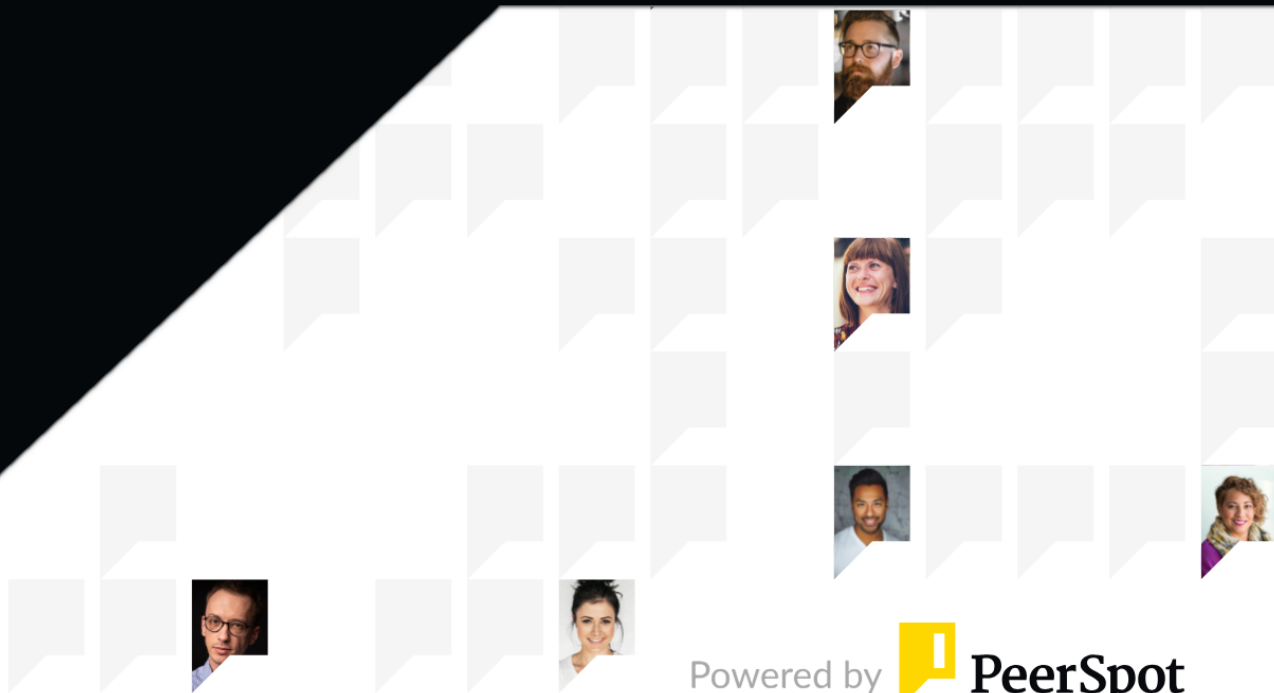




Portkey

Reviews, tips, and advice from real users



Powered by  **PeerSpot**

Contents

Product Recap..... 3 - 4

Valuable Features..... 5 - 12

Other Solutions Considered..... 13 - 14

ROI..... 15

Use Case..... 16 - 20

Setup..... 21 - 22

Customer Service and Support..... 23 - 24

Other Advice..... 25 - 27

Trends..... 28 - 29

About PeerSpot..... 30 - 31

Product Recap



Portkey

Portkey Recap

Portkey offers a robust digital solution catering to modern enterprises with a focus on ease of integration and comprehensive functionality, particularly for data-driven decision-making and process optimization.

Portkey stands out with its highly efficient data integration capabilities, allowing seamless connectivity across multiple systems. Designed for enterprises looking to streamline operations, it facilitates real-time analytics, empowering businesses to make informed choices quickly. It is tailored to support scalability, enabling companies to adapt swiftly to market changes. Known for its flexibility, it integrates easily with existing infrastructures, reducing downtime and enhancing productivity.

What are Portkey's key features?

- **Data Integration:** Connects disparate data sources for a unified view.
- **Scalability:** Adapts to growing demands without major overhauls.
- **Real-Time Analytics:** Provides immediate insights for fast decision-making.
- **Flexibility:** Easily integrates into existing infrastructures.
- **Customization:** Tailors functionalities to specific enterprise needs.

What benefits should users look for in reviews?

- **Increased Efficiency:** Streamlined processes reduce time and resource expenditures.
- **Cost Savings:** Optimized operations lead to lower operational costs.
- **Improved Decision Making:** Access to real-time data enhances strategic planning.
- **Enhanced Flexibility:** Supports diverse operational requirements without major changes.
- **Quick Implementation:** Reduces time to productivity.

Portkey is implemented across industries such as retail, where it helps in managing supply chains efficiently, and finance, where it provides robust analytical tools for real-time market analysis. In healthcare, it ensures data integration across platforms, contributing to improved patient care and operational efficiency.

Valuable Features

Excerpts from real customer reviews on PeerSpot:



“Overall, Portkey has had a huge business impact in terms of cost savings and operational efficiency.”



Shravan Revanna

Software Engineer at klydo.in



“Portkey has significantly improved the organization by streamlining the AI processes being followed.”



Singh Aman

Generative AI Engineer at Tata Consultancy



“Portkey is totally stable; it is one of the best software solutions in the market for this purpose.”



Akashkhurana Hirana

Senior Software Engineer 2 at Porch

- ✓ “Portkey's visualization is excellent, I particularly appreciate the multiple capabilities they have introduced, especially the MCP gateway, which is outstanding, the caching mechanism is very good, and the logs and traceability features are also excellent.”



Verified user

Gen Ai Architect at a tech vendor with 10,001+ employees

- ✓ “Portkey definitely provides a solid alternative solution for the agent and large model hosting platforms, and it is very helpful for us to explore the possibilities across the industry rather than staying with a few mainstream options.”



Support Pan

project manager and delivery owner

What users had to say about valuable features:

“Portkey provides the standard interfaces for the major features that interested the industry has. It has a very comprehensive feature support for the MCP agent and its playground, support for prompt templates, and allows the management of traceability analytics. It covers most agentic application usage very comprehensively.

Portkey definitely provides a solid alternative solution for the agent and large model hosting platforms, and it is very helpful for us to explore the possibilities across the industry rather than staying with a few mainstream options.

Portkey successfully provides a standard interface and large model based solutions, and we can see those features in many other platforms as well. This is solid, but so far I haven't found anything super exciting or very innovative on this platform. For our organization, it's a great alternative solution to consider..”

Support Pan

project manager and delivery owner

[Read full review](#) 

“The most valuable features are the unified AI gateway, observability, price calculations, routing controls, and guardrails. The universal API makes it easier to switch between and compare models without rewriting large parts of the application. For the organization, Portkey was providing two endpoints, one for the US and another for Europe, which was easily manageable. Some models were supported for the US, some not, some for the EU, and some for both. Switching between providers is very easy with Portkey. The fallback and retry capabilities are also useful because production LLM applications need resilience when a provider is slow, rate-limited, or temporarily unavailable. Cost and latency tracking are valuable because LLM usage can become difficult to manage across teams, but these can be easily managed through Portkey.

“Observability is the best feature because at the production grid application, the focus is usually on observability, such as how the end user is using it, what the latency is, how many errors occurred within time frames, which models were used, and how many tokens were consumed. These aspects are easy to manage in Portkey.

“Portkey has significantly improved the organization by streamlining the AI processes being followed. Portkey improves the workflow for centralizing LLM operations instead of every application team building its own logging, routing, fallback, and cost tracking logic. Portkey provides a shared layer for these capabilities so it can be synced between all teams. It helps reduce engineering efforts when experimenting with different providers and models. It also improves visibility into production behavior because request volumes, latency, cost, errors, and model usage can be seen in one place. For teams building multiple GenAI applications, the biggest improvement is standardization. Portkey makes it easier to enforce common practices across observability, reliability, and governance..”

Singh Aman

Generative AI Engineer at Tata Consultancy

[Read full review](#) 

“We have an application that uses LLM models such as Gemini, OpenAI models, or Claude models. We need to interact with these models, which we have deployed elsewhere, and we use Portkey as the AI/LLM gateway to check observability and monitor how our model is behaving. We also have custom fine-tuned trained models, and we want to check the observability or test these models using Portkey.

We primarily use three features: the AI Gateway, observability, and the guardrails.

The AI gateway and observability are the two features that are very important for the team. The AI gateway acts as a single point of application for all the AI model requests. It also enhances security since we do not need to provide the API key to all the developers. Additionally, it supports load balancing and automatic retries. When a lot of load is coming in, it balances the load among different models. If one of the models is not available, it handles the failover by redirecting the request to the other model. Regarding observability, it is very useful to monitor and understand how our AI application is behaving or performing. It usually logs all the requests and the responses. We can easily get an idea of the token cost, token usage, API cost, and error rates. It makes debugging easier and tracks the AI spending to show how much we have spent. These are the valuable aspects of both features.

Another feature of the AI Gateway is that it lets you switch between providers without changing your application code. Suppose we are using three models for three different purposes and one model is not available, or I want to switch the model within the application without restarting my code or doing a redeployment. We can easily do this with Portkey. Regarding observability, we usually monitor and check how our AI applications are performing and behaving. Other things that we usually track in observability include token usage, API cost, error rates, success and failure metrics, and response latency. This is an important aspect that we usually prioritize.

Guardrails are safety checks and quality control that we apply to the input and output of AI. Suppose I do not want certain personally identifiable information or personal information data of a user to be sent to the AI. Portkey has guardrails that

do not allow sending that data to the LLM gateway. It can also act as a middleware that checks all the data we are sending to the LLM before transmission, so we do not send any personally identifiable information or personal information of the user to the LLM gateway..”

Akashkhurana Hirana

Senior Software Engineer 2 at Porch

[Read full review](#) 

“The best feature that Portkey offers is a centralized AI gateway, which is the biggest advantage. Instead of integrating directly with different LLM providers, Portkey gives us a single, consistent interface. This results in a simpler architecture and has reduced maintenance. Another valuable feature is multi-model and multi-provider support, which allows us to work with different providers through the same integration. Observability and logging represent one of the most valuable features for production. We gain visibility into request latency, token consumption, errors, and success rates. Cost monitoring is critical since LLM costs can increase quickly, and we need centralized visibility into token usage and API consumption. Reliability features, including fallback mechanisms and routing to different LLMs when there is an issue with the provider, allow us to recover gracefully.

“Without question, observability is followed by the AI gateway itself. Whenever we are developing a new AI feature or investigating an issue, the first place we look is the logs and dashboards. We use it for monitoring request latencies and response times, tracking token usage, and identifying timeout requests. We compare how different models perform for the same use cases. Debugging prompts becomes seamless and easy. For example, when our customer support copilots started responding more slowly, we could quickly determine the bottleneck, whether it was the LLM provider or the application itself. Without a centralized observability system, we would have to piece that information together from multiple systems. Portkey itself is something we benefit from continuously, but it is largely transparent once configured.

“Some specific details that stand out include how it is vendor-agnostic. Since AI evolves very quickly and models are releasing every few weeks with frequent pricing changes and different performance characteristics, Portkey gives us the flexibility to evaluate and adapt to newer models without having to redesign our application architecture. Another valuable detail is how it encourages good engineering practices. Since every AI request goes through a centralized layer, it is much easier and standardized. If I had to summarize, Portkey gives engineering teams the flexibility to innovate with AI while maintaining the operational costs, control, visibility, and consistency needed to run AI applications at scale.

“There are specific outcomes and changes I have noticed since implementing Portkey. The biggest impact is operational efficiency and maintainability. Before Portkey, each AI service managed its own model integration. After adopting Portkey, we centralized those responsibilities into a single gateway. This reduced duplicate engineering efforts and gave every team a consistent way of consuming LLM services. The outcomes we noticed include faster development, quicker troubleshooting, better cost awareness, greater flexibility, and improved reliability..”

Shravan Revanna

Software Engineer at klydo.in

[Read full review](#) 

Other Solutions Considered

“Nothing has been purchased from the AWS marketplace. Roche is a partner for Portkey and the organization is an enterprise for Portkey, utilizing Portkey features..”

Singh Aman

Generative AI Engineer at Tata Consultancy

[Read full review](#) 

“We have evaluated several alternatives. Kong AI gateway is one alternative we have considered. Apigee from Google is another alternative we are aware of. Additionally, we are expecting another alternative from AWS called Gateway, which is expected to be released by Q4..”

Verified user

Gen Ai Architect at a tech vendor with 10,001+ employees

[Read full review](#) 

“Previous solutions were being used, but they were not considered optimal. AWS Marketplace is also being used for some LLM models, but AWS does not support all the models. Portkey is a better solution because it has everything in one place..”

Singh Aman

Generative AI Engineer at Tata Consultancy

[Read full review](#) 

“Previously, I did not have any dedicated gateway. We were integrating directly through SDKs and APIs. While that worked initially, it became difficult to manage. The main challenge was that comparing different LLMs required a lot of code changes, and there was no visibility on token usage. Troubleshooting was also challenging. I evaluated a few alternatives, but I chose Portkey because it has all the capabilities that I needed. It allowed me to standardize my infrastructure. The switch was not from commercial gateways; it was from direct integration to a centralized integration..”

Shravan Revanna

Software Engineer at klydo.in

[Read full review](#) 

ROI

Real user quotes about their ROI:

“Time has been saved for multiple developers because of having a single package to manage multiple models and multiple providers. This was the main ease gained..”

Singh Aman

Generative AI Engineer at Tata Consultancy

[Read full review](#) 

“I have seen a return on investment with Portkey. I can share relevant metrics: thirty to forty percent reduction in development time, fifty percent faster evaluation cycles, twenty to twenty-five percent reduction in unnecessary LLM spending, and twenty to twenty-five percent improvement in engineering productivity..”

Shravan Revanna

Software Engineer at klydo.in

[Read full review](#) 

Use Case

“Our team is searching for an alternative to LiteLLM, which is another monitoring management platform, and we found that Portkey is an alternative solution that we decided to try out.

We want to conduct prompt template testing with Portkey by splitting our business logic along with the prompt, so we need to test that against different models and perform evaluation accordingly. This is one major thing we are trying. The product owner can separately conduct their test for the prompt template on the playground.

There could be some advanced usage with Portkey such as setting up the MCP and guard rail. We are thinking about having that part as well, but so far we haven't touched it yet..”

Support Pan

project manager and delivery owner

[Read full review](#) 

“Our client is completely using AWS products. For AI gateways, we are using Portkey, which is the product we have implemented, and we are leveraging all AWS services for internal application development.

“Applications can interact with different multiple LLM models. We are restricting direct access from the application to LLMs. Instead, we are introducing Portkey as an AI gateway in between. We send the application's requests to the respective LLMs through Portkey, and we have enabled token usage, caching mechanisms, and guardrails at the AI gateway level. All applications must send requests via Portkey to leverage LLM capabilities..”

Verified user

[Read full review](#) 

Gen Ai Architect at a tech vendor with 10,001+ employees

“Portkey has been used as an AI gateway and observability layer for LLM applications. The main use case is to route requests across multiple LLM providers through a common API, monitor latency, cost tracking, prompts, and responses, and add reliability features such as fallback, retries, caching, and guardrails. It is especially useful when moving from experimentation to production because it provides a single control layer for model routing, prompt management, usage tracking, and operational visibility.

“Whenever an API is called through Portkey, it displays the total time consumption that has been taken. Observability is provided there. When it fails to connect to an LLM model, it attempts to use a fallback mechanism, which is helpful..”

Singh Aman

Generative AI Engineer at Tata Consultancy

[Read full review](#) 

“Portkey has many features for deploying an application to production, including an AI gateway, observability, guardrails, governance, and prompt management. We primarily use the AI Gateway, observability, and the guardrails.

We have an application that uses LLM models such as Gemini, OpenAI models, or Claude models. We need to interact with these models, which we have deployed elsewhere, and we use Portkey as the AI/LLM gateway to check observability and monitor how our model is behaving. We also have custom fine-tuned trained models, and we want to check the observability or test these models using Portkey. In this scenario, when any person wants to call the LLM or use the AI, they do not need the API key for that model. They can directly call Portkey and it will provide the response. The API keys for all the models are not exposed to the developers or the engineers. They are stored in the vault and can only be accessed by Portkey..”

Akashkhurana Hirana

Senior Software Engineer 2 at Porch

[Read full review](#) 

“Portkey serves as a centralized AI gateway for our organization. At Clydo, we are building multiple AI-powered features, including internal copilots, catalog intelligence, customer support assistance, and agentic workflows. This gives us a single layer for model routing, prompt management, and API key management.

“A primary example of our main use case is our internal AI assistant used by our operations and customer support team. Every user query goes through Portkey before reaching the language model. Portkey handles authentication and routes requests to appropriate models. One practical scenario involves our engineering team wanting to evaluate a newer model for better reasoning at a lower cost. Instead of changing code across multiple services, we update the routing configuration in Portkey and compare the responses and performance. Another benefit is observability. When users report that AI responses are slower or inconsistent, we use Portkey's logs to see the request latency. Overall, Portkey has become the control layer for our LLM infrastructure. It allows our engineers to focus on building AI features while providing us with centralized visibility, governance, and flexibility.

“The main use case is standardizing how AI is consumed across the organization. As more teams start building AI-powered features, we do not want every service to implement its own integration with retry logic or rate limiting, which becomes very difficult to manage. Portkey gives us a common layer that every AI application uses, whether it is an internal or customer-facing chatbot or any agentic flows. This makes the architecture much cleaner and easier to manage. It gives us the flexibility to experiment. We can compare different models and run A/B evaluations for specific workloads. From an engineering perspective, Portkey has evolved from being just an API gateway to becoming part of our platform. It is always a layer that provides consistency, observability, and governance across all our LLM-based applications..”

Shravan Revanna

Software Engineer at klydo.in

[Read full review](#) 

Setup

The setup process involves configuring and preparing the product or service for use, which may include tasks such as installation, account creation, initial configuration, and troubleshooting any issues that may arise. Below you can find real user quotes about the setup process.

“I would describe the learning curve for new users or teams adopting Portkey as seamless and very easy to adapt. Knowledge transfer becomes increasingly easier as team members become familiar with the platform..”

Shravan Revanna

Software Engineer at klydo.in

[Read full review](#) 

“The best aspect in terms of cost saving is using Portkey cache, which was built-in to the packages. This literally saved token costs because cache tokens cost less to process. This was the best aspect in terms of productivity. The setup was also smooth because only a single Portkey library is needed, and then any models can be used at once. Whether using AWS Marketplace models, Google models, or OpenAI models, everything is available in one place. There is no need to worry about multiple platforms or multiple packages..”

Singh Aman

Generative AI Engineer at Tata Consultancy

[Read full review](#) 

“The initial deployment of Portkey was straightforward. For a SaaS product, you can directly use it. For the SaaS version, registration with an email ID and purchase can be completed in approximately one minute.

“If deployed as a COTS product, the initial setup takes more time because the entire product needs to be installed into a Docker machine. Once that installation is available, the initial deployment takes time. After deployment, patching work takes approximately four hours..”

Verified user

[Read full review](#) 

Gen Ai Architect at a tech vendor with 10,001+ employees

Customer Service and Support

“Portkey's customer service is good. I remember when I was just setting up Portkey, I encountered a small issue. I tried to connect with customer support, raised a ticket, and they replied back within a few hours..”

Akashkhurana Hirana

Senior Software Engineer 2 at Porch

[Read full review](#) 

“For enterprise production, support quality is very important. Support was reached out to once on Discord. There is a support group on Discord where the team is very responsive. They provide a newsletter with updates and features. The support experience has been good..”

Singh Aman

Generative AI Engineer at Tata Consultancy

[Read full review](#) 

“We have scheduled support calls with Portkey every week. The support quality and responsiveness are good overall.

“We have received immediate responses when needed, and we have established a regular weekly call schedule. However, email communication has been somewhat lagging, and responses are not always provided within the proposed timeframes. Apart from email communication delays, the remaining support aspects are satisfactory..”

Verified user

[Read full review](#) 

Gen Ai Architect at a tech vendor with 10,001+ employees

“I have not reached out for customer support very often. Overall, the experience with customer support has been positive. Portkey's team has been responsive, technically knowledgeable, and understood the challenges I faced. Most of my questions were around configuring multiple provider routing and optimizing gateway settings. As the platform has matured, I have needed to contact them less frequently because of improved documentation. If I had one suggestion, it would be to expand the self-service knowledge base with more implementation guides. Overall, the customer experience rating is solid at nine out of ten..”

Shravan Revanna

[Read full review](#) 

Software Engineer at klydo.in

Other Advice

“It is definitely a private cloud for Portkey from their provider. Using their online services with Portkey is good. If you are seeking an alternative solution for the product, Portkey is definitely worth a look. My overall review rating for this product is 7..”

Support Pan

project manager and delivery owner

[Read full review](#) 

“Everything is perfect except for the documentation and the steep learning curve that I already mentioned. That is the part for which I deducted one mark. Everything else is perfect. Portkey is a very useful software if you are going forward with your AI application that uses the LLM gateway. I would recommend you use it and you will love it within a few days..”

Akashkhurana Hirana

Senior Software Engineer 2 at Porch

[Read full review](#) 

“Portkey should be recommended for teams that are moving GenAI applications into production and need more control over reliability, cost, observability, and governance. It is especially useful if multiple LLM providers are being used and if changes to models are expected over time. The overall rating given for Portkey is nine out of ten..”

Singh Aman

Generative AI Engineer at Tata Consultancy

[Read full review](#) 

“Regarding governance and security, I would say security is quite high. In fact, one of the reasons we chose to implement Portkey is that it gives us a central control point for AI usage rather than every application managing it independently. From a security perspective, I appreciate that sensitive provider credentials are managed centrally instead of being distributed across multiple services. On the governance side, having a single layer for routing, usage tracking, and policy enforcement helps us standardize how AI is used across different teams. As more applications adopt LLMs, consistency becomes increasingly valuable. Overall, I would describe Portkey as a strong differentiator that provides visibility and centralized control.

“Portkey is deployed in my organization as a hybrid cloud deployment. We use [AWS](#) as part of our hybrid cloud setup. Our applications and backend services run on [AWS](#), and all AI requests are routed through Portkey before reaching the underlying models. We have backend services and API databases on AWS. Portkey centralizes routing, observability, and cost management. This approach gives us the flexibility of managing foundation models while keeping our application logic simpler. Operational controls are within our own cloud environment, allowing us to change model providers without impacting the applications.

“In terms of customizing, I would say Portkey is very flexible because one of the reasons we adopted it is that it fits into the existing architecture. Most of my customization has been through configuring rather than custom development. For

example, I customized different routing rules for different AI workloads. I have also integrated Portkey into my internal monitoring and engineering workflows. The best part is that our application code remains largely provider-agnostic. If we want to evaluate a new LLM or change routing for a specific use case, we typically do not need to make significant changes. I have not had to build major custom extensions to make Portkey work for us. Most of what I have needed has been supported through existing configurations. Overall, customizability is highly rated.

“For my organization, compliance is important, though we are not heavily regulated. From my perspective, Portkey contributes to a centralized control layer. Since all LLM traffic passes through a single gateway, it is much easier. That said, we do not rely on Portkey alone to meet compliance requirements. We combine it with our own security controls.

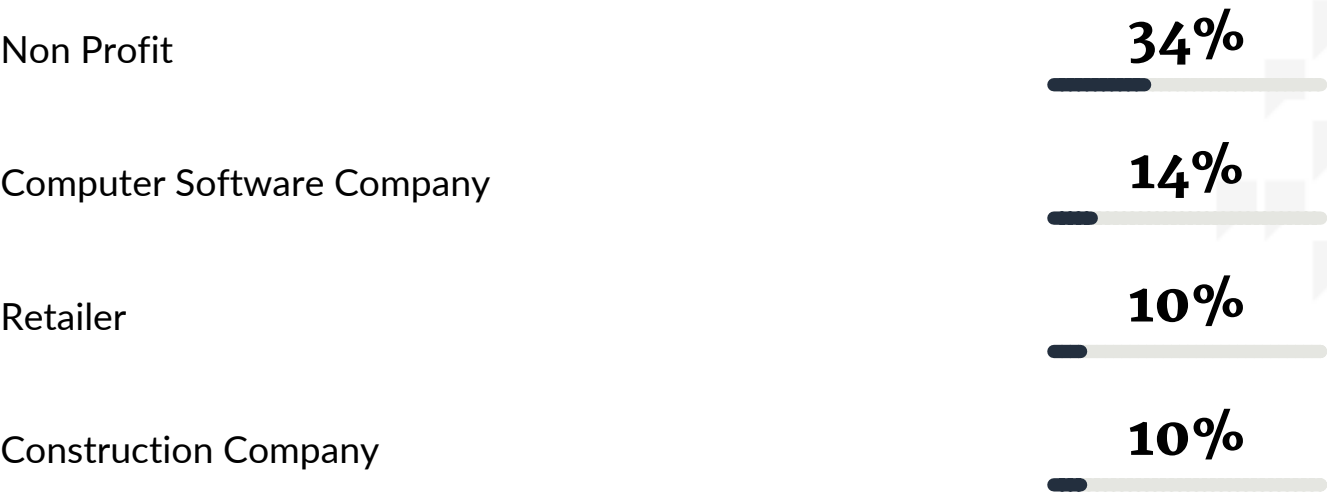
“I can share some approximate figures for faster development and better cost awareness. For development, I estimate that Portkey reduced engineering efforts by around thirty to forty percent. Previously, each team had to implement provider-specific authentication, retries, logging, and error handling. For model evaluation, we have seen about fifty percent reduction in time. From a cost management perspective, we have reduced unnecessary LLM consumption by around fifteen to twenty percent. I have also noticed thirty to forty percent reduction in troubleshooting time. Overall, Portkey has had a huge business impact in terms of cost savings and operational efficiency. I would rate this review overall at nine out of ten..”

Shravan Revanna
Software Engineer at klydo.in

[Read full review](#) 

Top Industries

by visitors reading reviews



About this buyer's guide

Thanks for downloading this PeerSpot report.

The summaries, overviews and recaps in this report are all based on real user feedback and reviews collected by PeerSpot's team. Every reviewer on PeerSpot has been authenticated with our triple authentication process. This is done to ensure that every review provided is an unbiased review from a real user.

Get a custom version of this report... Personalized for you!

Please note that this is a generic report based on reviews and opinions from the collective PeerSpot community. We offer a [customized report](#) of solutions recommended for you based on:

- Your industry
- Company size
- Which solutions you're already considering

The customized report will include recommendations for you based on what other people like you are using and researching.

Answer a few questions in our short wizard to get your customized report.

[Get your personalized report here](#)

About PeerSpot

PeerSpot is the leading review site for cloud, AI, and business software. We created PeerSpot to provide a trusted platform to share information about software, applications, and services. Since 2012, over 22 million people have used PeerSpot to choose the right software for their business.

PeerSpot helps tech professionals by providing:

- A list of products recommended by real users
- In-depth reviews, including pros and cons
- Specific information to help you choose the best vendor for your needs

Use PeerSpot to:

- Read and post reviews of products
- Access over 30,000 buyer's guides and comparison reports
- Request or share information about functionality, quality, and pricing

Join PeerSpot to connect with peers to help you:

- Get immediate answers to questions
- Validate vendor claims
- Exchange tips for getting the best deals with vendor

Visit PeerSpot: www.peerspot.com

PeerSpot

244 5th Avenue, Suite R-230 • New York, NY 10001

reports@peerspot.com

+1 646.328.1944