Generate a policy for API rate limiting.

**API Rate Limiting Policy**  
**Effective Date**: [Insert Date]  
**Last Updated**: [Insert Date]

**1. Purpose**

To ensure fair usage, prevent abuse, maintain system stability, and protect API resources from excessive or malicious traffic by implementing controlled access thresholds.

**2. Scope**

Applies to all users, applications, and third-party integrations accessing the **[Company Name]** API.

**3. Rate Limiting Strategy**

**3.1. Rate Limits**

* **Standard Limits**:
  + **Authenticated Users**: X requests per second/minute/hour (e.g., 100 requests/minute).
  + **Anonymous/Unauthenticated Users**: Lower threshold (e.g., 10 requests/minute).
* **Tiered Limits**:
  + **Free Tier**: Basic limits (e.g., 50 requests/minute).
  + **Premium Tier**: Higher limits (e.g., 1,000 requests/minute).
  + **Enterprise/Custom**: Negotiated limits based on contract.
* **Dynamic Adjustments**: Limits may adjust temporarily during high traffic or system stress.

**3.2. Throttling**

* Requests exceeding limits are throttled (delayed) or rejected.
* **Burst Allowance**: Short-term bursts (e.g., 10% over limit) permitted if system capacity allows.

**3.3. Rate Limit Headers**

API responses include headers to inform clients of their status:

* X-RateLimit-Limit: Total allowed requests in the window.
* X-RateLimit-Remaining: Remaining requests.
* X-RateLimit-Reset: Time (in seconds) until the limit resets.

**4. Exceeding Rate Limits**

* **HTTP Response Code**: 429 Too Many Requests.
* **Retry-After Header**: Informs clients when to retry (e.g., Retry-After: 60).
* **Penalties**: Repeated violations may trigger temporary blocks or account suspension.

**5. Exemptions**

* **Whitelisted IPs/Accounts**: Critical partners or internal services.
* **Emergency Overrides**: Temporarily lift limits during approved emergencies (managed by DevOps/API team).

**6. Monitoring & Reporting**

* **Real-Time Dashboards**: Track usage patterns, violations, and system health.
* **Alerts**: Notify users approaching limits via email or API responses.
* **Logging**: All rate-limiting events logged for audit and analysis.

**7. Compliance & Enforcement**

* Users must comply with limits; circumvention (e.g., IP rotation) is prohibited.
* Violations may result in warnings, throttling, or termination of API access.

**8. Support & Documentation**

* **Documentation**: Clearly publish rate limits in API docs, including examples.
* **Contact**: Provide support channels (e.g., [support@company.com](https://mailto:support@company.com/)) for limit inquiries or exemptions.
* **Feedback**: Users may request limit adjustments via a formal process.

**9. Policy Review**

This policy is reviewed annually or as needed to align with technical and business requirements.

**Approved By**: [Name/Title]  
**Contact**: [Email/Phone]

**Note**: Customize thresholds, tiers, and enforcement rules based on your API’s technical capacity and business needs. Use tools like API gateways (e.g., Kong, AWS API Gateway) or middleware (e.g., Redis) to enforce limits.