

Modernizing Systems Observability with AI and LLMs



SCaLE 22x

March 2025



DATADOG



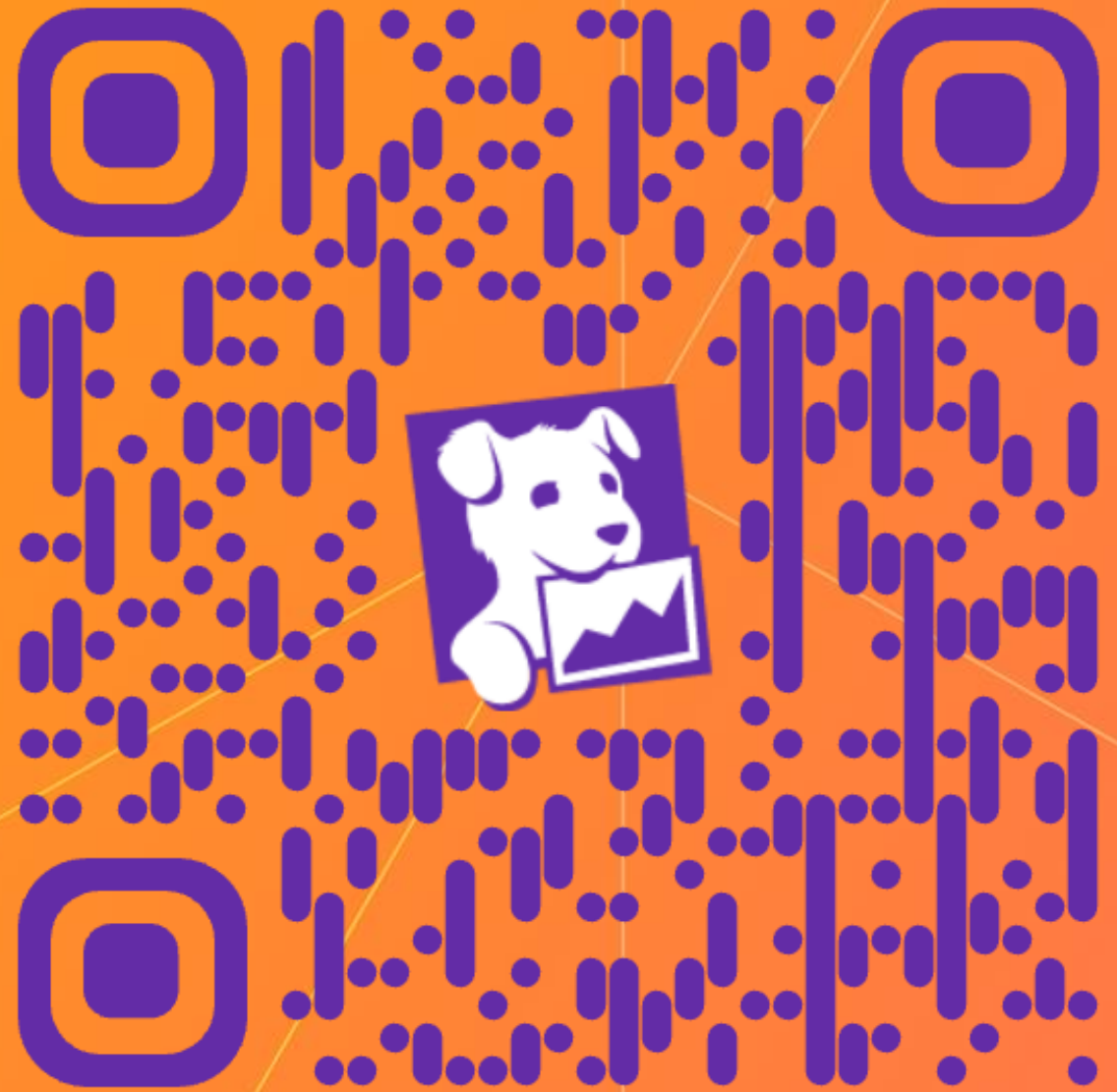
Jason Hand

Senior Developer Advocate : **Infra & Platform** – **AI** : Datadog

Resources

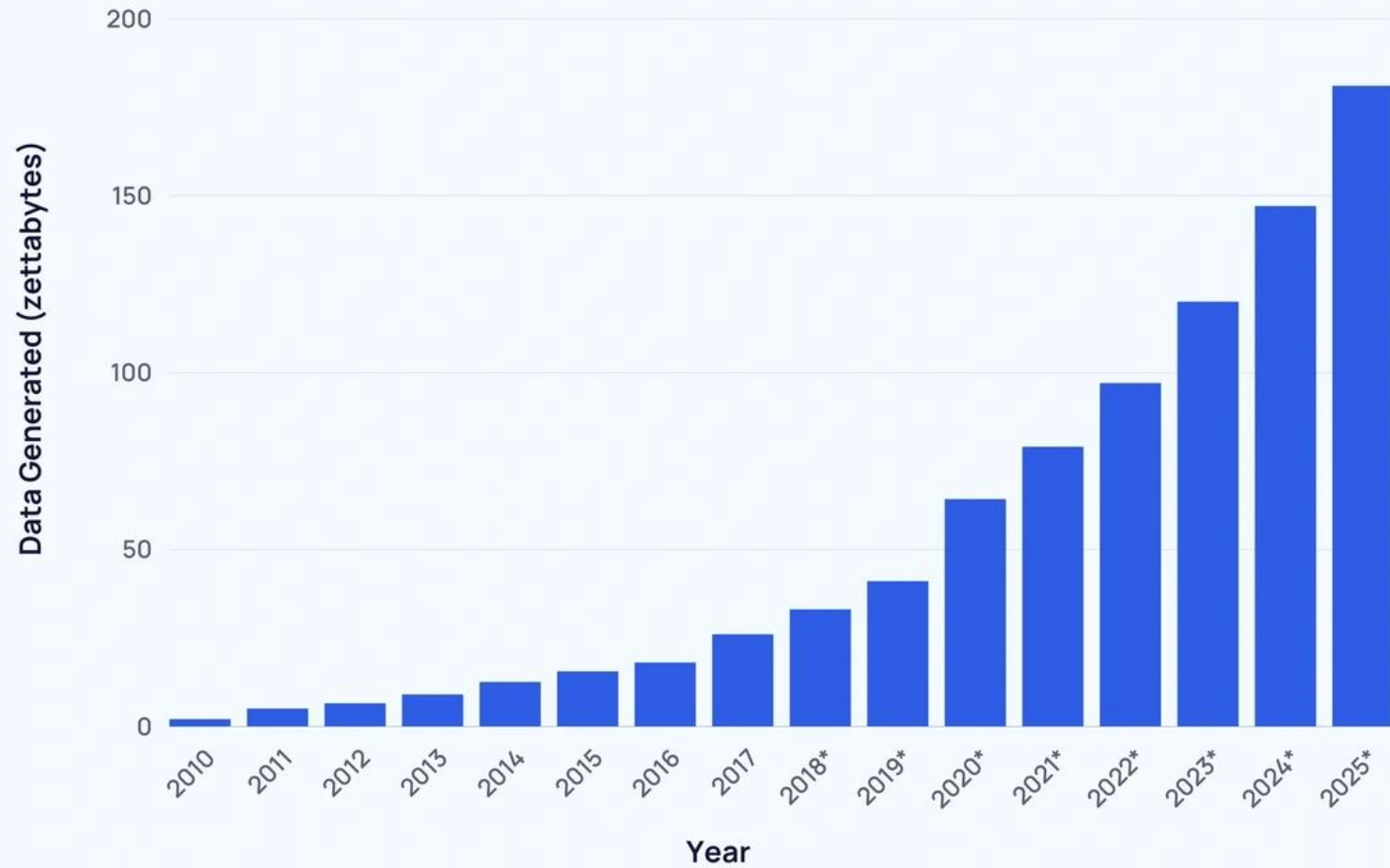
dtdg.co/SCaLE22x

NOTE: Authenticate GitHub first

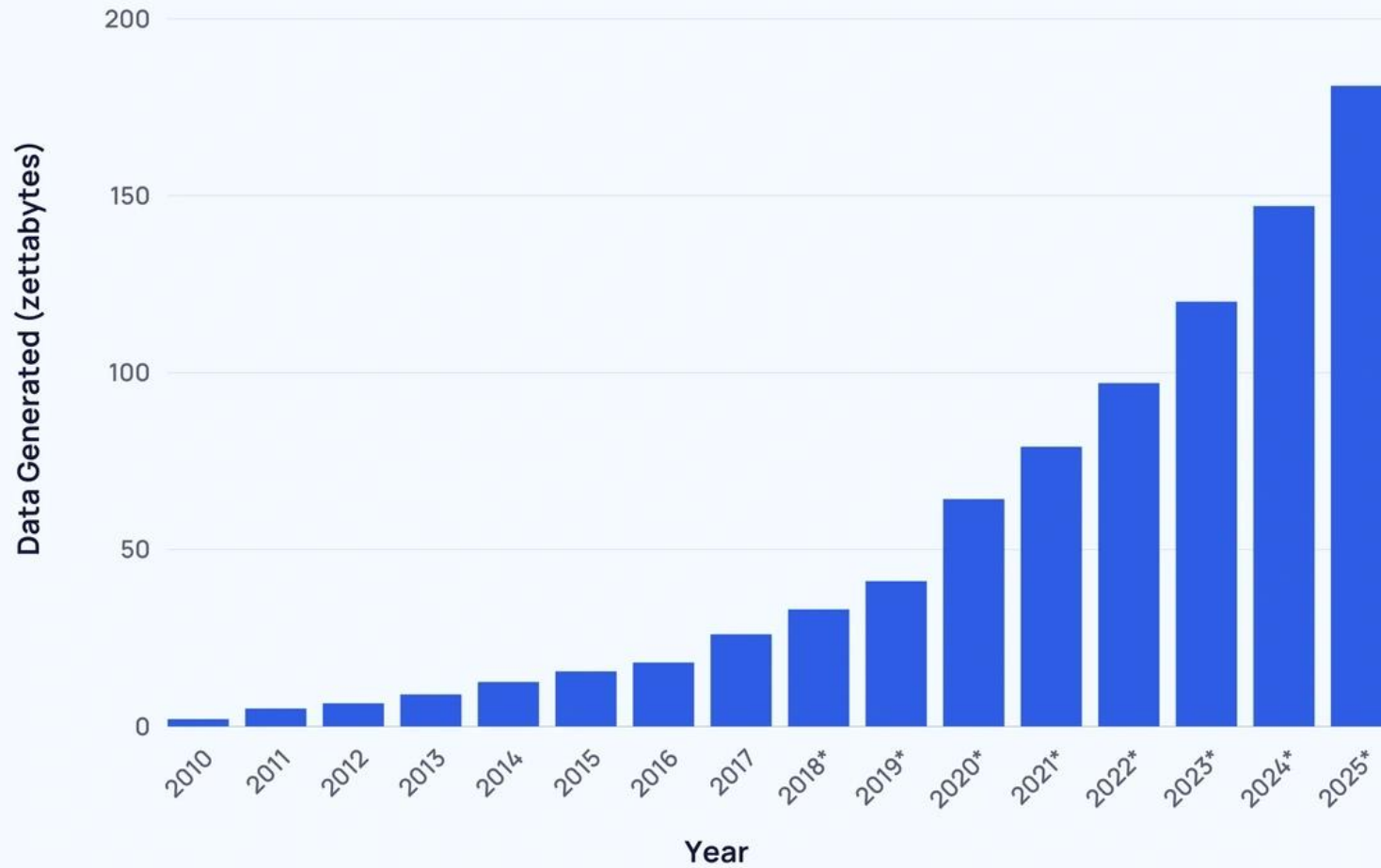


**How much data is
generated annually,
globally?**

Global Data Generated Annually



Global Data Generated Annually



90%

Produced in the past
two years

Data Size Breakdown

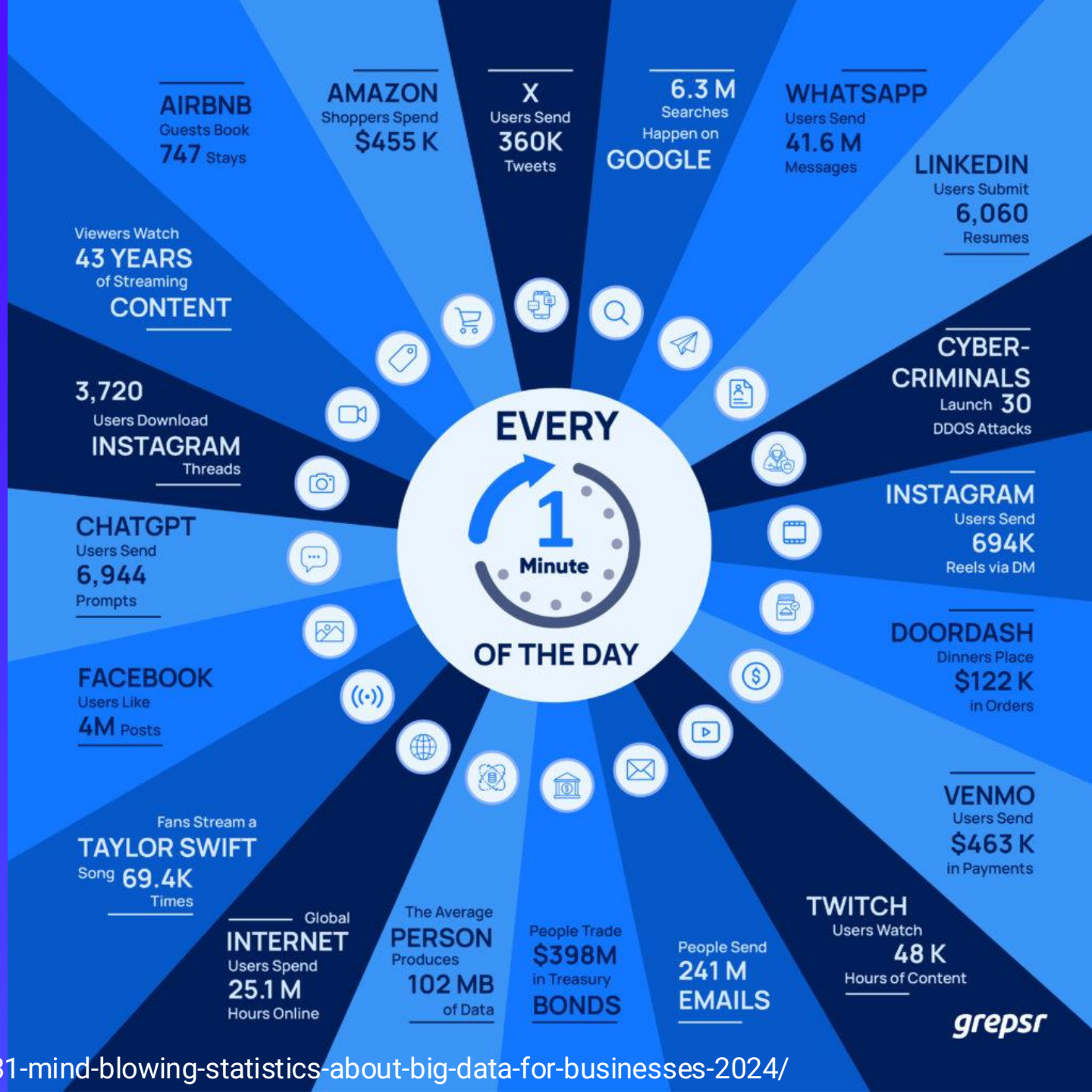
Petabyte = 1,000,000 gigabytes

Exabyte = 1,000 petabytes

Zettabyte = 1,000 exabytes

Yottabyte = 1,000 zettabytes

“By **2025**, the world will generate an
astounding **180 zettabytes** of data, with
approximately **30%** of this being **real-time
data** created by connected users having
digital interactions every **18 seconds**”





That's not my data

**Application
generated
data**



Data

Generated from servers



Data

Generated from the cloud



Data

Generated from your users



Data

Generated from your disruptions

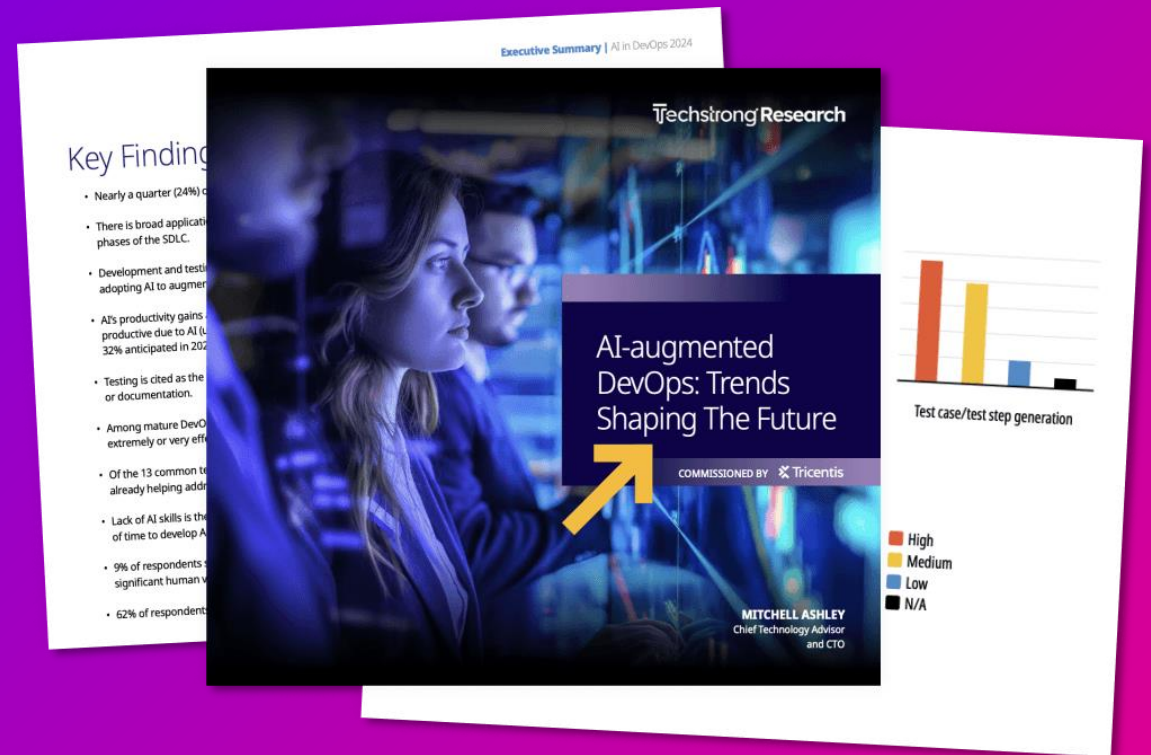
63%



will be making additional investments in observability over the next two years, with 21% describing those investments as significant..

46%

plan to adopt AI tools to augment DevOps teams in the next 12 months.



Adoption of AI tools

**Development
& Testing**

**Operations
& Monitoring**

“Expected” Benefits



Increased
productivity



Reduction in skills
gap



Improvement in
software quality



Lower operation
costs

Development



- **Improvements in writing code**
- **Gains in debugging and defect prevention**
- **Automated code reviews and optimization reduce errors**

AI copilots and agents help us understand and write code “better”

Testing



- **Enhanced testing efficiency**
- **Improvements in test script generation**
- **Automated testing and quality assurance processes**

AI helps detect, auto-heal, and predict defects

Adoption of AI tools

**Development
& Testing**

**Operations
& Monitoring**

Figure 1: Magic Quadrant for Observability Platforms



Current AI or LLM-enabled Observability Capabilities

Anomaly Detection	Root Cause Analysis	Predictive Analytics
Natural Language Query Processing	Automated Incident Correlation	Business Impact Analysis
Security Threat Detection	Log & Metric Analysis	Incident Clustering and Intelligence
Behavioral Analysis	Virtual Agents and Predictive Tools	Proactive Alerts

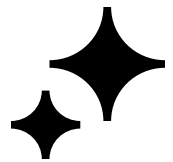
Categories of Capabilities



Incident Management

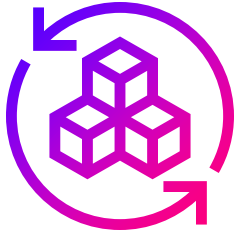


Predictive Capabilities



Query Intelligence

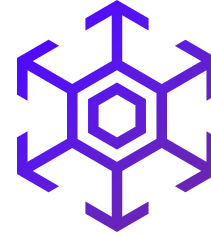
Incident Management



**Automated Root
Cause Analysis**



**Intelligent
Classification**



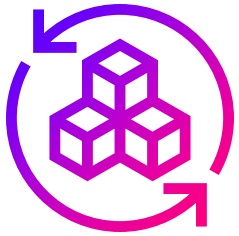
**Real-time
Response
Automation**



**Advanced Analytics
and Reporting**



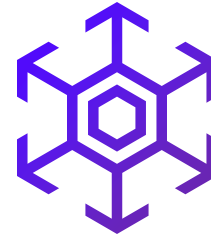
Predictive Capabilities



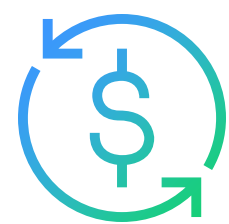
**Early Issues
Detection and
Prevention**



**Performance
Optimization**

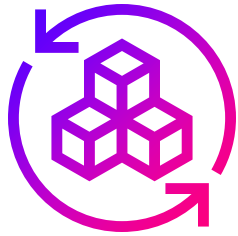


**Security and
Compliance
Enhancement**



**Cost Reduction and
Efficiency**

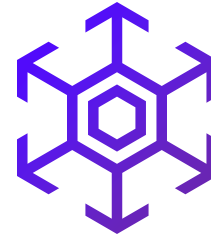
✦✦ Natural Language Querying



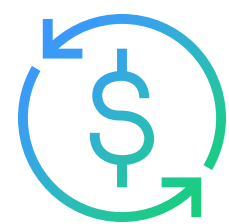
**Query
Transformation**



**Personalized
Experience**



Smart Assistance



Task Simplification

Toto: Time Series Optimized Transformer for Observability

Technical Report

Ben Cohen Emaad Khwaja Kan Wang Charles Masson
Elise Ramé Youssef Doubli Othmane Abou-Amal

{ben.cohen, emaad.khwaja, kan.wang, charles.masson, elise.rame, youssef.doubli, othmane}@datadoghq.com

This technical report describes the Time Series Optimized Transformer for Observability (Toto), a new state-of-the-art foundation model for time series forecasting developed by Datadog. In addition to advancing the state-of-the-art on generalized time series benchmarks in domains such as electricity and weather, this model is the first general-purpose time series forecasting foundation model to be specifically tuned for observability metrics.

Toto was trained on a dataset of one trillion time series data points – the largest among all currently published time series foundation models. Alongside publicly available time series datasets, 75% of the data used to train Toto consists of fully anonymous numerical metric data points from the Datadog platform.

In our experiments, Toto outperforms existing time series foundation models on observability data. It does this while also excelling at general-purpose forecasting tasks, achieving state-of-the-art zero-shot performance on multiple open benchmark datasets.

In this report, we detail the following key contributions:

- **Proportional factorized space-time attention:** We introduce an advanced attention mechanism that allows for efficient grouping of multivariate time series features, reducing computational overhead while maintaining high accuracy.
- **Student-T mixture model head:** This novel use of a probabilistic model that robustly generalizes Gaussian mixture models enables Toto to more accurately capture the complex dynamics of time series data and provides superior performance over traditional approaches.
- **Domain-specific training data:** In addition to general multi-domain time series data, Toto is specifically pre-trained on a large-scale dataset of Datadog observability metrics, encompassing unique characteristics not present in open-source datasets. This targeted training ensures enhanced performance in observability metric forecasting.

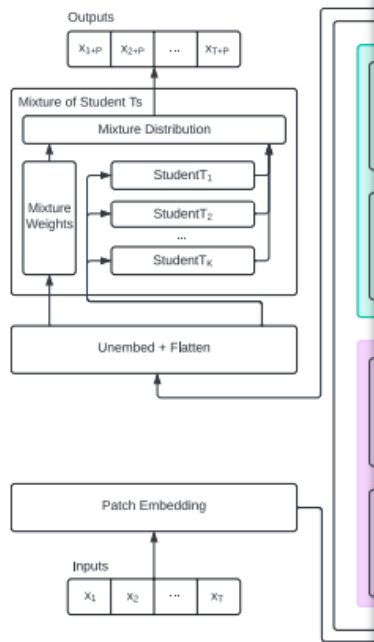


Figure 1. Toto architecture diagram. Input time series of T steps is embedded using the patch embedding layer. They then pass through each segment of the transformer. Each segment of the transformer consists of one space-wise transformer. Transformer outputs are projected to form the parameters of the Student-T mixture model head, which produces the forecasts for the input series, shifted P steps (the patch width).

1 Background

We present Toto, a groundbreaking time series forecasting foundation model developed by Datadog. Toto is specifically designed to handle the complexities of observability data, leveraging a state-of-the-art transformer architecture to deliver unparalleled accuracy and performance. Toto is trained on a massive dataset of diverse time series data, enabling it to excel in zero-shot predictions. This model is tailored to meet the demanding requirements of real-time analysis as well as compute and memory-efficient scalability to very large data volumes, providing robust so-

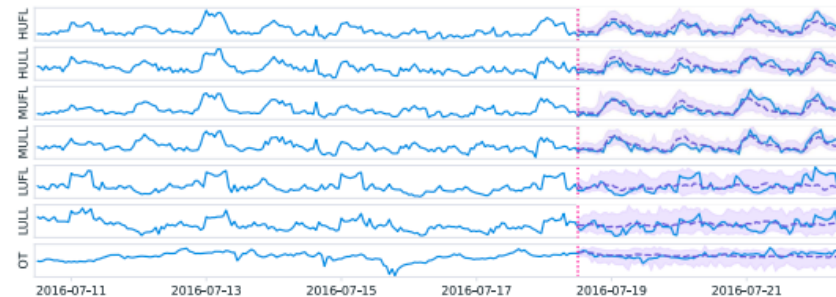


Figure 2. Example of Toto's 96-step zero-shot forecasts on the ETTh1 dataset, showing multivariate probabilistic predictions. Solid lines represent ground truth, dashed lines represent median point forecasts, and shaded regions represent 95% prediction intervals.

lating [1]. Additionally, Datadog integrates specific metrics from numerous SaaS products, cloud services, open-source frameworks, and other third-party tools. The platform allows users to apply various time series models to proactively alert on anomalous behavior, leading to a reduction in time to detection (TTD) and time to resolution (TTR) of production incidents [2].

The complexity and diversity of these metrics present significant challenges for time series forecasting. Observability data often requires high time resolution, down to seconds or minutes, and is typically sparse with many zero-inflated metrics. Moreover, these metrics can display extreme dynamic ranges and right-skewed distributions. The dynamic and non-stationary nature of the systems being monitored further complicates the forecasting task, necessitating advanced models that can adapt and perform under these conditions.

1.2 Traditional models

Historically, time series forecasting has relied on classical models such as ARIMA, exponential smoothing, and basic machine learning techniques [3]. While foundational, these models necessitate individual training for each metric, presenting several limitations [4]. The need to develop and maintain separate models for each metric impedes scalability, especially given the extensive range of metrics in observability

data. Moreover, these models often fail to generalize across different types of metrics, leading to suboptimal performance on diverse datasets [5, 6]. Continuous retraining and tuning to adapt to evolving data patterns further increase the operational burden. This scaling limitation has hindered the adoption of deep learning-based methods for time series analysis, even as they show promise in terms of accuracy [7].

1.3 Foundation models

Large neural network-based generative models, often referred to as "foundation models," have revolutionized time series forecasting by enabling accurate predictions on new data not seen during training, known as zero-shot prediction [8]. This capability significantly reduces the need for constant retraining on each specific metric, thus saving considerable time and computational resources. Their architecture supports the parallel processing of vast data volumes, facilitating timely insights essential for maintaining system performance and reliability [9, 10].

Through pretraining on diverse datasets, generative models exhibit strong generalization across various types of time series data. This enhances their robustness and versatility, making them suitable for a wide range of applications. Zero-shot predictions are particularly attractive in the observability domain, where the limitations of traditional methods are felt very acutely. The most common use cases for time series

Current AI or LLM-enabled Observability Capabilities

Anomaly Detection	AI-driven RCA	Predictive Analytics
Natural Language Query Processing	Automated Incident Correlation	Business Impact Analysis
Security Threat Detection	Log & Metric Analysis	Incident Clustering and Intelligence
Behavioral Analysis	Virtual Agents and Predictive Tools	Proactive Alerts

12 Leaders & Visionaries

Figure 1: Magic Quadrant for Observability Platforms





Datadog Platform

Infrastructure

Infrastructure Monitoring
Containers
Serverless
Network Performance Monitoring
Network Device Monitoring
Metrics
Cloud Cost Management
Cloudcraft

Applications

Application Performance Monitoring
Distributed Tracing
Continuous Profiler
Database Monitoring
Universal Service Monitoring
Data Streams Monitoring
Data Jobs Monitoring
LLM Observability

Digital Experience

Synthetics
Mobile App Testing
Browser Real User Monitoring
Mobile Real User Monitoring
Session Replay

Logs

Log Management
Observability Pipelines
Audit Trail
Log Forwarding
Error Tracking
Sensitive Data Scanner

Security

Cloud Security Management
Application Security Management
Software Composition Analysis
Cloud SIEM

Software Delivery

CI Visibility
Test Visibility
Intelligent Test Runner
Continuous Testing

Cloud Service Management

Incident Management
Event Management
Workflow Automation
App Builder



AI

Natural Language Querying • Root Cause Analysis • Anomaly Detection • Impact Analysis • Proactive Alerts • Autonomous Investigations • Bits AI

Shared Platform Services

Dashboards • CoScreen • Teams • Agent • OpenTelemetry • Notebooks • Service Catalog • IDE Plugins • ChatOps • SLOs • Case Management



UNIFIED METRICS, LOGS, TRACES, SESSIONS

800+ INTEGRATIONS

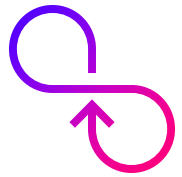
DORA

2024

Accelerate
State of DevOps

Google Cloud

DORA Metrics



**Deployment
Frequency**



**Lead Time for
Changes**



**Change Failure
Rate**



**Time to Restore
Service**

2024



The era of AI

- Most developers use AI for core work, boosting productivity.
- AI adoption is linked to improved team and organizational performance.
- Be relentlessly user-centric.

**Downstream
impacts of AI**

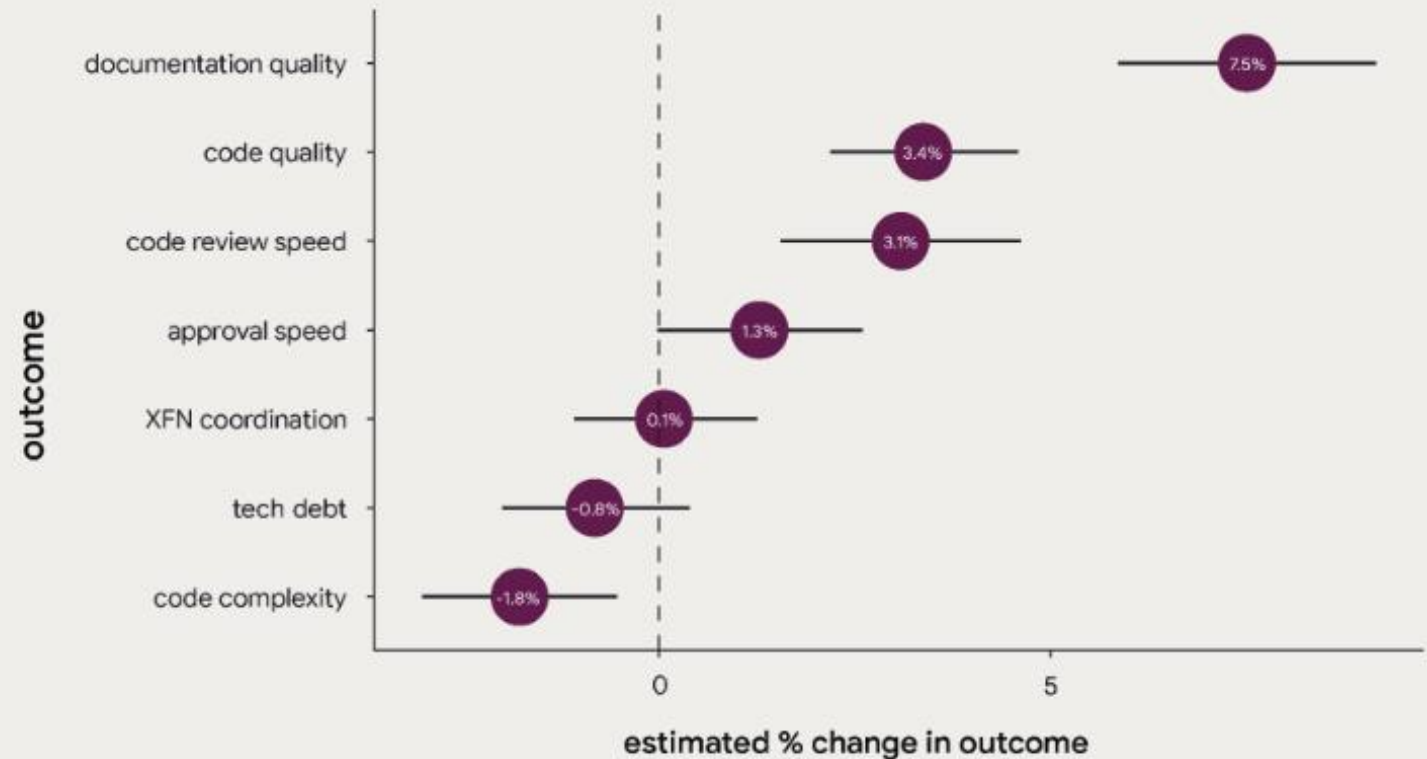
**DO
RA**

AI is removing obstacles in the developer workflow

A 25% increase in AI adoption is associated with a...

- 7.5% increase in documentation quality
- 3.4% increase in code quality
- 3.1% increase in code review speed
- 1.3% increase in approval speed
- 1.8% *decrease* in code complexity

If AI adoption increases by 25%..



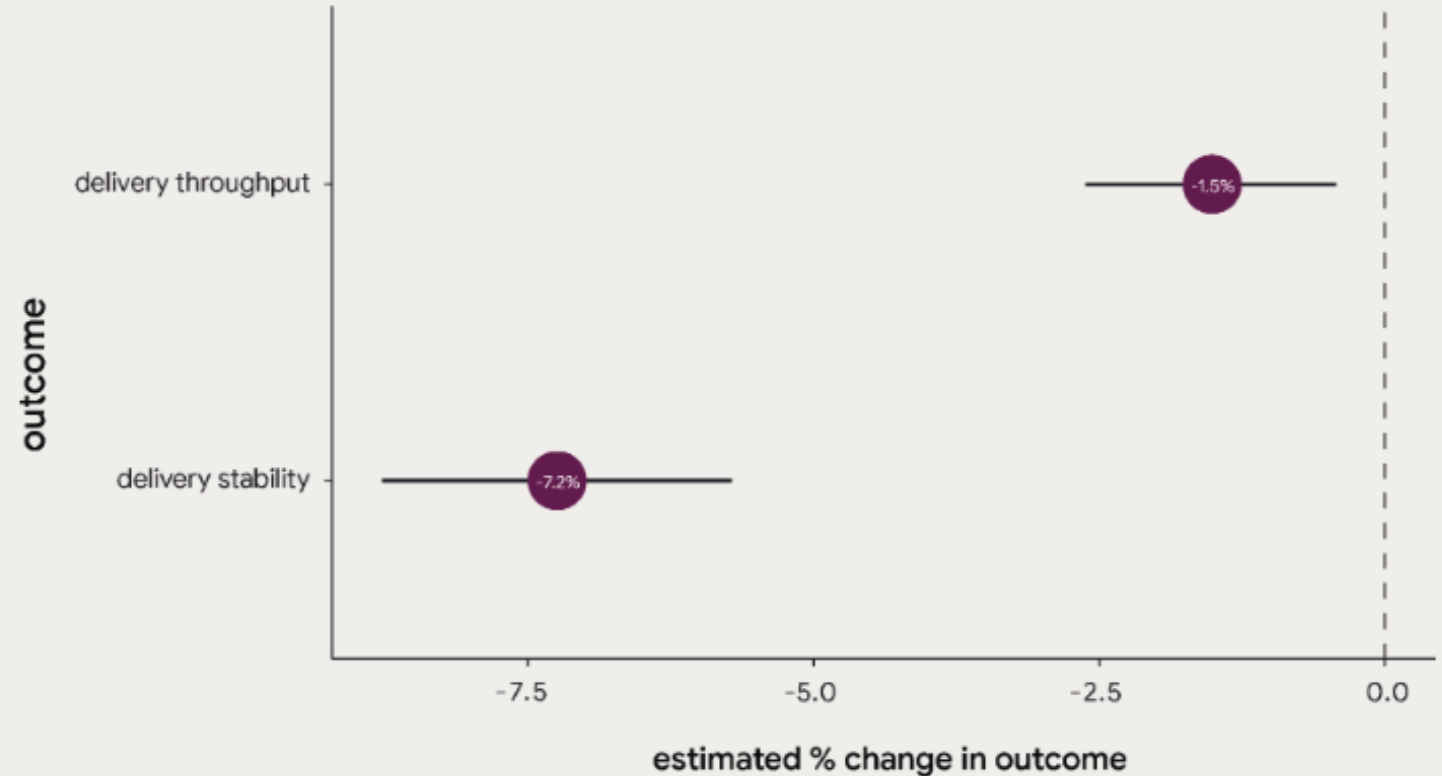
Point = estimated value
Error bar = 89% uncertainty interval

AI is hurting delivery performance

Contrary to our expectations, our findings indicate that AI adoption is negatively impacting software delivery performance.

These findings suggest gaining trust, perceived quality, and speed in one area of the development process can slow and destabilize later stages of software delivery.

If AI adoption increases by 25%..



Point = estimated value
Error bar = 89% uncertainty interval

AI Adoption



The report finds that AI adoption is rapidly increasing, with **76%** of respondents relying on AI for tasks like code writing, summarizing information, and code explanation. AI adoption is positively correlated with increased flow, productivity, job satisfaction, code quality, and internal documentation quality. However, it is associated with a decrease in delivery stability, suggesting a potential need for adapting change management practices in the age of AI-powered development.

Key metrics

81% of respondents say their company has shifted resources into developing AI.

67% of respondents report that AI is helping them improve their code.

39% of respondents reported having little or no trust in AI.

A **25%** increase in AI adoption is associated with a...

7.5% increase in documentation quality.

3.4% increase in code quality.

3.1% increase in code review speed.

1.3% increase in approval speed.

1.8% decrease in code complexity (this is a good thing!).

Change Management Techniques

Establish sense of urgency



Create a vision for the future



Communicate effectively



Data



Key Takeaways



Advanced Automated Data Analysis:
Instantly analyzing detecting anomalies



Identify hidden patterns in complex data:
Reduce false alarms, and accelerate issue correlation and resolution



Reduce manual troubleshooting:
Creating more intuitive and natural interactions

Where We Are Headed



Agentic AI:
Autonomous
Workflows



Unified AI/ML:
Monitoring End-to-end
visibility from app &
infra to model

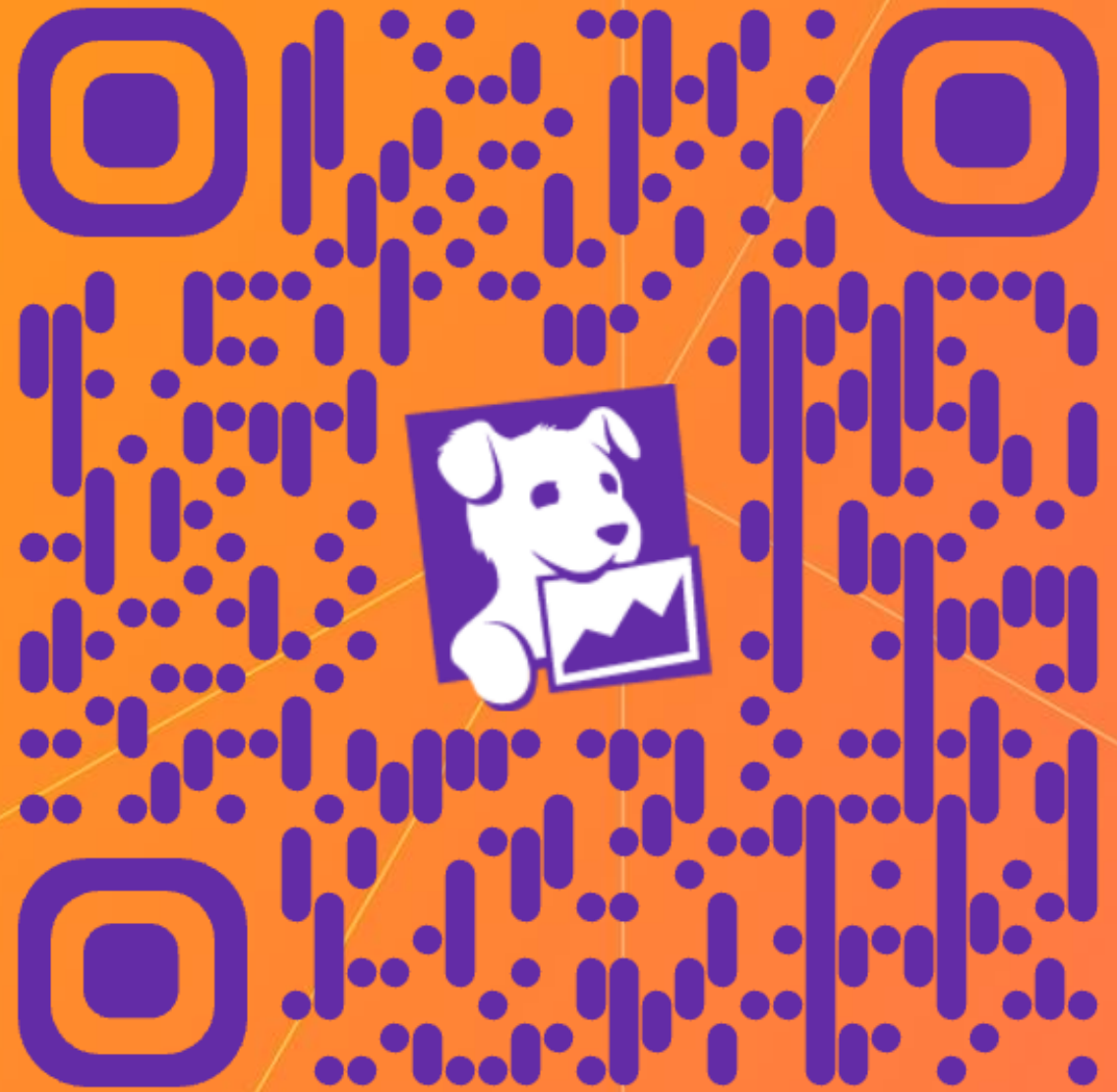


Cost Optimization:
Forecasting LLM token
usage

Explore More

dtdg.co/SCaLE22x

NOTE: Authenticate GitHub first



Thank you



SCaLE 22x



Jason Hand

Senior Developer Advocate : Infrastructure & Platform – AI : Datadog



DATADOG

