



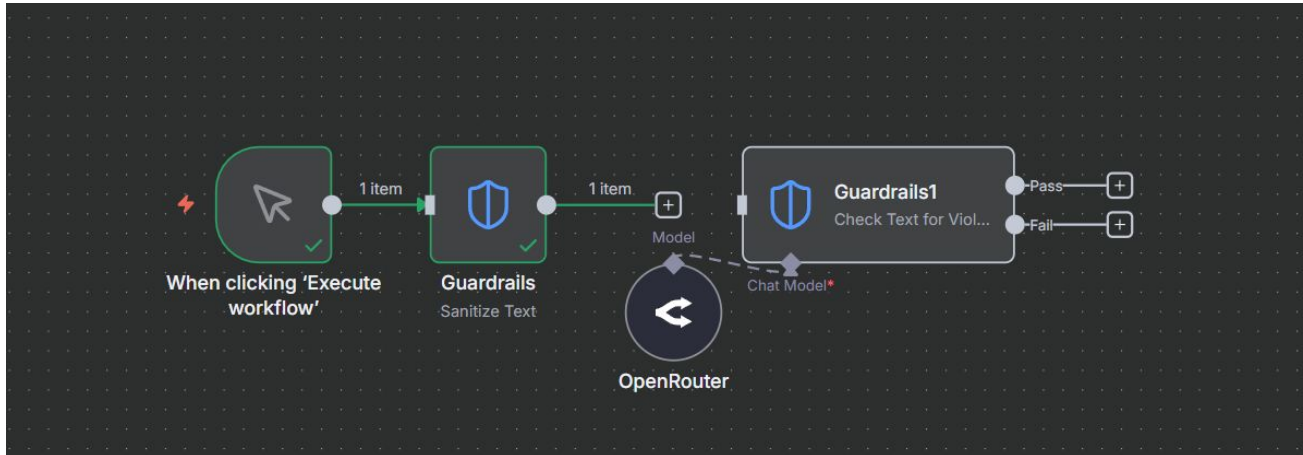
New node

**Guardrails**

# What Are Guardrail Nodes?

Guardrail nodes in n8n are specialized workflow components that enforce rules or “guardrails” on incoming or outgoing text. You can use them to:

- Validate user input before sending it to an AI model.
- Check AI-generated output before using it in your workflow.
- Automatically block, flag, or sanitize content that violates defined policies, such as NSFW content, PII, secret keys, or malicious URLs.



## How Do They Work?

Guardrail	Function
Keywords	Blocks specified words or phrases
Jailbreak	Detects prompt injection/exploit attempts
NSFW	Flags not-safe-for-work content
PII	Detects personally identifiable info (credit cards, emails, etc.)
Secret Keys	Flags API credentials/keys in text
Topical Alignment	Keeps conversation/content within a specified scope
URLs	Manages permitted URLs and blocks risky schemes or patterns
Custom	Lets you set LLM-driven or regex-based custom rules