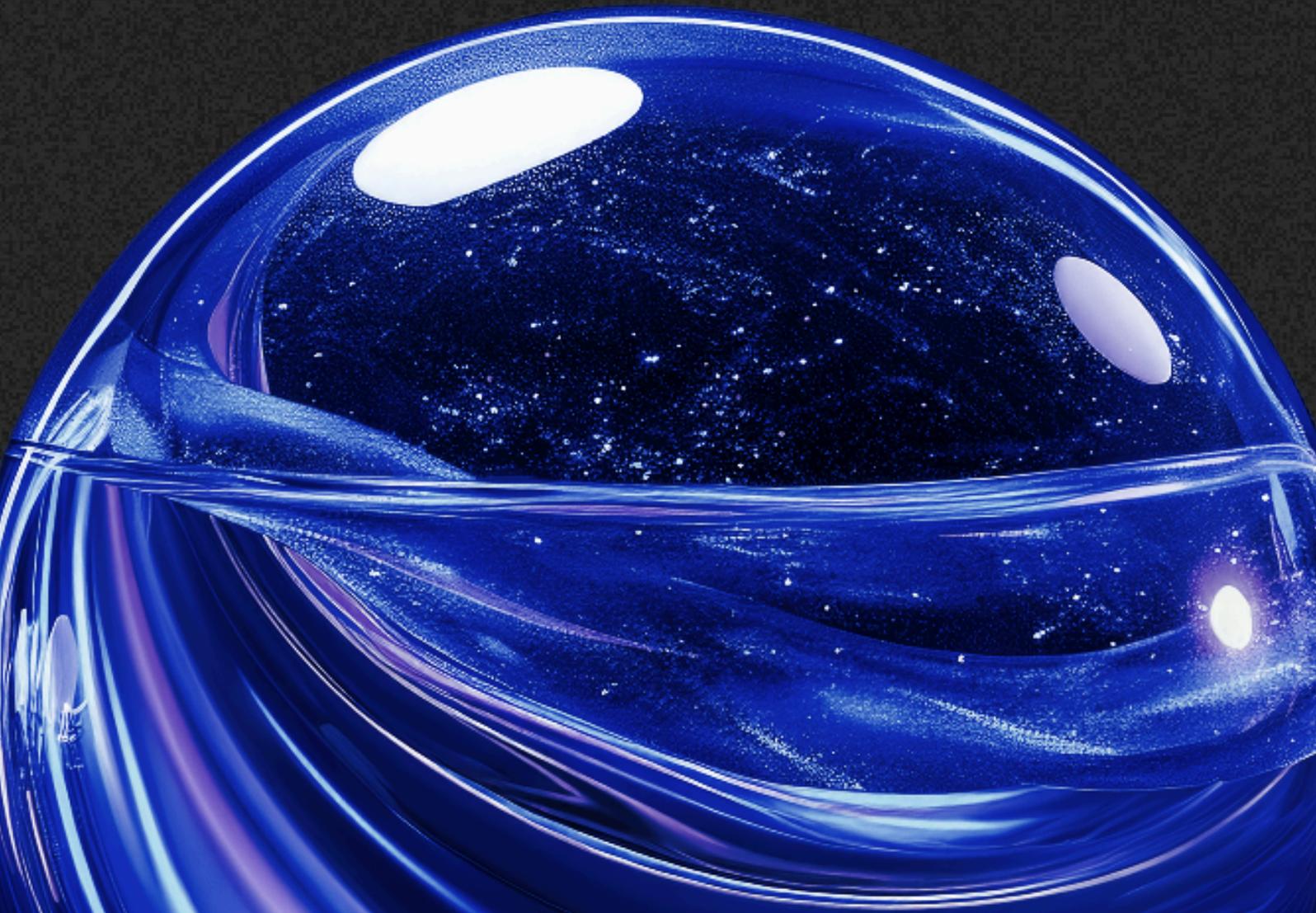




case study

Generative AI Article Summarization for MIT Technology Review



Executive Summary

MIT Technology Review is a leading global technology publication that reaches millions of readers with in-depth analysis and reporting on innovation, science, and technology. Founded in 1899, the publication has been at the forefront of technology journalism for over 120 years. Their editorial platform is powered by WordPress, hosted on a third-party cloud provider.

As part of their strategy to experiment with AI and improve reader engagement, **MIT Technology Review wanted to integrate Generative AI into their CMS, specifically to summarize long-form articles into concise, digestible formats.** Following successful production deployment, MIT Technology Review and Flatiron Software decided to **open-source the solution to benefit the broader media industry.**

The Challenge

MIT Technology Review needed a solution that would enhance their editorial capabilities without disrupting existing workflows or compromising editorial control. The requirements included:

Key Client Needs

- Enhance reader engagement by offering article summaries in widget form
- Experiment safely with GenAI without rebuilding their CMS stack
- Allow editorial teams to choose their preferred LLM for experimentation
- Build a scalable and secure bridge between WordPress and AWS AI/ML services

Technical Constraints

- Enhance reader engagement by offering article summaries in widget form
- Experiment safely with GenAI without rebuilding their CMS stack
- Allow editorial teams to choose their preferred LLM for experimentation
- Build a scalable and secure bridge between WordPress and AWS AI/ML services

Infrastructure Decision

- Amazon Bedrock - Multi-provider API access without infrastructure management, enabling model experimentation without vendor lock-in
- Security & Compliance - Protected sensitive editorial workflows throughout the AI pipeline
- Flexibility - Model selection and switching without code changes
- Scalability - Serverless architecture aligned with publishing schedules, no always-on infrastructure

The Solution

We designed and delivered a custom WordPress plugin that integrates with Amazon Bedrock for AI-powered summarization. The solution prioritizes editorial control while providing seamless access to multiple AI models.

How It Works

The workflow is straightforward and keeps editors in control at every step:

- 01** Editors can select any article in WordPress and trigger "Generate Summary" from the plugin panel
- 02** The plugin connects securely to Amazon Bedrock via API Gateway + Lambda
- 03** Bedrock invokes the chosen foundation model (customer-configurable: Amazon Titan, Anthropic Claude, or others)
- 04** The model produces a short, human-readable summary, which is returned to WordPress
- 05** The summary is displayed as a widget alongside the article (or embedded inline), configurable by the editorial team

Architecture

The technical architecture follows a serverless pattern that prioritizes security, scalability, and simplicity:

- **Amazon API Gateway:** Endpoint for WordPress plugin requests
- **AWS Lambda:** Serverless function to handle requests, call Bedrock, and process summaries
- **Amazon Bedrock:** Foundation model inference layer (Titan, Claude, or other supported models)
- **Amazon CloudWatch:** Monitors requests, usage, and logs
- **IAM Roles & Policies:** Secure integration and least-privilege access

Results & Benefits

After deployment, the system demonstrated measurable impact across editorial workflows and reader engagement:

- **Editorial Empowerment:** Journalists can now generate summaries on demand, without leaving WordPress.
- **Reader Engagement:** Articles now feature concise AI-generated summaries, improving time-on-page and mobile user experience.
- **Experimentation with Choice:** Editors can test multiple foundation models to determine tone, style, and performance fit.
- **Non-Disruptive Integration:** The plugin works seamlessly with the existing WordPress stack, regardless of hosting provider.
- **AWS Adoption:** Introduced AWS Bedrock as the AI backbone, creating opportunities for future GenAI use cases (translations, Q&A, personalization).

Open Source Release

Following successful validation in production, MIT Technology Review and Flatiron Software made the decision to open-source the WordPress plugin and AWS infrastructure code. Rather than commercializing the solution, both organizations chose to make it freely available to the media industry.

Why Open Source?

The decision was based on a shared philosophy: quality AI tooling should be accessible to publishers of all sizes, not locked behind enterprise licensing fees. Media organizations are experimenting with AI but often lack the resources to build production-grade tools from scratch. By open-sourcing a solution that has been proven in a demanding editorial environment, other publishers can benefit without repeating the same development and testing cycle.

Why Flatiron?

Flatiron delivers complex AI transformation projects at enterprise scale—fast, independently, and with no disruption to your existing teams. Our model lets you capture the value of AI while protecting your core business and product roadmap. From project management to technical delivery, we bring deep expertise, proven process, and clear accountability.

Ready to see what AI can do for your company?

Let's talk.

✉ marketing@flatiron.software

🌐 flatiron.software

📍 Miami, US