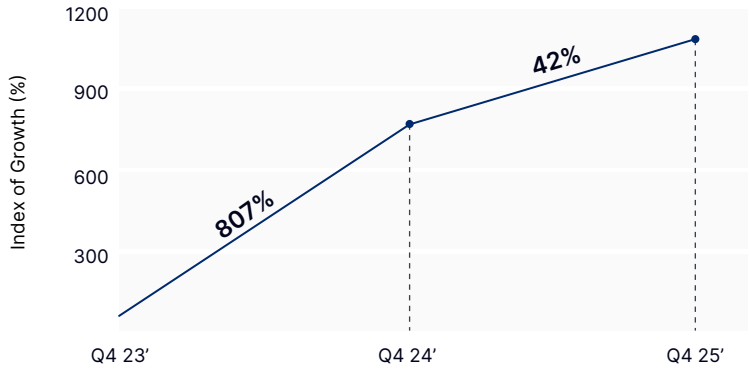
For Fortune 500 companies, this means rethinking API strategy entirely. The traditional model of APIs as read-heavy interfaces for accessing corporate data is obsolete. Today's APIs must be engineered for write-heavy workloads, with all the implications for database design, caching strategies, and infrastructure scaling that entails.

The DELETE and PATCH methods also tell an interesting story, with DELETE dropping 87% and PATCH falling 97%. This suggests enterprises are moving away from complex state management toward simpler, more robust patterns - likely creating new resources rather than modifying existing ones, a pattern that aligns with event-sourcing and immutable architecture principles gaining traction in enterprise environments.

The AI Revolution: APIs as the Nervous System of Intelligence

The relationship between APIs and AI isn't just symbiotic - it's existential. **There is no AI without APIs.** Every major AI breakthrough, every enterprise AI implementation, and every automated workflow depends entirely on APIs to function.

AI API Volume: Year-over-Year Comparison



AI API Volume YoY, *source: Treblle; sample data 2025*

While the 42% growth from 2024 to 2025 might seem like a deceleration compared to the explosive 807% growth the previous year, this normalization actually signals market maturity. We've moved from experimental adoption to production deployment.

Consider what this means in practical terms for enterprise IT leaders:

- **Large Language Models** require APIs for every interaction - OpenAI, Anthropic, and Google's models all operate exclusively through API interfaces
- **Model Context Protocol (MCP) servers** are emerging as the standard for connecting AI systems to enterprise data through APIs
- **RAG (Retrieval-Augmented Generation)** implementations depend on APIs to access knowledge bases, databases, and document stores

- **AI Orchestration platforms** like LangChain and AutoGPT coordinate multiple APIs to accomplish complex tasks

The cumulative 1,185% growth since 2023 represents one of the fastest technology adoptions in enterprise history, surpassing even the cloud migration wave of the 2010s. For CIOs and CTOs, this raises critical questions about API governance, security, and scalability. When every AI interaction is an API call, API performance directly determines AI performance. A 100ms delay in API response time could mean the difference between a usable and unusable AI assistant.

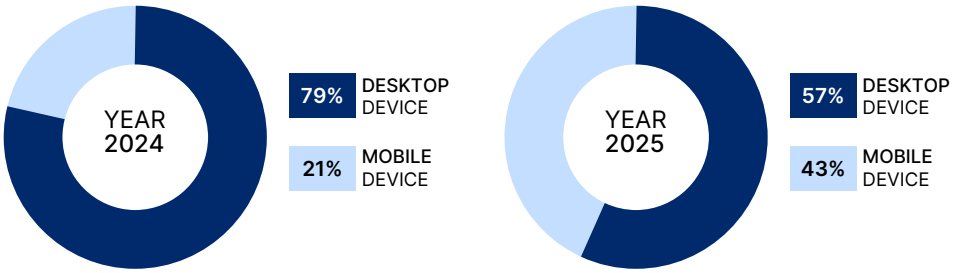
The normalization of growth from 807% to 42% isn't a slowdown - it's a consolidation. Enterprises are moving from "we need AI" to "we need AI that works reliably at scale." This maturation phase demands:

- **Enterprise-grade reliability:** AI APIs need 99.99% uptime
- **Predictable latency:** Consistent sub-second response times
- **Scalable architecture:** Handling millions of API calls daily
- **Cost optimization:** AI API calls can quickly become the largest line item in IT budgets

Mobile Dominance: The Untethered Enterprise

The doubling of mobile API traffic from 21% to 43% in a single year represents the fastest platform shift we've observed in enterprise computing.

Mobile vs. Desktop Traffic (2024 - 2025)



Mobile vs. Desktop Traffic, *source: Treblle; sample data 2025*

This isn't just about employees checking email on phones. The mobile explosion reflects fundamental changes in how enterprises operate:

Field Operations Revolution Manufacturing, logistics, healthcare, and retail enterprises are deploying mobile-first applications for frontline workers. These aren't simplified versions of desktop apps - they're sophisticated platforms handling complex transactions, real-time data processing, and AI-powered decision support.

Executive Decision Platforms C-suite executives increasingly expect full operational visibility and decision-making capability from mobile devices. API-powered executive dashboards, approval workflows, and analytics platforms must deliver desktop-class functionality with mobile-optimized performance.

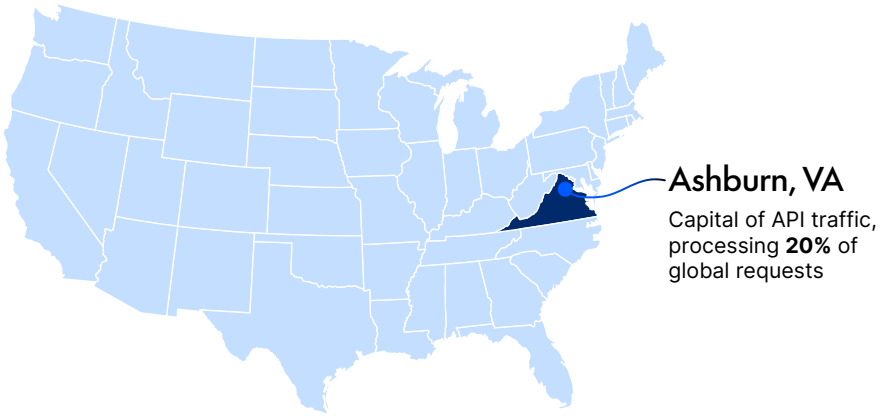
Customer Experience Transformation B2B customers expect the same mobile experience they get from consumer apps. Whether it's a procurement platform, a banking interface, or a healthcare portal, mobile-first is now table stakes.

For enterprise architects, this mobile surge demands rethinking API design principles:

- **Payload optimization:** Mobile networks demand smaller, more efficient data transfers
- **Offline resilience:** APIs must gracefully handle intermittent connectivity
- Battery efficiency: Polling strategies and connection management affect device battery life
- **Security challenges:** Mobile devices operating outside corporate networks require a zero-trust architecture

Geographic Infrastructure: The Ashburn Phenomenon

For the third consecutive year, Ashburn, Virginia, maintains its position as the undisputed capital of API traffic, processing 20% of global requests - more than Europe or Silicon Valley combined.



Most Popular API Traffic Origin, *source: Treblle; sample data 2025*

This isn't a coincidence - it's infrastructure determinism. Ashburn has evolved into what industry insiders call "Data Center Alley" hosting:

- **70% of global internet traffic** routes through Ashburn
- **AWS US-East-1** is the oldest and largest AWS region
- **Microsoft Azure's** primary East Coast presence
- **Google Cloud Platform's** us-east4 region
- **Over 75 data centers** within a 50-mile radius

For enterprise IT leaders, Ashburn's dominance raises strategic questions:

Latency Implications: If your APIs are hosted elsewhere but most traffic routes through Ashburn, you're adding unnecessary latency. Consider edge deployment strategies or Ashburn-based infrastructure.

Redundancy Concerns: Concentration creates vulnerability. The 2025 AWS US-East-1 outage that took down Netflix, Instagram, and Pinterest serves as a warning. Geographic distribution isn't just about performance - it's about resilience.

Cost Optimization: Ashburn's economies of scale often translate to lower costs, but data transfer fees between regions can quickly erode savings. Understanding

your API traffic patterns relative to Ashburn is crucial for cost optimization.

Data Sovereignty: With increasing regulatory requirements around data localization, Ashburn's dominance complicates compliance for global enterprises. The EU's GDPR, China's data localization laws, India's DPDPA, and emerging regulations worldwide require careful API routing strategies.

From Backend Tools to Strategic Business Enablers

The convergence of these trends - the shift to transactional APIs, AI integration, mobile dominance, and infrastructure concentration - fundamentally repositions APIs in the enterprise technology hierarchy.

APIs are no longer middleware. They're not connectors, integrations, or technical utilities. They've become the primary interface through which businesses operate, innovate, and compete. Consider the strategic implications:

Revenue Generation APIs directly generate revenue through:

- API-as-a-Product offerings (Stripe, Twilio, Plaid)
- Marketplace integrations enabling new sales channels
- Partner ecosystems are creating network effects
- Usage-based pricing models for service

Operational Excellence APIs drive operational efficiency through:

- Real-time supply chain orchestration
- Automated workflow management
- Predictive maintenance systems
- Dynamic resource allocation

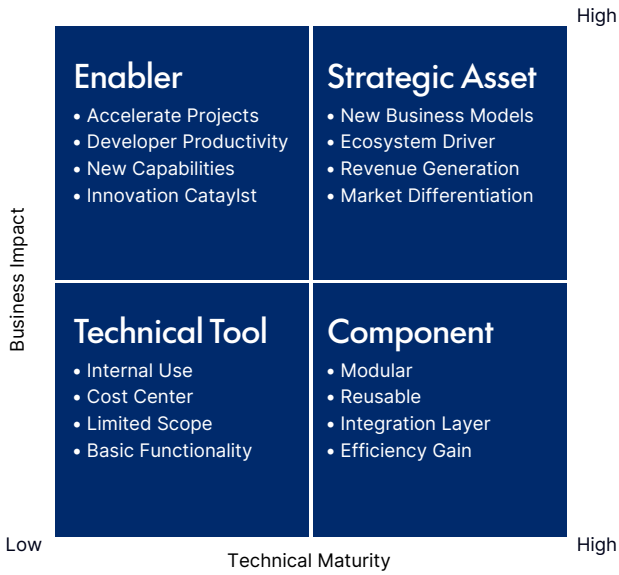
Innovation Acceleration APIs enable rapid innovation through:

- Plug-and-play AI capabilities
- Microservices architecture enabling independent scaling
- Third-party integrations expand functionality
- Rapid prototyping and experimentation

Competitive Differentiation APIs create competitive advantages through:

- Superior developer experience attracting partners
- Faster time-to-market for new features
- Network effects from ecosystem growth
- Data monetization opportunities

API Strategic Value Framework



API Strategic Value Framework, *source: Treblle; sample data 2025*

For Fortune 500 executives, this fundamental shift demands a corresponding shift in thinking. APIs require:

- **Board-level visibility:** API performance metrics should be KPIs in board reports
- **C-suite ownership:** API strategy needs executive sponsorship, not just IT ownership
- **Infrastructure-scale investment:** APIs deserve the same investment as data centers or ERP systems
- **Business-IT collaboration:** API design must reflect business strategy, not just technical requirements

The enterprises that recognize and act on this fundamental shift, treating APIs as strategic business infrastructure rather than technical tools, will be the ones that thrive in the API-powered economy. Those that continue to view APIs as backend utilities will find themselves increasingly unable to compete in a world where every business interaction, every customer experience, and every competitive advantage flows through an API.

> The fact that API traffic does not coincide with workday hours means we've moved out of the simple app era and into an age of automation - possibly signaling the coming autonomous API wave.
>
> 
>
> **Emmanuel Paraskakis**
> CEO @ Level 250

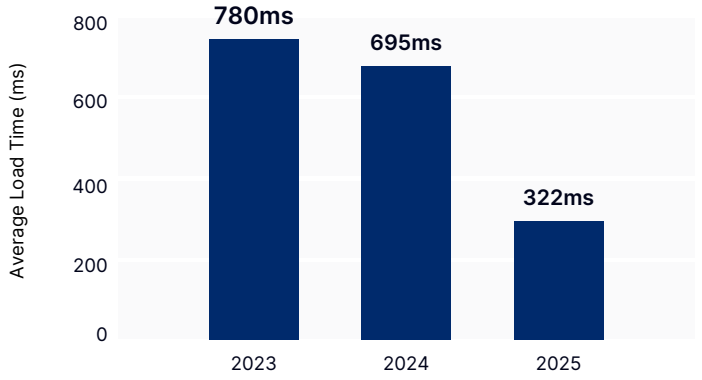
The Performance & Reliability Revolution

In 2025, APIs crossed a critical threshold: they achieved the utility-grade reliability that mission-critical infrastructure demands. This isn't incremental improvement - it's a fundamental transformation in how enterprises engineer, operate, and think about API performance. The data reveals a coordinated industry-wide push toward excellence, driven by the recognition that API failures now directly impact business operations, customer experiences, and competitive positioning.

The 322 Millisecond Barrier: Breaking Through to Excellence

The headline number is striking: average API load time plummeted 54% from 695ms to 322ms in a single year. To put this in perspective, this is the largest year-over-year performance improvement in the history of API monitoring. But the real story isn't just the speed - it's what enabled it and what it means for enterprise operations.

Load Time Evolution (2023-2025)



Load Time Evolution, *source: Treblle; sample data 2025*

This dramatic acceleration reflects multiple converging factors that enterprise IT leaders orchestrated simultaneously:

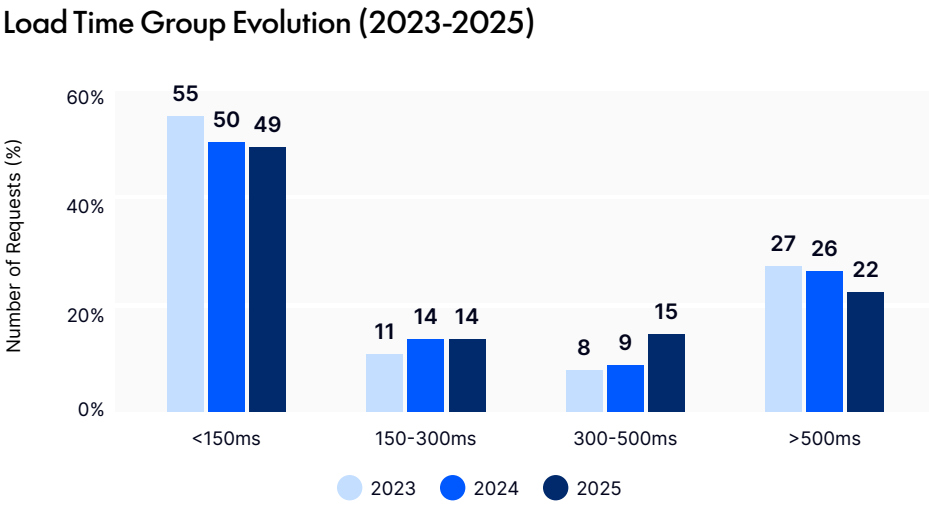
The POST Request Paradox Counterintuitively, the surge in POST requests, typically more resource-intensive than GET requests, coincided with massive performance improvements. This apparent contradiction reveals a sophisticated truth: enterprises have fundamentally reengineered their API architecture.

The AI Performance Imperative With AI APIs growing 42%, performance became existential. Large Language Models have strict timeout requirements; OpenAI's GPT-4, for example, expects responses within specific time windows. An API that takes 700ms to respond simply cannot participate in the AI economy.

Consider the mathematics of AI orchestration:

- AI request to your API: 322ms (average)
- Your API to database: 50ms
- LLM processing: 1-2 seconds
- Total user experience: Under 3 seconds

At 695ms average response time, the same flow would exceed 4 seconds, crossing the threshold where users abandon interactions. The performance improvement isn't just a technical achievement; it's business survival.



Load Time Group Evolution, *source: Treblle; sample data 2025*

The Customer Experience Awakening Enterprise leaders finally connected the dots between API performance and customer satisfaction. Every mobile app freeze, every website spinner, every timeout error traces back to API performance. The data suggests a coordinated response to years of accumulated technical debt.

CIOs and CTOs report that performance improvements were driven by:

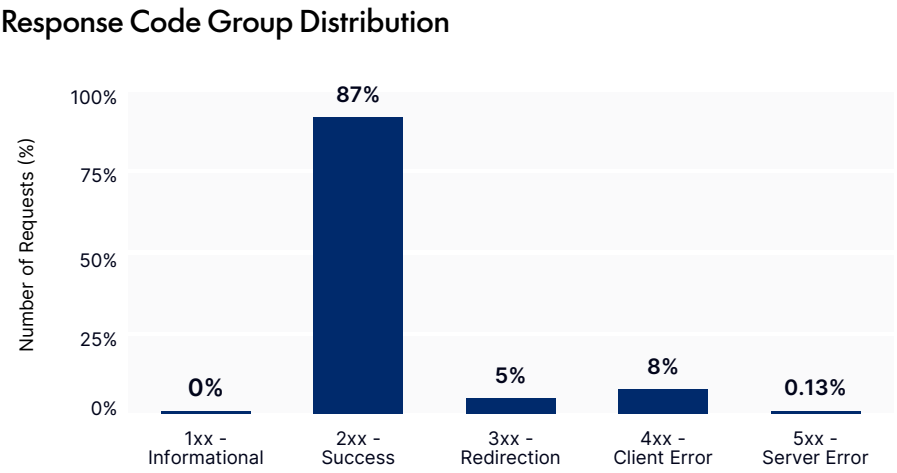
- **Executive mandates:** CEO-level directives to improve digital experience
- **Competitive pressure:** Customers comparing response times across vendors
- **Cost optimization:** Faster APIs require less infrastructure to handle the same load
- **Developer productivity:** Faster APIs accelerate development and testing cycles

The distribution of load times reveals both achievement and opportunity. While 50% of requests complete in under 150ms, the gold standard for user experience, 26% still exceed 500ms. For enterprise architects, this bimodal distribution suggests a two-speed IT environment: modern, optimized APIs coexisting with legacy systems requiring modernization.

> The most frequently encountered scenario requires securing an externally exposed API. However, there are a multitude of use cases that security leaders need to study individually to identify a solution. For example: Internal connectivity of east-west APIs.
>
> 
>
> **Mark O'Neill**
> Chief of Research @ Gartner

The Success Paradox: When 200 OK Isn't Really OK

On the surface, a 97% success rate appears to validate the reliability revolution. Only 3% of API requests result in errors - a remarkable achievement in distributed systems. But deeper analysis reveals a troubling disconnect between reported success and actual reliability.



Response Code Group Distribution, *source: Treblle; sample data 2025*

The data exposes a fundamental design flaw plaguing enterprise APIs: 10% of requests return HTTP 200 OK status codes while actually containing errors. This means the true error rate isn't 3% - it's closer to 13%. For enterprise systems processing millions of requests, this represents millions of hidden failures.

This practice, returning success codes for failure conditions, creates cascading problems:

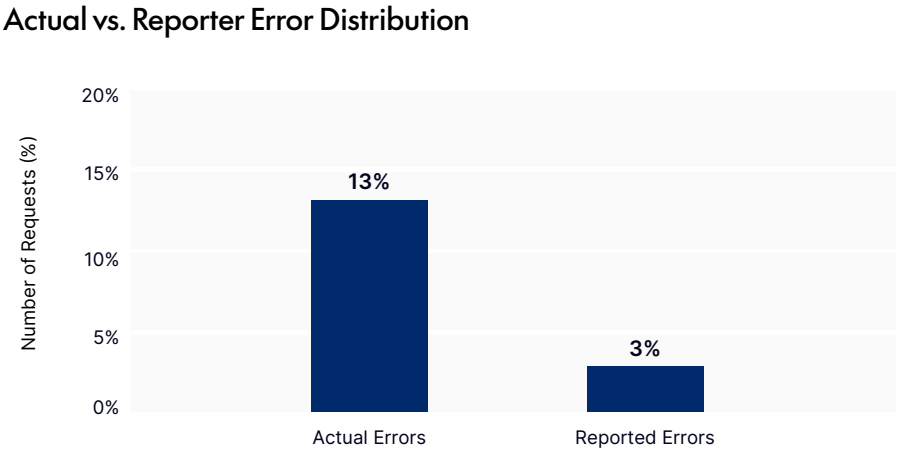
Customer Experience Degradation When an API returns 200 OK with an error message in the response body, client applications can't handle the failure appropriately. Mobile apps don't retry, web applications don't show error messages, and automated systems continue processing bad data. The result is a silent failure - the worst kind of user experience.

Monitoring Blind Spots Enterprise monitoring systems rely on HTTP status codes to track API health. When errors masquerade as successes, they become invisible to:

- Application Performance Monitoring (APM) tools
- Site Reliability Engineering (SRE) dashboards
- Automated alerting systems
- Service Level Agreement (SLA) tracking

AI and Automation Failures This problem becomes critical in the AI era. Large Language Models and automation platforms perfectly understand HTTP status codes and have built-in error handling. But when APIs lie about their status, AI systems can't adapt:

- LLMs continue conversations with corrupt data
- Automation workflows proceed despite failures
- Retry logic never triggers
- Fallback mechanisms remain dormant



Actual vs. Reported Error Distribution, *source: Treblle; sample data 2025*

The persistence of this anti-pattern despite widespread awareness suggests organizational rather than technical challenges. Enterprise architects understand

the problem, but fixing it requires:

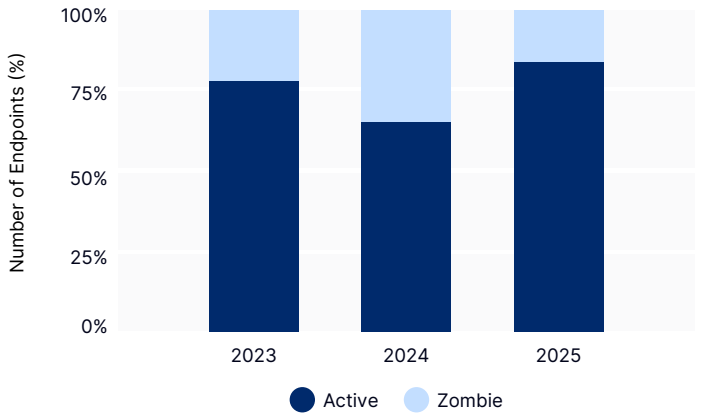
- Updating legacy systems with embedded error patterns
- Retraining development teams on REST principles
- Modifying client applications expecting the current behavior
- Managing backward compatibility during transitions

For CIOs and CTOs, this hidden error epidemic represents both risk and opportunity. Organizations that achieve true 97% success rates, with proper error codes, will deliver superior customer experiences and more reliable AI integrations.

The Zombie Apocalypse Reversed: 51% Fewer Undead Endpoints

The dramatic reduction in zombie endpoints, from 36% to 17%, represents one of 2025's most encouraging trends. This 51% improvement suggests enterprises finally gained control over API sprawl, a problem that has plagued organizations since the microservices revolution began.

Zombie vs. Active Endpoints (2023-2025)



Zombie Endpoint Trends, *source: Treblle; sample data 2025*

The turnaround is particularly impressive considering zombie endpoints actually increased from 2023 to 2024, suggesting 2025 marked a watershed moment in API lifecycle management. What changed?

The Gartner Effect Industry analysts, particularly Gartner, elevated API sprawl and zombie endpoints to boardroom conversations. Their research showed that zombie endpoints were responsible for 38% of API security breaches, creating executive urgency. When Gartner declares something a "critical risk," Fortune 500 executives listen.

Regulatory Pressure New regulations requiring API inventory management forced organizations to catalog their endpoints. The EU's Digital Operational Resilience Act (DORA) and similar regulations worldwide mandate that financial institutions maintain accurate API inventories. You can't manage what you can't measure, and regulation forces measurement.

Tool Maturation API observability platforms now automatically detect zombie endpoints through:

- Traffic analysis identifying unused endpoints
- Code coverage tools mapping endpoint utilization
- Dependency analysis reveals orphaned services
- Automated sunset recommendations

The True Cost Recognition Enterprises finally calculated the total cost of zombie endpoints:

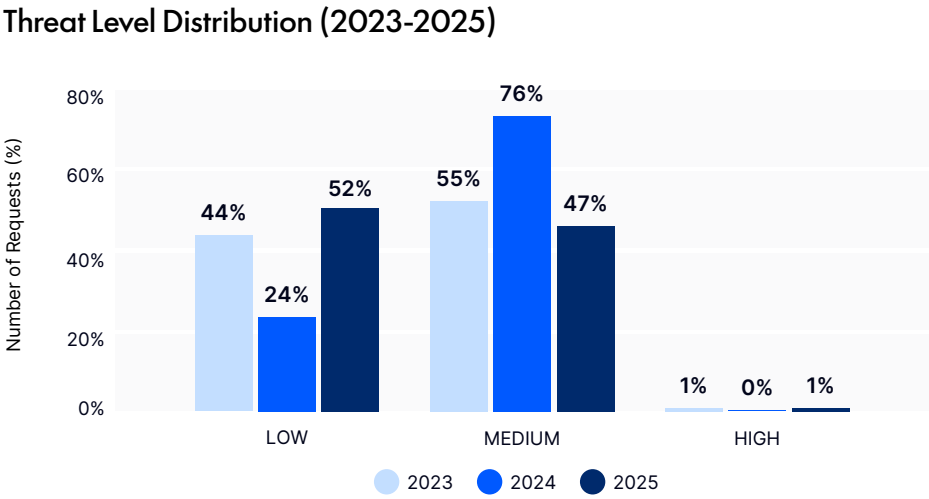
- **Security exposure:** Each endpoint is a potential attack vector
- **Maintenance burden:** Developer time spent maintaining unused code
- **Infrastructure waste:** Compute resources serving no purpose
- **Compliance risk:** Undocumented endpoints failing audit requirements

For enterprise architects, the zombie endpoint reduction provides a template for technical debt management. The approach that worked here: analyst pressure, regulatory requirements, tooling improvements, and cost analysis can be applied to other systemic problems.

However, 17% zombie endpoints still represent a significant risk. In a typical Fortune 500 company with 10,000 endpoints, this means 1,730 unmaintained, unmonitored potential vulnerabilities. The journey from good to great requires eliminating these remaining zombies.

The Threat Level Transformation: From Red to Green

The security posture of APIs underwent dramatic improvement, with low-threat traffic increasing from 24% to 52%, while medium-threat traffic decreased from 76% to 48%. This shift represents billions of API requests moving from vulnerable to secure.



Threat Level Distribution, *source: Treblle; sample data 2025*

This transformation didn't happen by accident. It reflects coordinated enterprise action on multiple fronts:

OWASP API Security Top 10 Adoption The data shows enterprises systematically addressing the OWASP API Security Top 10 vulnerabilities:

- Broken Object Level Authorization (addressed through proper access controls)

- Broken Authentication (improved from 48% to 53% authenticated requests)
- Excessive Data Exposure (reduced through response filtering)
- Rate Limiting (though still only 15% implementation)

Security-First Architecture The improvement suggests enterprises are building security into APIs from the design phase:

- Threat modeling during API design
- Security testing in CI/CD pipelines
- Automated vulnerability scanning
- Runtime protection through API gateways

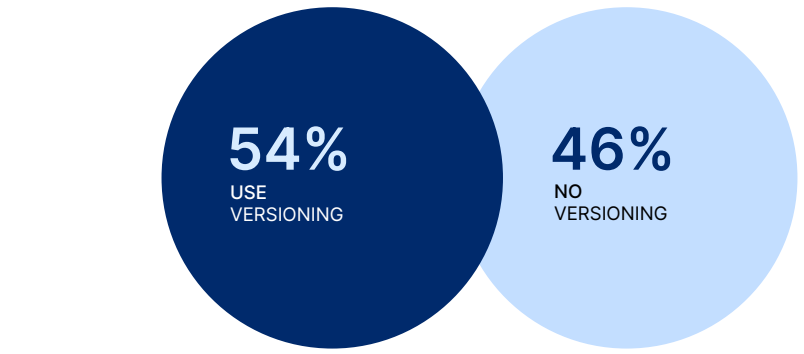
The Analyst Influence Gartner, Forrester, and IDC's focus on API security created board-level awareness. When Gartner predicts that "by 2025, API attacks will become the most frequent attack vector," executives authorize security investments.

However, the near-tripling of high-threat traffic (0.003% to 0.9%) raises concerns. While still minimal, this represents millions of high-risk requests in enterprise environments.
These could indicate:

- Sophisticated attack attempts
- Legacy system vulnerabilities
- Supply chain API compromises
- Insider threat activities

The Versioning Gap: Why 46% of APIs Live Dangerously

Despite all improvements, a glaring weakness remains: only 54% of APIs implement versioning. This means nearly half of all enterprise APIs cannot evolve without potentially breaking existing integrations.



API Versioning Usage, *source: Treblle; sample data 2025*

The absence of versioning creates a cascade of problems for enterprise agility:

Innovation Paralysis Without versioning, any API change risks breaking production systems. This creates:

- Fear of improvement (teams avoid optimizations)
- Accumulation of technical debt (problems persist rather than being fixed)
- Competitive disadvantage (inability to rapidly deploy new features)

The Backward Compatibility Trap APIs without versioning must maintain backward compatibility forever, leading to:

- Bloated responses including deprecated fields
- Confusing documentation mixing old and new patterns
- Performance degradation from legacy support
- Security vulnerabilities in outdated implementations

Integration Nightmares For Fortune 500 companies with thousands of API consumers, a lack of versioning means:

- Coordinating simultaneous updates across all consumers
- Extended maintenance windows for changes
- Higher risk of production incidents
- Increased testing complexity

The preference for URL-based versioning (v1, v2 in the path) among those who do version reveals pragmatism over purism. While REST purists prefer header-based versioning, URL versioning provides:

- Clear visibility in logs and monitoring
- Simpler client implementation
- Better cache management
- Easier documentation

For CTOs and enterprise architects, the versioning gap represents a critical vulnerability in digital transformation strategies. Modern practices like continuous deployment, A/B testing, and rapid experimentation all require robust versioning strategies.

The Reliability Revolution: From Startup Agility to Enterprise Resilience

The convergence of these improvements, 54% faster performance, 51% fewer zombie endpoints, improved security posture, and growing versioning adoption, marks a fundamental shift in API maturity. APIs have evolved from startup-style "move fast and break things" to enterprise-grade "move fast and don't break things."

This evolution reflects APIs' new role as critical infrastructure. When APIs were middleware connecting internal systems, occasional failures were inconvenient. When APIs power customer experiences, enable AI operations, and drive revenue, failures become existential threats.

The data reveals that leading enterprises have internalized this reality and responded accordingly:

Investment in Reliability Engineering

- Dedicated API reliability teams (following Google's SRE model)
- API-specific observability platforms
- Chaos engineering for API resilience
- Formal API lifecycle management

Operational Excellence:

- 99.99% uptime SLAs are becoming standard
- Sub-second response time requirements
- Automated performance testing
- Continuous optimization cycles

Governance Maturity:

- API Centers of Excellence
- Standardized design patterns
- Automated compliance checking
- Version lifecycle management

For Fortune 500 leadership, the performance and reliability revolution validates a crucial insight: treating APIs as critical infrastructure requiring commensurate investment pays dividends. The organizations achieving 322ms response times and 97% success rates aren't lucky - they're disciplined.

The remaining gaps: hidden errors, unversioned APIs, and residual zombie endpoints - represent the difference between good and great. As APIs become even more central to business operations, closing these gaps transitions from technical optimization to a business imperative.

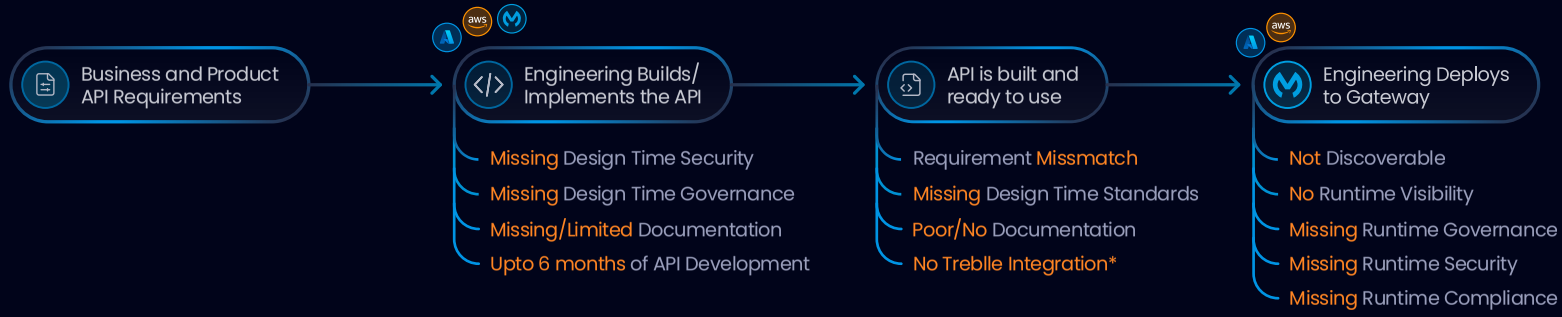
The Governance & Security Paradox

If the "Fundamental Shift" described in section one represents the acceleration of the API economy, and the Performance Revolution represents the engine upgrades we have made to our infrastructure, then this section uncovers a startling reality: we are driving faster than ever, but many of us have forgotten to put on our seatbelts.

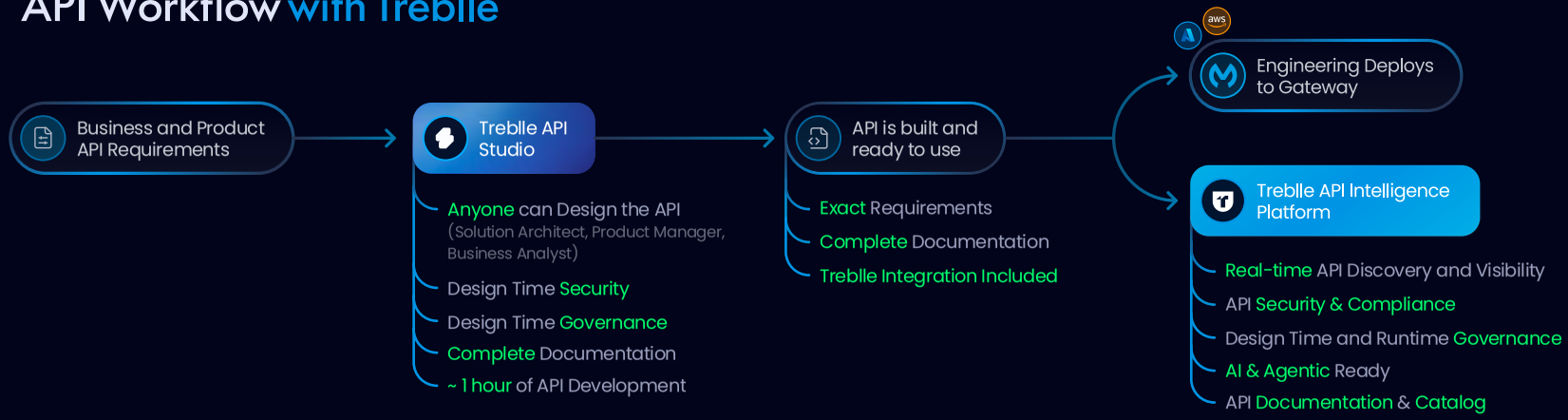
While organizations have invested heavily in speed, latency reduction, and AI integration, the foundational pillars of governance and security have not kept pace. In fact, in several key metrics, they show signs of regression.
For the executive reader, this is not merely a technical issue; it is a strategic liability. The data paints a picture of an ecosystem where "Implicit Trust", the assumption that internal traffic is safe, still dominates architectural decision-making. As APIs evolve from simple data pipes to the central nervous system of the AI-driven enterprise, this lack of rigorous governance creates a blast radius that extends far beyond the IT department.

> Digital businesses are getting clearer about maturing their API portfolio — by reducing zombie endpoints, by improving performance, and by extending functionality. But there is still work to do to make sure APIs drive new revenue and bring in new customers — by making APIs secure, by mapping what APIs they have and who uses them, and by reducing the complexity of monolithic APIs. Trebllle's data shows where you can benchmark yourself against the industry and top performers.
>
> 
>
> **Mark Boyd**
> Director @ Platformable

API Workflow without Treblle



API Workflow with Treblle



The Unencrypted Elephant in the Room

In an era of Zero Trust architectures and regulatory mandates (GDPR, CCPA, PCI-DSS), one would assume that unencrypted HTTP traffic would be a relic of the past. The 2025 data suggests otherwise: 42% of all API traffic in 2025 remains unencrypted (HTTP). This represents an improvement from 2024 (where unencrypted traffic spiked to 55%), but it remains significantly worse than 2023, where nearly 74% of traffic was secure (HTTPS).

Why, in 2025, is nearly half of all API traffic traversing networks in plain text? The data points to a prevalent architectural habit: **The Perimeter Defense Fallacy.**

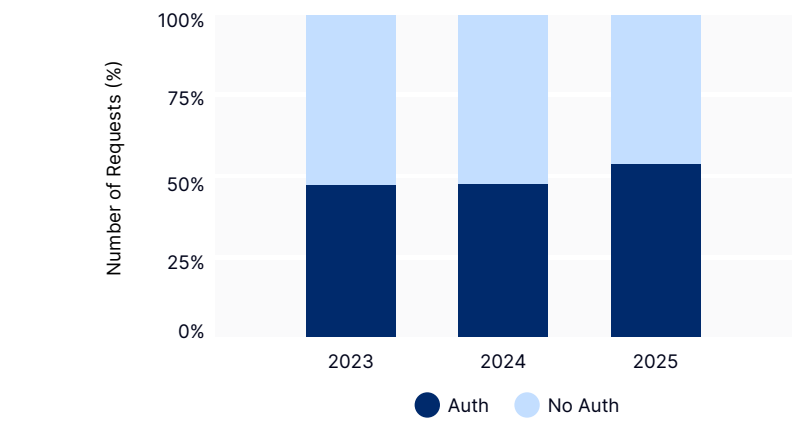
A granular look at the source of this traffic suggests that the bulk of unencrypted requests are not public-facing mobile apps or partner integrations, which are almost universally encrypted by app store requirements, but rather **internal microservices** and **legacy backend systems.**

Many organizations continue to operate under the assumption that once a request passes the API Gateway or the corporate firewall, it is "safe." Consequently, service-to-service (East-West) traffic often defaults to HTTP to save on the marginal computational overhead of TLS handshakes or, more commonly, to avoid the operational complexity of managing internal certificates.

This 42% represents a massive attack surface for **lateral movement.** If a threat actor breaches the perimeter, a scenario that is statistically inevitable, unencrypted internal APIs allow them to sniff credentials, inject malicious payloads, and map the network without decryption. In an AI-driven world where APIs are autonomously calling other APIs, this lack of encryption breaks the chain of custody for data privacy.

The Identity Crisis

If transport security is the walls of the castle, authentication is the guard at the gate. Unfortunately, 2025 data indicates that for nearly half of the API ecosystem, the gate is wide open. **47% of APIs process requests without any form of Authentication.** While this is a slight improvement over 2024 (53% No Auth), it indicates that essentially **1 in 2 API calls** is anonymous.



Auth vs. No Auth, *source: Treblle; sample data 2025*

It is important to nuance this statistic. "No Auth" does not always mean "Public Data."

1. **Public/Open Data:** A portion of this traffic is legitimately public (e.g., weather data, product catalogs, public AI models).
2. **The Hidden Risk:** The more concerning segment involves APIs that rely on Network-Level Trust (IP whitelisting) or Obscurity (hidden endpoints) rather than cryptographic identity (OAuth, JWT, API Keys).

The 2025 dataset shows a clear correlation: frameworks that enforce opinionated security structures (like NestJS) see much higher authentication adoption than unopinionated, flexible frameworks (like ExpressJS or plain PHP).

Another interesting angle on the auth story is the HTTP response codes. Only 6% of all recorded errors were 401 Unauthorized; this relatively low number in the context of 47% unauthenticated traffic suggests a lack of enforcement. Systems aren't rejecting unauthenticated users because they aren't checking for them.

Anonymous APIs are the preferred training ground for malicious AI agents. Without authentication, rate limiting becomes difficult (tracking by IP is insufficient in the age of botnets), and attribution becomes impossible. If you cannot identify who is calling your API, you cannot govern it.

Operational Governance: The Zombie Apocalypse

Security is often a byproduct of good governance. You cannot secure what you do not know exists. The 2025 data reveals a significant "Inventory Problem" within the enterprise, with **17% of all tracked endpoints being "Zombie APIs."**

A "Zombie API" is an endpoint that is technically live and accessible but has seen no legitimate active development or traffic patterns consistent with production use for an extended period, yet it hasn't been decommissioned.

Compounding the Zombie issue is the lack of formal versioning.

- **46% of APIs lack any versioning strategy** (e.g., /v1/, headers)
- **54% use versioning**

The Paradox: We are building new APIs at a record pace (38% more endpoints on average than in 2024), but we are failing to retire the old ones.

Zombie APIs are the **Path of Least Resistance** for attackers. These endpoints often rely on older, vulnerable libraries (Log4j remnants), lack modern monitoring, and are frequently excluded from WAF (Web Application Firewall) rules because "nobody uses them anymore." In 2025, retaining 160,000+ zombie endpoints is the digital equivalent of leaving the keys in the door of your old office building.

The Threat Landscape

For the first time, this report can quantify the "Temperature" of API security threats based on request analysis.

It is tempting for a leader to look at "1% High Threat" and dismiss it as negligible. However, scale matters. In a dataset of **1 billion requests,** a 1% rate implies **9 million highly malicious requests.**

These "High" threat requests are not merely malformed data; they represent:

- SQL Injection (SQLi) attempts.
- Cross-Site Scripting (XSS) payloads.
- Remote Code Execution (RCE) probes.

Interestingly, we observed a shift from 2024, where "Medium" threats dominated (76%), to 2025, where "Low" threats (52%) took the majority share. This suggests two possibilities:

1. **Better Filtering:** Gateways are blocking obvious threats earlier, so they don't register as deep API threats.
2. **Stealthier Attacks:** Attackers are moving away from "noisy" medium-level scans to "low and slow" traffic that mimics legitimate user behavior to evade AI detection.

The API Scorecard

To synthesize our findings on an API level, we built a real-time scoring algorithm that rates APIs on a scale of 0-100 or A-F. It does that across 4 different categories: AI Readiness, Security, Design, and Performance. Our data shows that the **Global API Scorecard in 2025 was 58/100.**

While this is an increase from 2023 (50/100) and 2024 (57/100), a score of 58 - a failing grade in any academic setting - reflects the paradox perfectly. We have world-class performance (Grade A) but elementary-school security hygiene (Grade F).

> We've been doing API first wrong and now the growth of AI is revealing what many of us have known for some time: APIs are for more than data sharing, they enable desired outcomes. Expect to see more transactional API operations rather than pure data sharing APIs as we engage in deeper partner and customer opportunities.
>
> 
>
> **James Higginbotham**
> API Strategist @ LaunchAny

Governance decisions made at the start of a project (Technology Selection) dictate long-term security scores. The data shows a massive disparity based on the SDK/Framework used:

Top API Governance Scores per SDK

Nest JS	83	The Gold Standard. Opinionated structure forces security defaults.
Spring Boot	73	Strong enterprise defaults (Java ecosystem).
Laravel	65	Good defaults, widely used in rapid dev.
.NET Core	63	Solid, but configuration dependent.
NodeJS	54	The Risk Zone. Highly flexible, requires manual security setup.
Django	49	Surprisingly low, likely due to misconfigured debug modes/headers.

SDK/Frameworks by API Governance Score, *source: Treblle; sample data 2025*

The choice of technology is a governance choice. Using flexible, unopinionated frameworks like raw NodeJS or Django requires a significantly higher investment in security tooling and training to achieve the same baseline as opinionated frameworks like NestJS.

The Governance & Security Paradox presents a clear warning: **We are building technical debt faster than we are paying it down.** As we move into 2026, the strategy must shift from "Speed at all costs" to "Secure by Design."

The Technology & Platform Ecosystem

If governance is the "conscience" of the API ecosystem, the technology stack is its "muscle." In 2025, that muscle is growing larger, faster, and more opinionated.

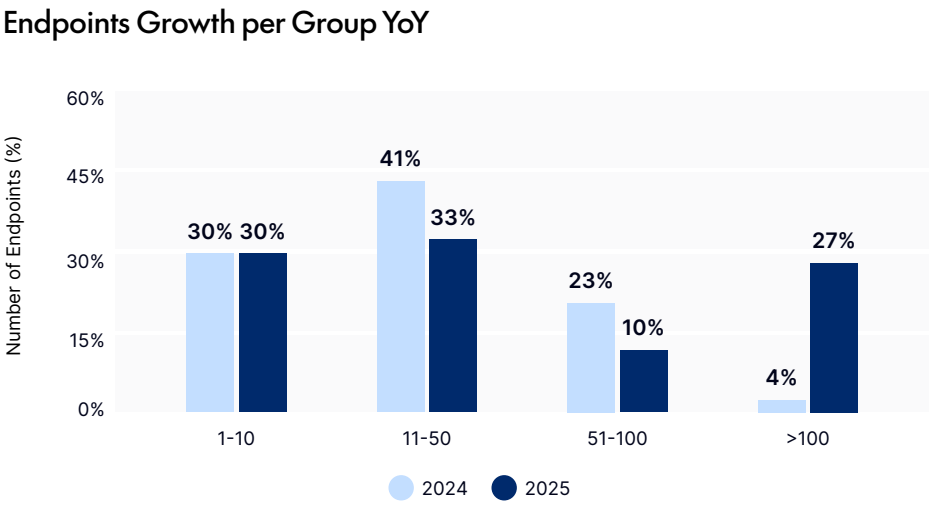
Our analysis of the technology powering over 1 billion requests reveals a definitive split in the market. On one side, we see the rise of "Enterprise Guardrails", frameworks that enforce security and structure by default. On the other hand, we see the chaotic growth of "Vibe Coding" - rapid, AI-generated implementations in flexible languages that often sacrifice long-term stability for immediate velocity.

For the CIO, the data presents a clear choice: The tools your teams select today will mathematically determine your security posture tomorrow.

The Complexity Explosion

Perhaps the most telling statistic regarding the complexity of modern software is the explosion in API surface area.

In 2024, APIs with **100 or more endpoints** accounted for only **4%** of our sample. In 2025, that number has skyrocketed to **38%.**



Endpoints Growth per Group YoY, *source: Treblle; sample data 2025*

This data suggests a reversal of the "Microservices" trend. We are seeing the consolidation of services into larger, more complex "Macro-services" or "Modular Monoliths."

We hypothesize that this explosion in endpoint count is driven by **AI.** For an AI agent to be useful, it needs granular access to data and actions. Developers are exposing more discrete functions as endpoints to allow LLMs to "do" more things.

This rapid expansion correlates directly with the **17% Zombie Endpoint** rate mentioned in the Governance section. As teams rapidly add endpoints to feed AI models or support new features, they are failing to deprecate the old ones, leading to bloated, difficult-to-maintain interfaces.

Maintenance costs are about to skyrocket. Documenting 12 endpoints is a chore; documenting 56 is a full-time job. If we are moving toward APIs with hundreds of endpoints, "Documentation-as-Code" and automated spec generation (OpenAPI) become the only way to survive the complexity.

The Corporate Heartbeat

The internet may never sleep, but the API economy definitely follows a schedule. Our analysis of traffic distribution by day of the week reveals that the digital world is still tethered to the physical work week.

The Mid-Week Peak - we observe a distinct bell curve centered on the middle of the working week:

- **Wednesday:** 16.3% of total traffic (Peak).
- **Thursday:** 16.2% of total traffic.
- **Tuesday:** 15.7% of total traffic.

The Weekend Drop-off - despite the rise of consumer apps and 24/7 global connectivity, traffic drops significantly on weekends (Saturday ~13%, Sunday ~12%).

This pattern suggests that the bulk of API consumption is **B2B (Business-to-Business)** or **workflow**-related. It is systems talking to systems during business hours. This challenges the assumption that "Global/Remote work"

has flattened time zones. The "Heartbeat" of the API economy is still synchronized with the Western business week.

Leaders of the API Economy

Our analysis of traffic patterns and security scores strongly correlates with the top three industries defining the API economy in 2025: **Financial Services, Travel,** and **Telecom.**

1. Financial Services: The Fortress

Banking, Insurance, and Fintech remain the undisputed heavyweights of the API world.

- **The Driver:** Regulatory pressure (Open Banking, PSD3) and the shift to Real-Time Payments have forced this sector to mature faster than any other.
- **The Data Connection:** The surge in traffic from major financial cloud hubs and the dominance of high-scoring frameworks like **Spring Boot** and **NestJS** directly reflect this sector's "Security-First" mandate. They are the primary drivers behind the 53% drop in latency, as algorithmic trading and fraud detection require millisecond precision.

2. Travel: The Aggregator

The Travel industry was the original API economy, and it continues to lead in complexity.

- **The Driver:** The fragmentation of airlines, hotels, and experiences requires massive aggregation layers to present a unified booking experience to the user.
- **The Data Connection:** The explosion of **"Mega-APIs" (100+ endpoints)** is a hallmark of this sector. Travel APIs are not just moving data; they are orchestrating complex, multi-step transactions (search, book, pay, confirm) across disparate legacy systems, necessitating the large endpoint counts we observed this year.

3. Telecom: The Awakening Giant

Telecom is shifting from providing just "connectivity" to providing "programmable networks."

- **The Driver:** The rollout of 5G and the CAMARA alliance initiatives are turning network capabilities (like Quality on Demand or SIM Swap checking) into consumable APIs.
- **The Data Connection:** The massive volume of unauthenticated or "internal" traffic often mirrors the telco architectural style, where network-level trust (IP whitelisting) has historically been preferred over token-based auth.

> AI is triggering a second wave of 'Taking API Landscaping Seriously.' The first came when organizations worked to scale their API practices, attempting to move beyond bespoke integrations. The second comes now as teams move from successful but isolated AI experiments to the challenge of scaling them so they can build and roll out useful systems repeatedly and improve over time.
>
> 
> **Erik Wilde**
> Head of Enterprise Strategy @ Jentic

The 2026 Imperative: From Infrastructure to Strategic Advantage

If 2025 taught us one thing, it is that the era of the "Technical API" is over. We have entered the era of the "Business API."

For the last decade, APIs were viewed primarily as IT infrastructure, plumbing to be maintained, cost centers to be optimized, and integration points to be managed. The data from this year's report signals a fundamental inversion of that model. With POST requests (transactional actions) overtaking GET requests (passive reading), and with AI agents emerging as a new primary consumer, **APIs have become the most valuable strategic assets in the modern enterprise.**

They are no longer just how you *connect* your business; they are how you do business.

As we look toward 2026, the mandate for enterprise leadership is to stop managing APIs as code and start managing them as products. This requires a shift in focus from "Uptime" to "Revenue," from "Monitoring" to "Observability," and from "Implicit Trust" to "Explicit Governance."

The Visibility Mandate: Why Run-Time Data Wins

The most glaring risk, the Governance Gap, exists because of a lack of visibility. **You cannot secure what you cannot see.**
For years, organizations have relied on **Surveys** ("Do we use SSL?") and **Static Analysis** ("Does the code look secure?") to judge their posture. The 2025 data proves these methods are insufficient.

- Developers say they use Versioning, but **46%** of traffic is unversioned.
- Architects design for HTTPS, but **42%** of traffic flows over HTTP.

The Strategic Shift: In 2026, **Run-Time Observability** is non-negotiable. Leaders must demand real-time dashboards that show what is actually happening on the wire, not what is supposed to happen in the documentation. This is the only way to bridge the gap between "Design" and "Reality."

The Investment Roadmap: A 2026 Execution Plan

To close the gaps identified in this report and transform your API landscape, we recommend a focused, quarterly execution strategy for 2026.

Q1: Visibility & Discovery

- **Objective:** Turn on the lights.
- **Action:** Implement Run-Time Security and Observability tools. You cannot fix what you do not know exists.
- **Key Result:** Achieve 100% visibility into your API inventory, uncovering all shadow APIs and internal microservices that are currently flying under the radar.

Q2: Close the Security Gap

- **Objective:** Remediation of critical risks.
- **Action:** Now that visibility is established, use that data to audit and fix the foundation. Gate all APIs that currently have "No Auth" (47%), enforce strict HTTPS encryption for internal traffic, and decommission identified zombie endpoints.
- **Key Result:** A secured perimeter where no known vulnerabilities exist in the transport or identity layers.

Q3: Improve Governance & Design

- **Objective:** Prevention through standardization.
- **Action:** Shift from "Cleanup" to "Prevention." Implement design standards (Style Guides) and enforce run-time governance policies that automatically block non-compliant traffic before it hits production.
- **Key Result:** A standardized ecosystem where new APIs are "Secure by Design," preventing technical debt from re-accumulating.

Q4: Monetization & AI Expansion

- **Objective:** The Offensive Shift.
- **Action:** Start exposing your certified, secure APIs to AI Agents via Model Context Protocols (MCP) and identify high-value internal APIs to monetize with external partners.
- **Key Result:** Transformation of the API platform from a cost center into a revenue generator.

API Intelligence Platform

Discover, Understand, and Manage APIs

API Discovery

- Azure APIM
- AWS
- Mulesoft
- .NET AWS
- Spring
- GitHub
- +10 more

API Traffic

- Azure APIM
- Spring
- AWS
- Mulesoft
- +30 more

API Security

- OWASP Top 10
- Automated API Design Time and Runtime Checks
- Alerting

API Visibility

- Realtime API Data across Infrastructure
- API Catalog
- Business Group level Dashboards

API Governance

- Design, Performance Security, AI Readiness
- Specific Custom Rules
- Automated Governance at Design Time and Runtime

Agentic AI

- AI Agent Detection
- Automated API Refactoring
- MCP Server

API Documentation

- Autogenerated OpenAPI Specs
- Internal and External Real-time Dev portals
- AI Assistant for Integration

API Compliance

- GDPR, HIPPA, CCPA PCI-DSS
- Compliance Dashboards
- Audit Logs on every request

API Analytics

- 100+ API Specific Metadata on Every Request
- Customizable API Dashboards
- Customer Dashboards

Conclusion

The Anatomy of an API in 2025 is a picture of immense power and significant fragility. We have built a Ferrari engine and put it in a go-kart chassis. The data is more than clear: The technology works. The performance is there. The AI is ready. The only thing holding us back is **Governance.**

The winners of 2026 will not be the ones who write the fastest code; they will be the ones who have the **visibility** to see where they are going and the **discipline** to steer the ship.

To repeat something we mentioned before: **Today, every business is a digital business, and every digital business runs on APIs.** Whether you are a bank, a retailer, or a startup, your APIs are your products.

Methodology

The insights in this report are not derived from surveys, interviews, or theoretical models. They are generated from **run-time data,** the actual digital footprint of the global API economy.

This data is sourced directly from Treblle, an API Intelligence platform that powers over 15,000 APIs worldwide. Each month, our infrastructure processes approximately 3 billion API requests and 10 terabytes of data.

For the purpose of this 2025 edition, we extracted a statistically significant sample of **1 billion randomly selected requests.**

The data is ingested via Treblle's lightweight SDKs, which are integrated directly into our customers' APIs. These integrations span over **30 different technologies** including the most popular programming languages and API gateways.

Once ingested, every single request is enriched in real-time. Our engine generates over **50 unique data points** from a single transaction, calculating load times, identifying geolocation data, analyzing headers, and grading architectural quality.

We adhere to the strictest standards of data privacy and ethics.

- **Anonymization:** All data used in this report was completely anonymized. No specific API, customer, endpoint, or user can be identified from the dataset.
- **Data Masking:** Treblle's SDKs feature a built-in masking engine that strips PII and sensitive values before data ever leaves the customer's infrastructure.

Exclusions: This report relies exclusively on data from our Public Cloud Deployment (SaaS) and not our Private Cloud and On-prem Deployments