Urgent: Response to API Policy Warning - Need Clarification

Briggs from OpenAI <trustandsafety@openai.com> Mon, 10 Feb at 9:13 PM
Reply-To: Briggs from OpenAI <trustandsafety@openai.com>
To: <ramzi.sub1@gmail.com>

Hello Ramzi,

Thank you for reaching out to OpenAI support.

We appreciate your patience while we reviewed your case. We've reinstated your organization (org-iBgww4CC3tp4kXVYpMjW5jdt), and you should now have access to your account. We understand how critical this is for your work, and we appreciate your commitment to ensuring compliance with our policies.

While we can't provide specific details about the requests that triggered the warning, here are some possibilities that could have led to this situation:

1. **Policy Violations**: The warning may have been triggered by requests that generated content violating our Usage Policies. This includes harmful, illegal, or prohibited content, or attempts to bypass safety systems.
2. **User Inputs or Outputs**: It's possible that specific user inputs or outputs from your implementation inadvertently violated our policies. Using the Moderation endpoint can help you proactively detect and handle such content.
3. **Implementation Issues**: If your API implementation allows unfiltered user inputs or lacks safeguards, it could lead to unintended violations. Reviewing your implementation to ensure it aligns with our guidelines is a good step forward.

To address your questions:

1. **Specific Examples of Triggered Requests**: Unfortunately, we cannot provide specific examples of the requests that triggered the warning. However, reviewing your recent API usage and logs may help identify any patterns or issues.
2. **Relation to User Inputs or Implementation**: Both user inputs and implementation design can contribute to policy violations. We recommend using the Moderation endpoint to screen inputs and outputs for compliance. You can find more details in our Moderation documentation.
3. **Guidance on Moderation Endpoint**: The Moderation endpoint is a powerful tool to help you detect and handle unsafe content. It can be used to check both user inputs and model outputs. For implementation details, please refer to the Moderation guide.

To help protect your API key and prevent unauthorized usage in the future, here are some best practices:

- **Rotate Your API Key**: If you suspect your API key may have been compromised, immediately rotate it in your API key settings.
- **Use Environment Variables**: Store your API key securely in environment variables rather than hardcoding it into your application.

- **Set Usage Limits**: Set a monthly usage limit to prevent unexpected overages.
- **Monitor Usage**: Regularly review your API usage to detect any unusual activity.
- **Enable Two-Factor Authentication (2FA)**: Add an extra layer of security to your account by enabling 2FA. You can find instructions here.

If you have further questions or encounter any issues, please don’t hesitate to reach out. We’re here to help ensure your experience with OpenAI is smooth and productive.

Best,
Briggs
OpenAI Support

OpenAI

[Quoted text hidden]