

# Appendix A. Public Issue Sample for LLM Gateway Product Research

Retrieved: 2026-05-25 | Corpus: 75 public GitHub issue observations | Unit of analysis: one issue observation

## A1. Purpose and Scope

This appendix supports the article's issue-taxonomy section. The sample is designed for qualitative product research: it identifies recurring engineering friction around provider/schema fidelity, streaming parity, deployment, logging, reliability, agent/tool state, security, and multimodal resource handling. It is not a probability sample and should not be read as a market-wide defect-rate estimate.

The project frame covers APIPark, LiteLLM, Higgs, New API, Portkey, Bifrost, APISIX, Envoy AI Gateway, Helicone, vLLM, SGLang, and llama.cpp. Sources are public GitHub issue pages; row-level URLs are kept in the CSV dataset.

## A2. Coding Protocol

Code	Definition
provider_adapter_schema	Provider compatibility, model/channel configuration, OpenAI-compatible schema and parameter preservation.
streaming_protocol_parity	Streaming vs non-streaming behavior, SSE/chunk handling, Responses API deltas, reasoning/thinking fields.
agent_context_state	MCP, tool/function calling, agent context, parser/renderer, state offload and context compression.
deployment_ops	Offline install, Docker/Compose/Kubernetes/Helm, Gateway API compatibility and private-network deployment.
logging_audit_cost	Request log archive, audit traces, metrics, usage, cost calculation, billing and token accounting.
reliability_routing_key	Retry, fallback, timeout, latency, rate limits, key rotation/disablement and health checking.
multimodal_resource	Images, audio, video, files, URLs, base64, object storage and persistent resource handling.
security_guardrails	Auth, permission, tenant isolation, secrets, guardrails, prompt/tool safety and abuse prevention.

L3 relevance does not refer to OSI Layer 3. It estimates whether an issue touches the L3 business asset and compliance layer: durable resources, object storage/CDN persistence, permissions, audit evidence, long-term logs, cost ledger, and policy enforcement.

## A3. Aggregate Counts

Primary category	Count	Share
provider_adapter_schema	18	24.0%
deployment_ops	12	16.0%
agent_context_state	11	14.7%
logging_audit_cost	9	12.0%
streaming_protocol_parity	9	12.0%
reliability_routing_key	7	9.3%
security_guardrails	5	6.7%
multimodal_resource	4	5.3%

Build locus	Count	Share
community_build_in	39	52.0%
self_operated	21	28.0%
mixed	15	20.0%

L3 relevance	Count	Share
medium	34	45.3%
high	29	38.7%
low	12	16.0%

## A4. Project Coverage

Project	Count	Share
APIPark	8	10.7%
Bifrost	8	10.7%
Envoy AI Gateway	8	10.7%
Helicone	8	10.7%
Higress	8	10.7%
LiteLLM	8	10.7%
Portkey	8	10.7%
New API	7	9.3%
APISIX	6	8.0%
vLLM	3	4.0%
SGLang	2	2.7%
llama.cpp	1	1.3%

## A5. Row-Level Issue Sample

Full URLs and machine-readable fields are available in llm-gateway-issue-sample.csv. Titles are truncated in the PDF table for layout.

Project	#	Primary category	Locus	L3	Title
APIPark	438	deployment_ops	community_build_in	low	bug: 项目在离线网络环境下部署失败，有无离线部署教程
APIPark	428	logging_audit_cost	self_operated	high	feat: 希望增加请求日志归档功能
APIPark	427	streaming_protocol_parity	community_build_in	medium	help request: 本地模型发布后，在调用的时候，非流式返回内容为空，流式正常
APIPark	437	provider_adapter_schema	community_build_in	medium	help request: API 参数无法维护
APIPark	434	reliability_routing_key	self_operated	medium	Bug Report: APIpark Gateway 504 Timeout Issue
APIPark	433	provider_adapter_schema	community_build_in	medium	help request: 自定义的rest openAPI访问问题
APIPark	432	deployment_ops	community_build_in	low	help request: 使用 Docker-Compose部署 失败
APIPark	430	provider_adapter_schema	community_build_in	medium	bug: 添加自定义供应商之后，删除时Apinto未关联删除
Bifrost	3719	logging_audit_cost	self_operated	high	[Bug]: HTTP access log accidentally disabled in CorsMiddleware since transports/v1.5.3
Bifrost	3710	provider_adapter_schema	community_build_in	medium	[Bug]: Custom OpenAI-compatible providers with non-/v1 API paths are unroutable — request_path_overrides igno...
Bifrost	3701	reliability_routing_key	self_operated	medium	[Feature]: Automatic retry req with different api key
Bifrost	3696	agent_context_state	mixed	high	[Feature]: Make retry backoff context-aware and honor Retry-After for rate limits
Bifrost	3688	provider_adapter_schema	community_build_in	medium	[Bug]: Bedrock provider emits reasoning after tool use in assistant content, breaking Anthropic multi-turn to...
Bifrost	3667	security_guardrails	self_operated	high	[Bug Report] /anthropic/v1/messages?beta=true indicates unauthorized.
Bifrost	3666	logging_audit_cost	self_operated	high	[Bug Report] Why is there no model price?
Bifrost	3663	provider_adapter_schema	community_build_in	medium	[Bug Report] There is a problem with the provider alias.
Envoy AI Gateway	2151	provider_adapter_schema	community_build_in	medium	anthropic→openai translator: response surfaced as HTTP 500 (empty body) when upstream returns OpenAI error wi...
Envoy AI Gateway	2150	agent_context_state	mixed	high	MCPRoute controller mints one redundant mcp-proxy Backend per route — consider sharing one
Envoy AI Gateway	2149	streaming_protocol_parity	community_build_in	medium	SSE parser silently drops fields when colon is not followed by a space
Envoy AI Gateway	2145	agent_context_state	mixed	high	bug: MCP session encryption seed is exposed as plaintext in controller Deployment args
Envoy AI Gateway	2141	agent_context_state	mixed	high	bug: MCP backend API Keys are embedded as plaintext in HTTPRoute
Envoy AI Gateway	2137	provider_adapter_schema	community_build_in	medium	Feature Request: shadow traffic mirroring for AIGatewayRoute with per-mirror model rewriting
Envoy AI Gateway	2124	provider_adapter_schema	community_build_in	medium	Anthropic built-in tools (web_search_20260209) are stripped when routed through OpenAI-compatible /v1/chat...
Envoy AI Gateway	2119	provider_adapter_schema	community_build_in	medium	bug: RegularExpression model header match silently omits models from /v1/models with no validation error
Helicone	5684	provider_adapter_schema	community_build_in	medium	[Feature]: Question: how do I add an OpenAI compatible provider? I need to add my self hosted model
Helicone	5672	provider_adapter_schema	community_build_in	medium	ProviderClient.ts — 429 and 5xx use same exponential backoff, Retry-After never consulted
Helicone	5658	deployment_ops	community_build_in	low	[Bug]: 不是，我自己部署的本地服务，怎么访问内网链接结果到你他妈的官网里面了
Helicone	5657	deployment_ops	community_build_in	low	[Bug]: docker image broken?
Helicone	5650	agent_context_state	mixed	high	Clarvia: Pre-flight tool quality scoring as an LLM observability signal
Helicone	5645	logging_audit_cost	self_operated	high	Integration: Cryptographic signing for LLM logs with asqav
Helicone	5643	agent_context_state	mixed	high	MCP server: support configurable base URL for EU instances

Project	#	Primary category	Locus	L3	Title
Helicone	5597	security_guardrails	self_operated	high	Security: Cross-Organization Provider Key IDOR in Vault Endpoint
Higress	3877	provider_adapter_schema	community_build_in	medium	DeekSeek官方服务商使用/v1/responses端口请求失败问题    DeekSeek official service provider failed to use /v1/responses port r...
Higress	3868	deployment_ops	community_build_in	low	v2.1.9本地镜像构建gateway报supercronic-linux-amd64包404    v2.1.9 local image build gateway reports supercronic-linux...
Higress	3866	reliability_routing_key	self_operated	medium	自定义插件连接redis失败问题    The custom plug-in fails to connect to redis.
Higress	3865	deployment_ops	community_build_in	low	higress安装key-auth插件失败，造成网关崩溃
Higress	3863	deployment_ops	community_build_in	low	关于higress在k8s安装后通过ingress访问higress-console界面打开是Thanks for using Higress!的问题    Thanks for using Higress! Rega...
Higress	3861	agent_context_state	mixed	high	添加mcpbridge启用mcpServer后，higress-gateway崩溃，Envoy exited with error: signal: trace/breakpoint trap
Higress	3858	deployment_ops	community_build_in	low	The v2.2.2 arm64 image is corrupted.
Higress	3857	deployment_ops	community_build_in	low	Higress gatewayclass 显示unknown，无法使用gateway api与http route
LiteLLM	28781	reliability_routing_key	self_operated	medium	LOW: Missing SECURITY.md - AI gateway handling LLM API keys
LiteLLM	28780	deployment_ops	community_build_in	low	HIGH: Hardcoded DB password dbpassword9090 in docker-compose - stores all LLM API keys
LiteLLM	28778	agent_context_state	mixed	high	[Bug]: When an agent with tools turns on either redis-semantic or qdrant-semantic cache mode and carries tool ...
LiteLLM	28777	streaming_protocol_parity	community_build_in	medium	[Bug]: streamGenerateContent passthrough always returns SSE format, breaking google-genai SDK
LiteLLM	28756	reliability_routing_key	self_operated	medium	use_chat_completions_url_for_anthropic_messages config key not loaded from proxy YAML
LiteLLM	28747	provider_adapter_schema	community_build_in	medium	[Bug]: /eu.assemblyai passthrough resolves to default US AssemblyAI endpoint
LiteLLM	28734	deployment_ops	community_build_in	low	[Bug]: Docker image pins glibc version in /etc/apk/world, preventing security upgrades
LiteLLM	28689	reliability_routing_key	self_operated	medium	[Bug]: max_budget_in_team: null via /team/member_update causes 401 auth crash — LiteLLM_TeamMembership.litell...
New API	5101	logging_audit_cost	self_operated	high	fix: show actual image quality value in log instead of always 'standard'
New API	5096	logging_audit_cost	self_operated	high	使用日志页面过几秒就总报500错误
New API	5091	deployment_ops	community_build_in	low	feat:如果newapi部署容器环境中是否增加网关代理
New API	5089	multimodal_resource	mixed	high	<b>【Bug】</b> 有人在NewAPI中正确调用gpt-image-2吗？
New API	5080	logging_audit_cost	self_operated	high	Async task refund logs miss request_id and upstream_request_id
New API	5069	logging_audit_cost	self_operated	high	feat: data panel tokens 统计增加 cache_tokens 和 cache_creation_tokens
New API	5064	logging_audit_cost	self_operated	high	模型测试计费漏乘分组倍率：日志显示 default 1.7x，但实际扣费按 1x 计算
Portkey	1663	agent_context_state	mixed	high	Request: List Longbridge MCP on portkey.ai/mcp-servers directory
Portkey	1659	streaming_protocol_parity	community_build_in	medium	Streamed responses fail under concurrent load with "Response object has been garbage collected"
Portkey	1648	provider_adapter_schema	community_build_in	medium	Feature: Add MiniMax as a built-in catalog tile in the Dashboard
Portkey	1640	multimodal_resource	mixed	high	[Feature] OpenRouter Image Generation
Portkey	1639	streaming_protocol_parity	community_build_in	medium	reasoning_effort silently no-ops for Vertex gemini-2.5-flash-lite on /v1/responses
Portkey	1636	security_guardrails	self_operated	high	[Feature] Lakera Guard guardrail plugin
Portkey	1630	multimodal_resource	mixed	high	[Feature] Add ChatGPT Images 2 Support
Portkey	1627	streaming_protocol_parity	community_build_in	medium	Unhandled exception in fireworks-ai streaming requests when stream chunk has empty choices

Project	#	Primary category	Locus	L3	Title
vLLM	32685	multimodal_resource	mixed	high	Responses API supports multimodal inputs
vLLM	34857	agent_context_state	mixed	high	vLLM H1 2026 roadmap / lookahead: MCP, tool calling, parser/renderer and state offload
SGLang	10177	agent_context_state	mixed	high	Responses API lacks custom function tools
SGLang	9681	streaming_protocol_parity	community_build_in	medium	OpenAI tool calling / responses API behavior clarification
llama.cpp	15231	provider_adapter_schema	community_build_in	medium	llama-server tool_calls parameter type breaks OpenAI compatibility
APISIX	13351	security_guardrails	self_operated	high	ai-proxy auth.header re-encrypted on every PATCH, causing 401 from upstream LLM
APISIX	13290	reliability_routing_key	self_operated	medium	add ai-cache plugin for LLM semantic caching
APISIX	13256	streaming_protocol_parity	community_build_in	medium	ai-proxy: streaming SSE loop can starve worker CPU under bursty upstreams
APISIX	13233	security_guardrails	self_operated	high	reference a secret in ai-rag plugin instead of writing the apikey in plain text
APISIX	13213	provider_adapter_schema	community_build_in	medium	about ai-multi-proxy, rewrite the endpoint
APISIX	13150	provider_adapter_schema	community_build_in	medium	dynamically pass parameters from Consumer to an ai-proxy plugin
vLLM	43267	streaming_protocol_parity	community_build_in	medium	Support streaming output for tool_calls arguments

## A6. Interpretation and Limits

The sample supports three cautious observations. First, provider adapter/schema fidelity is the largest class in this corpus, so adapter tests, schema preservation and OpenAI-compatible edge cases should be treated as infrastructure rather than incidental bugs. Second, deployment, streaming parity and agent/tool state are also recurring concerns, suggesting that production gateways need stronger smoke tests and protocol regression checks. Third, multimodal resource and security samples are fewer but have higher L3 relevance because they often imply persistence, permission, audit or policy behavior outside a pure HTTP proxy.

Limitations: the corpus is purposeful, not probabilistic; GitHub issue volume reflects project size, reporting norms, maintainer workflow and search ranking. The appendix should therefore be used for taxonomy-building and product reasoning, not defect-rate estimation.