Service Level Agreement

> Draft for review. This document is not binding until incorporated into an executed Enterprise Agreement or Order Form.

Provider	ONTECH DIGITAL PTE. LTD. ("RouteAPI", "we", "us"), 186 Depot Road, #11-13 The Interlace, Singapore
Service	RouteAPI API Gateway (the "Service")
Version	1.0
Effective Date	The effective date of the Enterprise Agreement or Order Form that references this SLA.

This Service Level Agreement ("SLA") applies to customers that have executed an Enterprise Agreement or Order Form referencing this SLA ("Enterprise Customers", "you"), whether the account is billed on a prepaid or postpaid basis. It does not apply to self-service accounts without an Enterprise Agreement, free tiers, trials, or beta features. This SLA also does not apply to dedicated (private) deployments; service levels for dedicated deployments are set out in the applicable deployment agreement.

1. Definitions

1.1 "Eligible Usage Charges" means: (a) for prepaid accounts, the charges actually deducted from your account balance for use of the Service during the calendar month, excluding unused top-up balance, taxes, refunds and previously issued credits; and (b) for postpaid accounts, the charges calculated for that calendar month under the applicable Enterprise Agreement, before credits.

1.2 "Monthly Uptime Percentage" means, for each calendar month:

```
Monthly Uptime Percentage = (Total Minutes − Downtime Minutes) / Total Minutes × 100
```

where "Total Minutes" means every minute of that calendar month.

1.3 "Downtime" means any period of at least five (5) consecutive minutes during which the Gateway Error Rate exceeds ten percent (10%). Downtime is measured in whole minutes and does not include any period attributable to an Exclusion under Section 4.

1.4 "Gateway Error Rate" means, for any one-minute interval, the number of Eligible Requests across the entire gateway that result in a Gateway Error, divided by the total number of Eligible Requests across the entire gateway in that interval. The Gateway Error Rate is not calculated on a per-model basis.

1.5 "Gateway Error" means either: (a) a response with HTTP status code 500, 502, 503 or 504 that originates in RouteAPI infrastructure; or (b) no response within thirty (30) seconds, including connection refusal or reset, TLS handshake failure, DNS failure, or any other transport-level failure that yields no HTTP status. The following are not Gateway Errors: responses caused by the request or its authorisation (including 400, 401, 403 and 422), rate limiting (429), and errors that originate with an upstream model provider, even where RouteAPI returns them with the same status code.

1.6 "Eligible Request" means a well-formed API request to a production endpoint of the Service, authenticated with a valid API key, that is not rejected for reasons attributable to you (see Section 4.3).

1.7 "Measurement Source" means RouteAPI's synthetic availability probes, which issue requests to the Service through the same public ingress used by customers, each subject to a thirty (30) second response timeout. Results are published at https://routeapi.ai/status.

2. Service Commitment

2.1 RouteAPI will use commercially reasonable efforts to make the Service available with a Monthly Uptime Percentage of at least **99.9%** in each calendar month (the "Service Commitment").

2.2 Gateway-level availability. The Service Commitment applies to the gateway as a whole, across all models and upstream channels, and not to any individual model. The Service is considered available at any given time if at least one model supported by the gateway is capable of accepting and successfully processing requests.

2.3 Per-model information. RouteAPI publishes availability information for individual models at https://routeapi.ai/status for transparency. Service Credits under this SLA are calculated solely on gateway-level availability as defined in this SLA, and not on the availability of any individual model.

2.4 For reference, 99.9% corresponds to a maximum of approximately 43.2 minutes of Downtime in a 30-day month and 44.6 minutes in a 31-day month.

3. Service Credits

3.1 If the Service Commitment is not met in a calendar month, you are eligible for a Service Credit calculated as a percentage of your Eligible Usage Charges for that month:

Monthly Uptime Percentage	Service Credit
Less than 99.9% but greater than or equal to 99.0%	**10%**
Less than 99.0% but greater than or equal to 95.0%	**25%**
Less than 95.0%	**50%**

3.2 Service Credits:

(a) are applied as a credit to your account balance within thirty (30) days of approval and may be used for future use of the Service; they are not offset against a future invoice;

(b) are your sole and exclusive remedy for any failure to meet the Service Commitment;

(c) may not exceed 50% of the Eligible Usage Charges for the affected month; and

(d) are not cumulative with any other credit for the same period.

4. Exclusions

The Service Commitment does not apply to, and Downtime does not include, any unavailability, suspension or degradation caused by:

DRAFT

4.1 Scheduled maintenance announced at least forty-eight (48) hours in advance via the status page and by email to your designated contact, not exceeding four (4) hours per calendar month in aggregate. RouteAPI performs maintenance on production nodes one at a time so that routine maintenance does not interrupt traffic to the Service.

4.2 Upstream model provider failure, meaning unavailability, rate limiting, errors or degraded output of the third-party model provider selected for the affected request, but only to the extent that such failure is confirmed by that provider's publicly available status page or incident report. RouteAPI bears the burden of demonstrating such confirmation. Failures within RouteAPI's own aggregation, routing or supply layer, including the relays, accounts or endpoints through which RouteAPI reaches a model provider, are not excluded and count as Downtime.

4.3 Customer-attributable causes, including invalid requests (4xx), exceeded quota or rate limits, suspended or unfunded accounts, misuse, or use of the Service in violation of the Terms of Service.

4.4 Force majeure, meaning events beyond RouteAPI's reasonable control, including natural disasters, war, terrorism, labour disputes, governmental action, Internet backbone or DNS failures outside RouteAPI's network, and denial-of-service attacks that cannot reasonably be mitigated.

4.5 Beta, preview or deprecated features, including any model marked "preview", "experimental" or "deprecated" in the model listing.

4.6 Your network, software or equipment, or third-party services not under RouteAPI's control.

5. Credit Request Procedure

5.1 Service Credits are not applied automatically. To receive a Service Credit, you must submit a request by email to support@routeapi.ai within thirty (30) days after the end of the calendar month in which the Downtime occurred.

5.2 The request must include: (a) your account identifier; (b) the dates and times of each Downtime incident claimed; and (c) request IDs and timestamps for affected requests, or server logs documenting the errors.

5.3 RouteAPI will review the request against the Measurement Source and its own request logs and will respond within fifteen (15) business days. If RouteAPI confirms that the Service Commitment was not met, the Service Credit will be applied in accordance with Section 3.2(a).

6. Status Page and Measurement Records

RouteAPI publishes availability information at https://routeapi.ai/status. The status page is provided for transparency. In the event of any discrepancy between the status page and RouteAPI's internal measurement records, RouteAPI's internal measurement records govern.

7. Changes to this SLA

RouteAPI may modify this SLA by giving at least thirty (30) days' notice. No modification will reduce the Service Commitment for an Enterprise Customer during the then-current term of its Enterprise Agreement.

ONTECH DIGITAL PTE. LTD. · 186 Depot Road, #11-13 The Interlace, Singapore
Current version: https://routeapi.ai/sla · support@routeapi.ai

DRAFT