CASE STUDY: SMARTLY

Streamlining communications for technical support teams

Llama-powered generative AI agents automate support ticketing and generate resolution messages

Meta SMARTLY

At a glance

Smartly used Llama to create AI agents that automate ticket generation and generate resolution messages for its Support Operations team. The agents retrieve information, summarize complex technical issues and create professional, on-brand customer communications.

Industry
Advertising

Use case
Automating customer support ticketing and communications

Goal
Enable better and faster customer support for a growing customer base

Llama version
Llama 3.1 8B

Deployment
On-premises Kubernetes

Results*

80%
less time to create support tickets

50%
less time to draft customer resolution messages

*All results are self-reported and not identifiably repeatable. Generally, expected individual results will differ.

The challenge

Manual ticketing, complex communications

The Smartly Support Operations team works behind the scenes to provide world-class technical and customer support. In addition to serving customers, the team managed communications, including manually creating tickets and composing resolution messages.

The ticketing process involved multiple platforms, causing duplicated efforts and constant context switching. Crafting resolution messages required a thorough analysis of engineering notes posted across multiple communications platforms plus a review of ticket histories and details. Resolution messages also demanded careful attention to ensure that highly technical issues were clearly explained to the customer.

As the Smartly platform grew, the combination of manual work and pressure from scaling demand became unmanageable, driving the team to look for ways to reduce workloads and save time with AI.

Better, faster, secure communication with less effort

The Smartly Support Operations team wanted to use generative AI to automate ticket generation and create resolution messages including populating multiple platforms, retrieving engineering and support notes, and summarizing complex technical issues. Once a customer's issues were resolved, the AI application would then have to generate customer resolution messages that maintained a professional, on-brand tone while clearly explaining technical topics.

SMARTLY

Smartly is an AI-powered advertising technology company that manages nearly five billion dollars in ad spending for more than 700 brands worldwide. Its platform seamlessly integrates the capabilities of media, creative and intelligence to reach and engage consumers and learn what performs best.

- **Industry:** Advertising
- **Company size:** 650+ employees in 13 countries

"We needed a model that excelled in summarization and had flexible deployment options. Our main priority was to find a solution that could integrate easily with our infrastructure, ensuring customer data remained secure on our servers while minimizing resource usage."

Tatiana Kuumola
Support Solutions Engineer, Smartly

The solution

Llama-powered AI agents for automated ticketing and communications

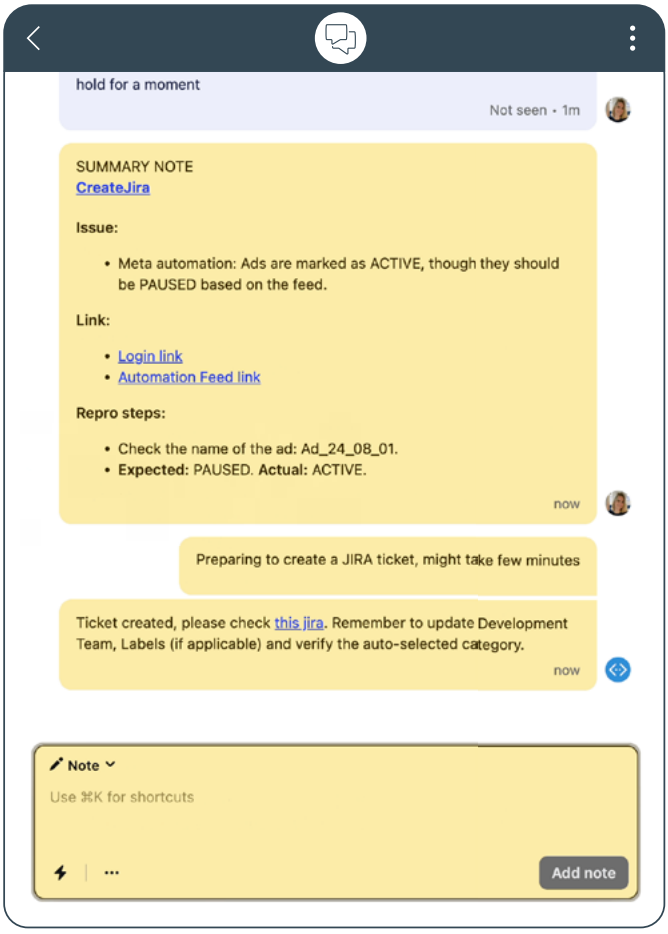
After testing several competing models for the application, Smartly landed on Llama, which showed promising performance, flexibility and ease of integration.

“Llama 3.1’s token capacity, ability to handle complex summarization tasks and seamless integration with our infrastructure convinced us to continue with Llama without testing other models,” says Erik Karsten, DevOps Engineer at Smartly.

With the release of Llama 3.1 8B, the Smartly team had a production-ready candidate. Llama 3.1 8B required no training or fine-tuning to create tickets and generate customer messages. They used basic prompt engineering and few-shot learning to provide clear instructions, including formatting examples (such as ticket titles and descriptions), subject matter context and writing style.

Open-source Llama checked all the boxes

Smartly deployed Llama on its private, on-premises hosting environment within a Kubernetes-based orchestration platform. Deploying locally kept them in complete control of their data and their customers’ information. The model and prompt template integrated seamlessly with their existing Kubernetes platform so it could be easily accessed by other services and applications. This setup drastically improved performance and reduced manual workloads.



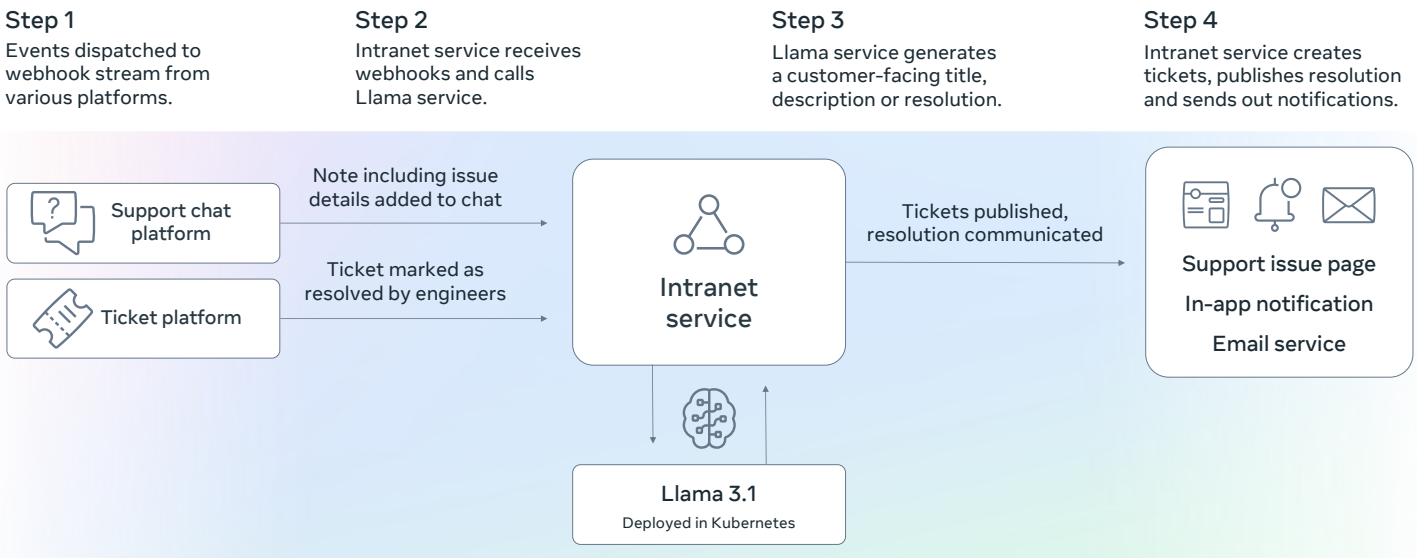
The agent ingests customer issues and generates trouble tickets automatically.

> “As an open-source platform, Llama gives us the flexibility to run on our private infrastructure and stay competitive without the high costs of proprietary software. With Llama, organizations like ours can innovate rapidly, create custom solutions and adapt quickly.”
>
> Erik Karsten
> DevOps Engineer, Smartly

As the project moved from pilot to production, the Llama open-source community and knowledge base helped the Smartly team work through challenges, including optimizing the model’s performance on CPUs and GPUs and tuning their infrastructure to improve throughput and scale efficiently.

“The open-source approach allowed us to stay updated with the latest developments, including switching from Llama 2 to Llama 3 mid-project,” says Tatiana Kuumola, Support Solutions Engineer. “It also helped us find answers to running Llama on both CPU and GPU, optimizing our infrastructure for better performance.”

Generative AI agent workflow



Smartly used Llama in a private, local deployment that automates ticketing and resolution messages.

The outcome

Automated ticketing and resolution messaging save time and streamline operations

The results have exceeded expectations. The Llama 3.1-powered agent fully automated support ticket creation, including customer communication. Smartly estimates Llama has reduced time spent creating tickets by up to 80%.

When an issue is resolved, the AI agent drafts a resolution message based on the standards embedded in the prompt template. Now, all resolution messages deliver consistent information in a uniform, professional tone and take 50% less time to create.

Reactions to the new service have been positive. Customer satisfaction scores for technical support are up, and Smartly support agents are happier. By taking on repetitive, meticulous tasks, the Llama-powered AI agents have given the support team time back to help customers one-on-one.

80%
less time to create support tickets

50%
less time to draft customer resolution messages

> “In addition to time savings and increased customer satisfaction, we’ve seen a notable boost in our support agents’ satisfaction. Our team is more motivated and confident, strengthening our ability to deliver exceptional customer support.”

Tatiana Kuumola
Support Solutions Engineer, Smartly

Conclusion

A robust outlook for new Llama-based solutions

Ticket and resolution automation are only the first of many generative AI applications for Smartly. The Smartly Operations Support team plans to develop Llama-based solutions for ticket categorization, product support analytics and error message creation.

How can Llama help your business?

See how open-source Llama brings unmatched control, customization and flexibility to generative AI application development and deployment.

Learn More

Related Stories ▸

© 2024 META

Meta