

# Service Level Agreement (SLA)

**Version:** 1.0

**Effective Date:** June 2026

**Contact:** naeem@cliniqueue.com

---

## Overview

This Service Level Agreement defines CliniQueue's commitments to customers regarding service availability, performance, and support response times. These commitments apply to all paid plans (Compliant, Clinical, Critical, and Enterprise).

---

## Service Availability

### Uptime Commitment

Plan	Monthly Uptime Target
Compliant	99.5%
Clinical	99.5%
Critical	99.5%
Enterprise	99.9% (negotiated separately)

**99.5% uptime** means no more than approximately **3.6 hours of downtime per month**.

Uptime is measured as the percentage of minutes in a calendar month during which the `/classify` endpoint successfully responds to valid requests.

### What Counts as Downtime

Downtime is defined as any period longer than 5 consecutive minutes during which the `/classify` endpoint returns 5xx errors or does not respond.

## What Does Not Count as Downtime

- Scheduled maintenance windows (announced 48 hours in advance)
- Errors caused by invalid API keys or malformed requests (4xx errors)
- Downtime caused by third-party services outside CliniQueue's control (OpenAI outages, Railway infrastructure incidents)
- Force majeure events

## Scheduled Maintenance

Scheduled maintenance windows:

- Announced at least 48 hours in advance via email
- Scheduled outside peak hours (preference for 2:00 AM - 6:00 AM UTC on weekdays)
- Typically last less than 30 minutes
- Do not count against uptime calculations

---

## Performance

### Response Time Target

Percentile	Target
p50 (median)	< 800ms
p95	< 2,000ms
p99	< 5,000ms

Response time is measured from receipt of the HTTP request to delivery of the HTTP response, including PHI scrubbing and AI model inference.

**Note:** Response times are dependent on OpenAI's API performance. During periods of elevated OpenAI latency, CliniQueue response times will be higher. This is outside CliniQueue's direct control.

## Current Performance

As of June 2026, average observed response time is approximately 3,200ms. This is being actively optimized. Target p50 < 800ms will be achieved as the system matures and the fine-tuned model is deployed (target: Q1 2027).

---

## Support Response Times

Priority	Definition	First Response Target
P0 — Critical	API completely down, potential PHI breach	1 hour
P1 — High	API degraded, classification errors at scale	4 hours
P2 — Medium	Single customer issue, billing question	24 hours
P3 — Low	General question, feature request	3 business days

**Support channel:** [naeem@cliniqueue.com](mailto:naeem@cliniqueue.com)

**Support hours:** Support is provided via email at [naeem@cliniqueue.com](mailto:naeem@cliniqueue.com). Responses are typically within 24 hours, often sooner.

For P0 incidents outside business hours, email with subject line **URGENT:** and a response will be provided as soon as possible.

## How to Classify Your Support Request

Use these subject line prefixes to ensure correct prioritization:

- [P0] — API down, potential breach
  - [P1] — Degraded service, multiple tickets affected
  - [P2] — Account issue, billing, single ticket error
  - [P3] — General question, feature request
- 

## SLA Credits

If CliniQueue fails to meet the monthly uptime target, affected customers are eligible for service credits.

Monthly Uptime Achieved	Credit
99.0% – 99.5%	10% of monthly invoice
95.0% – 99.0%	25% of monthly invoice
Below 95.0%	50% of monthly invoice

## How to Request a Credit

Credits must be requested within 30 days of the incident. Send an email to [naeem@cliniqueue.com](mailto:naeem@cliniqueue.com) with:

- Subject: `SLA Credit Request – [Month Year]`
- The dates and times of the downtime experienced
- Your customer name and API key (last 4 characters only)

Credits are applied to the next monthly invoice. Credits are not redeemable for cash and do not apply to Enterprise plans negotiated separately.

## Credit Exclusions

Credits are not issued for downtime caused by:

- OpenAI service outages (outside CliniQueue's control)
- Railway infrastructure incidents (outside CliniQueue's control)
- Customer-side issues (network problems, misconfigured webhooks)
- Force majeure events
- Scheduled maintenance announced in advance

---

## Data Retention SLA

<b>Data Type</b>	<b>Retention Period</b>
Raw ticket text	Never stored — zero retention
Classification metadata (usage table)	Retained for the life of the customer account
Scrubbed embeddings	Retained for the life of the customer account
Automation rules	Retained until deleted by customer or account closure
Draft settings	Retained until changed or account closure

Upon account cancellation:

- Customer data is retained for 30 days after cancellation to allow for export requests
  - After 30 days, data is permanently deleted upon request
  - Customers may request immediate deletion at any time
- 

## **Accuracy SLA**

CliniQueue does not guarantee classification accuracy as part of this SLA. Current measured accuracy on the regression test suite is 99% category accuracy and 100% escalation accuracy.

However, CliniQueue commits to:

- Maintaining a publicly documented accuracy benchmark updated with each model change
  - Responding to accuracy complaints within the support response times above
  - Reprocessing tickets at no charge if a systematic accuracy issue is identified and confirmed
- 

## **Changes to This SLA**

CliniQueue reserves the right to update this SLA. Customers will be notified via email at least 30 days before any reduction in SLA commitments takes effect. Improvements to SLA commitments take effect immediately.

---

## **Entire Agreement**

This SLA is incorporated by reference into the CliniQueue Terms of Service and Business Associate Agreement. In the event of a conflict between this SLA and those documents, those documents take precedence.

---

*Questions about this SLA? Contact [naeem@cliniqueue.com](mailto:naeem@cliniqueue.com)*