

Triangle: Empowering Incident Triage with Multi-Agent

A Multi-Agent System Reducing Incident Engagement Time by up to 91%

Multi-Agent System | Incident Management | Cloud Operations

Zhaoyang Yu¹ • Aoyang Fang⁵ • Minghua Ma² • Jaskaran Singh Walia³ • Chaoyun Zhang³ • Shu Chi¹

Ze Li¹ • Murali Chintalapati² • Xuchao Zhang² • Ruija Wang² • Chetan Bansal² • Saravan Rajmohan²

Qingwei Lin³ • Shenglin Zhang⁴ • Dan Pei¹ • Pinjia He⁵

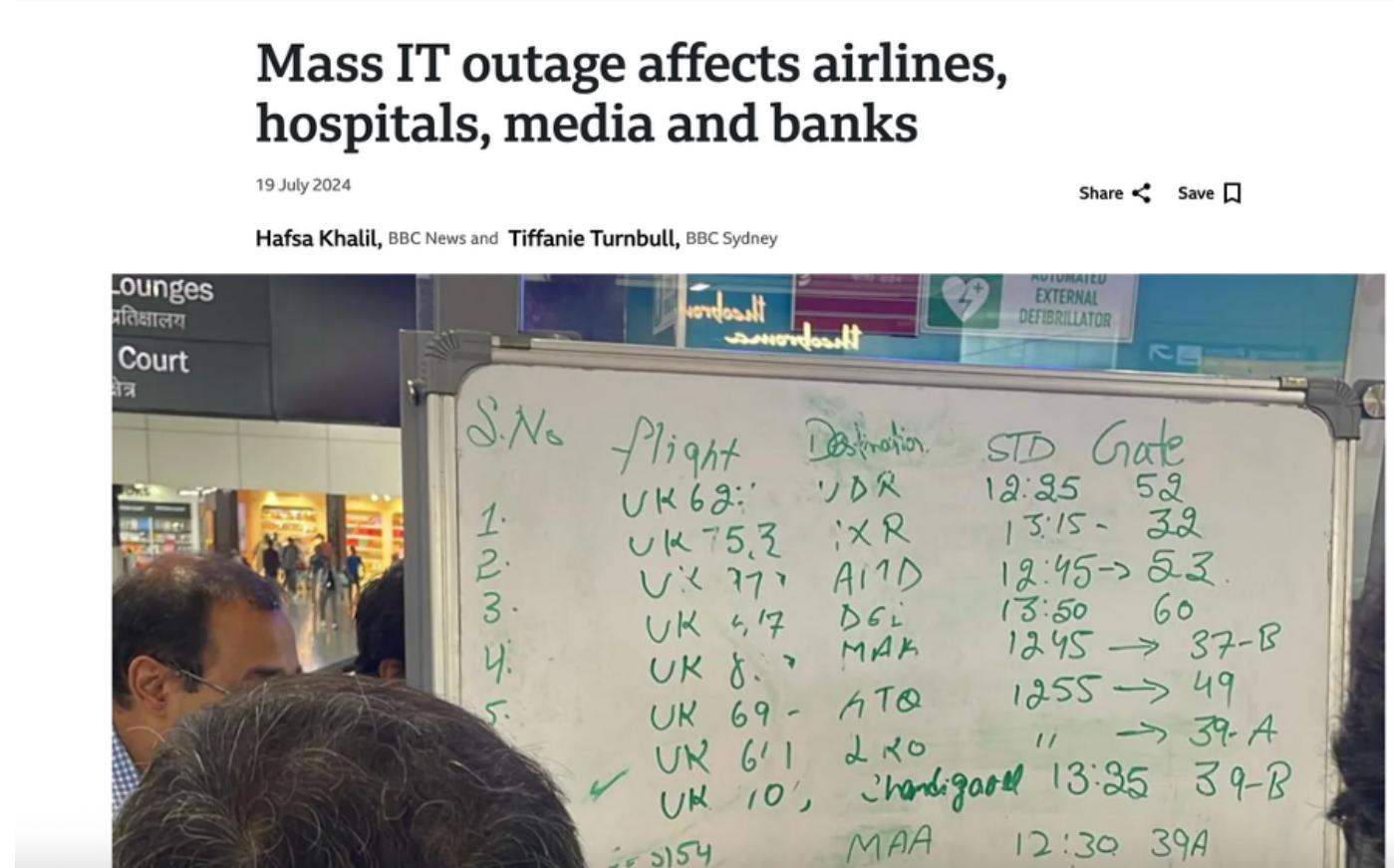
¹Tsinghua University • ²Microsoft • ³Microsoft, Beijing

⁴Nankai University • ⁵The Chinese University of Hong Kong, Shenzhen



1. Why Incident Triage Matters

High-profile outages demonstrate the critical importance of rapid incident response. Even minutes of downtime translate to millions in revenue loss and severe customer impact.



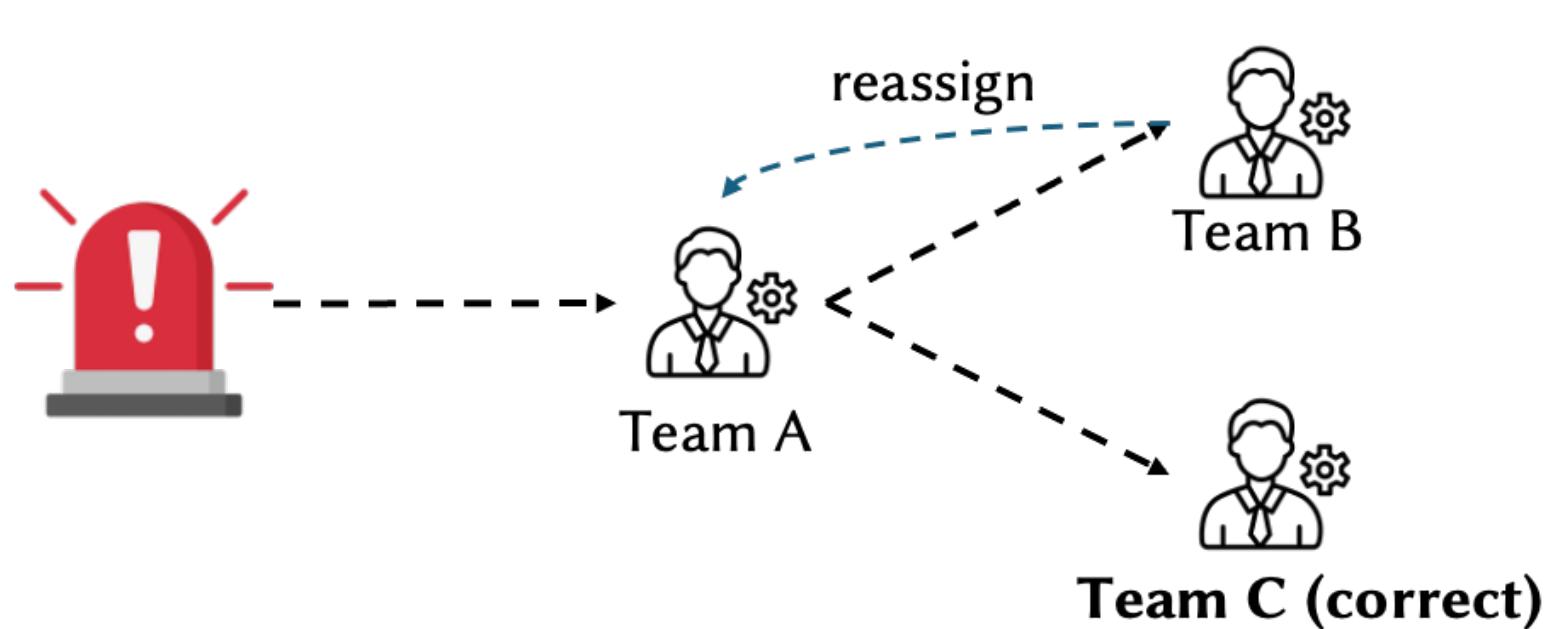
CrowdStrike and Windows Outage (July 19, 2024): Caused massive global service disruptions and economic losses.



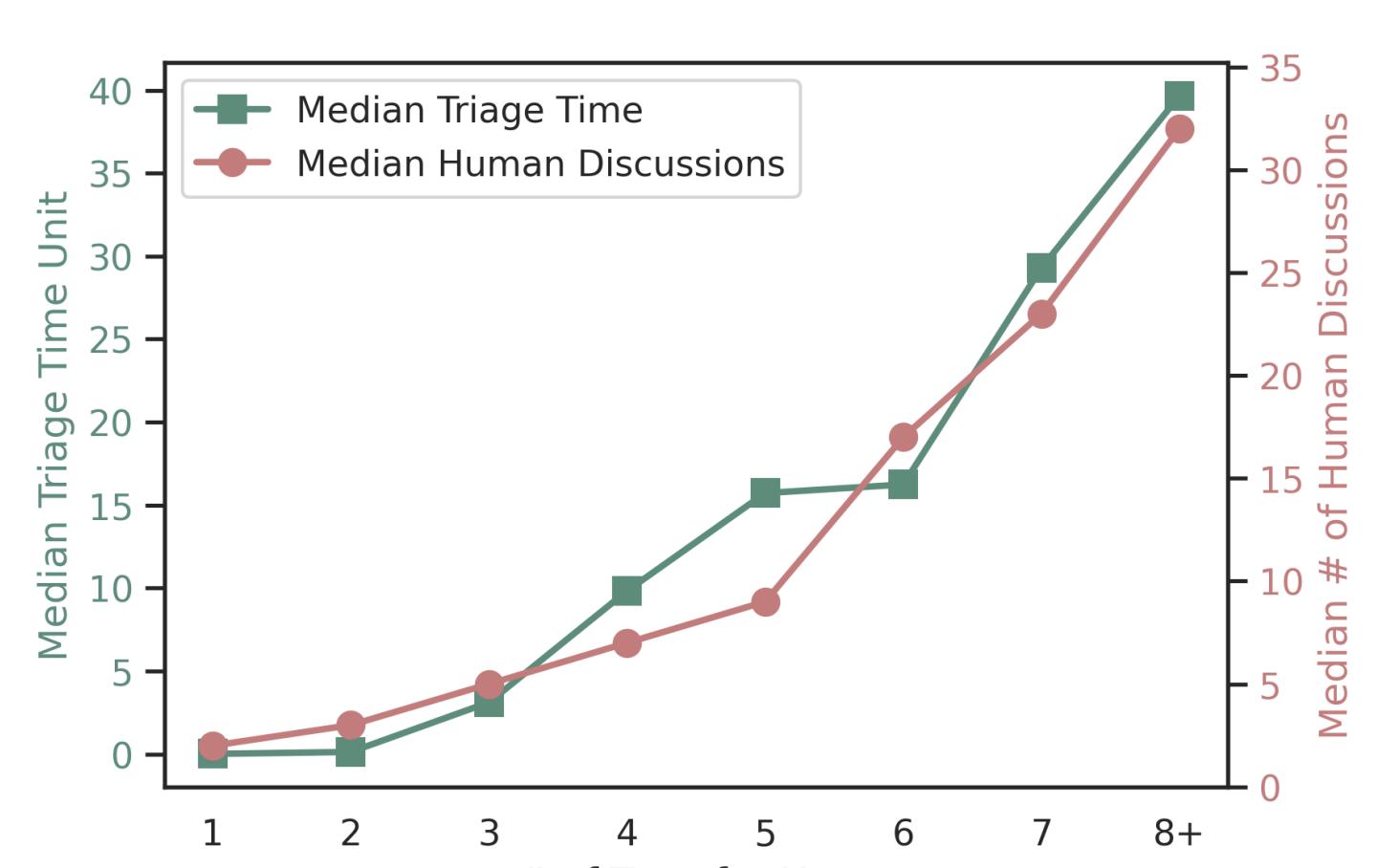
Amazon Web Services Outage (Oct 19, 2025): Impacted millions of users and businesses worldwide.

2. The Core Challenge: Triage Cycles

At the heart of incident management lies **Incident Triage**: assigning incidents to the correct team. A wrong assignment triggers costly "triage cycles" where incidents bounce between teams.



The triage process relies on slow manual operations, vague alerts, and has high business impact from delays.



Our study of 3,000+ teams: Each hop exponentially increases triage time and communication overhead.

Incident metadata: Customers are encountering a sign-in issue with the *** desktop client on Mac



Team A: Clearing the cache didn't resolve the issue. Testing on both Mac and Windows shows that works fine on Windows..



Team B: Confirm the prerequisites MacOS and perform the keychain clean up. The customer is still having issues...



Team C: Collect a HAR file while the client is starting. Identify the issue ***, apply a hot fix, and resolve the problem

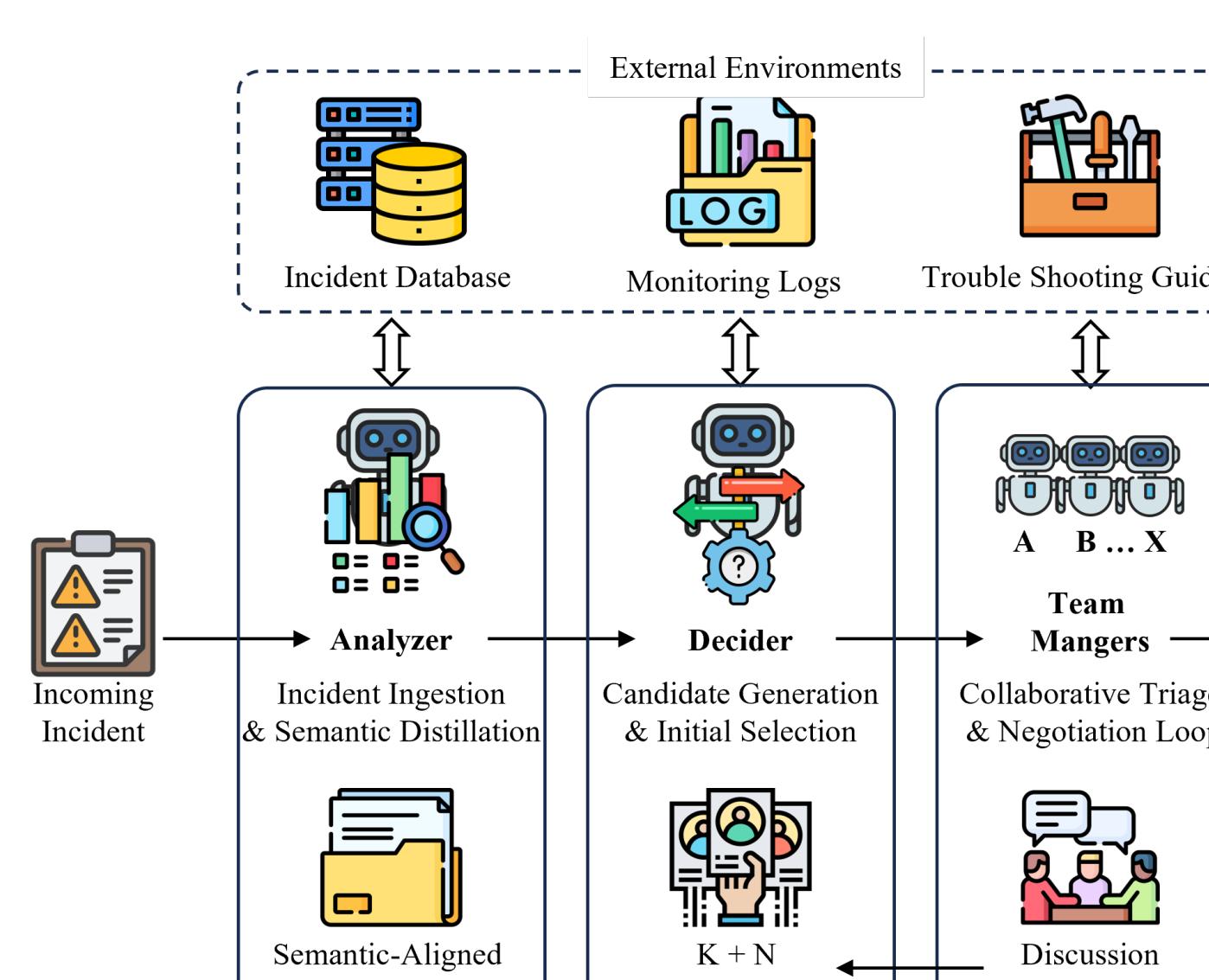
Real incident that manually escalated across teams

Real example: A Mac sign-in issue escalated across Teams A, B, and C before reaching the team with correct expertise.

3. Our Solution: Triangle

Triangle is a multi-agent system that automates triage by simulating expert team collaboration:

- **Semantic Distillation:** Analyzer Agent extracts key information from noisy data.
- **Collaborative Negotiation:** Team Manager Agents discuss, enrich with tools, and vote.
- **Automated Enrichment:** Agents query logs and monitoring systems automatically.

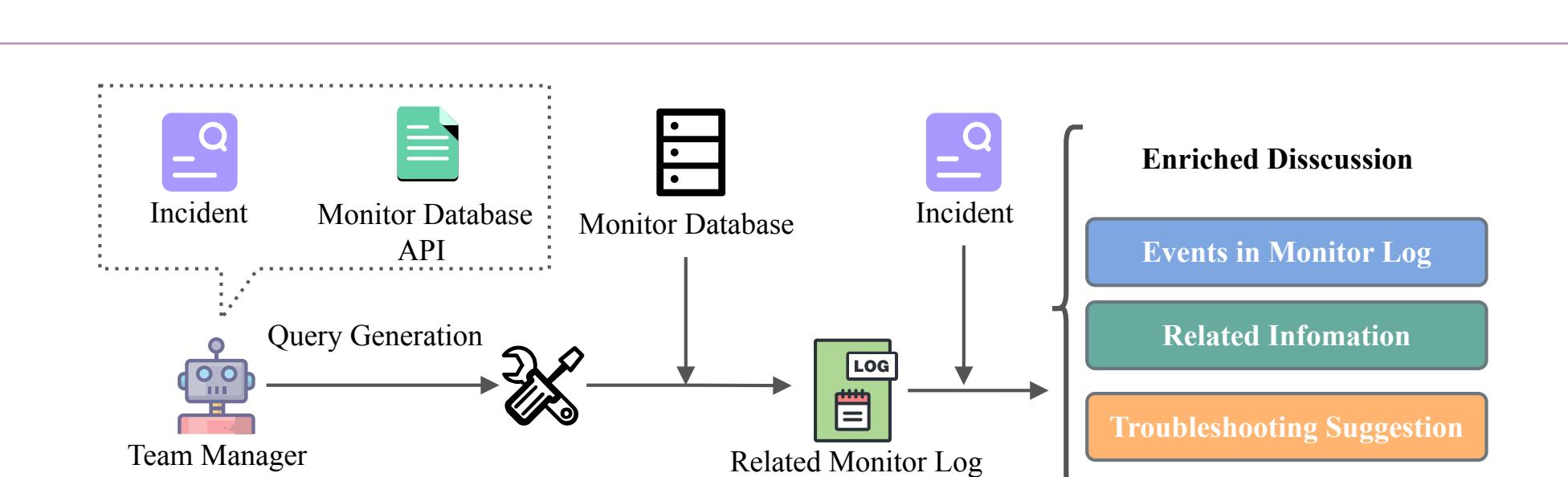


The Triangle Framework: Agents collaborate to analyze, propose, and negotiate the correct team assignment.

4. Key Results

We evaluated Triangle on real-world incident data from a large-scale cloud provider.

- Outperforms the state-of-the-art (DeepCT) by a relative margin of 26-42%.
- Achieves 91.7% accuracy after 5 hops, without relying on manual discussion data.
- Generalizes to other tasks, outperforming baselines on the MSR 2013 Bug Dataset by an average of 51%.



Team Manager Agents use tools to automatically enrich incidents with log data and analysis.

5. How It Works

Initial Incident

An incident is ingested. It's often a vague symptom (e.g., "slow response"). Traditional manual triage is slow, error-prone, and leads to costly "triage cycles."

Raw Alert: "ID#98765: Users report 'Payment Failure: Error 503' during checkout."

Phase 1: Semantic Distillation (Analyzer Agent)

The Analyzer Agent tackles "Incident Semantic Heterogeneity." It uses TF-IDF and LLMs to normalize terminology and extract key phrases, aligning the raw text with team functional documents.

Semantic-Aligned Output:

- Location: Checkout Flow
- Symptom: Payment Failure, Error 503 (Service Unavailable)
- Capability: API Gateway Diagnostics

Phase 2: Candidate Generation (Decider Agent)

The Decider Agent uses a two-pronged approach: 1) Matching against historical incidents (TF-IDF), and 2) Matching against team functional documents (LLM) to find the best candidates.

Candidate Teams (Top 3):

1. Team-PaymentGW (Historical Match)
2. Team-BillingDB (Document Match)
3. Team-Frontend (Historical Match)

Phase 3: Collaborative Triage & Negotiation

The system activates Team Manager Agents for the 3 candidates to simulate an expert consultation, handling decentralized domain knowledge.

Team Information Enrichment (TIE)

Each agent automatically generates and executes queries against its team-specific monitoring databases (logs, metrics). An LLM then summarizes the findings.

- Frontend Agent: "Queried UI logs. No errors found. Request was successfully sent to payment-api."
- PaymentGW Agent: "Queried API logs. Detected 1,500 503 errors. Reason: Timeout connecting to BillingDB cluster."
- BillingDB Agent: "Queried DB metrics. Confirmed: 14:29 CPU spiked to 100% due to a batch job."

Voting & Consensus

All agents share their enriched information. Based on the new evidence (the 100% CPU), they vote. A majority consensus is reached, concluding the triage.

Vote Result: 3 / 3 votes for Team-BillingDB

Final Assignment

The incident is automatically and accurately assigned to the root cause team without manual intervention or re-assignment hops, drastically reducing Time to Engage (TTE).

Responsible Team: Team-BillingDB

Triangle's three-phase workflow: from initial incident to final assignment through analyze, propose, and negotiate phases.

6. Real-World Business Impact

Triangle is deployed in production, serving tens of millions of users. Across six major services, it has delivered significant operational improvements.

97% Peak Triage Accuracy

91% Max TTE Reduction

20TB+ Data Processed Daily