

# Fandom Improves Community Safety and Reduces Cost by 50% with Coactive AI

Learn how the world's largest fan platform uses AI to automate the removal of offensive content, taking removal time from 24 hours to mere seconds.

## Opportunity | From Westeros to Ponyville: Accepting the Quest to Improve the Fan Experience

Fandom is the world's largest fan platform, reaching 350 million fans worldwide each month. From Game of Thrones to My Little Pony, Fandom's mission is to power fan experiences by providing fans with the largest online library of information on anything and everything in entertainment, gaming, and pop culture. There are three unique types of clients that Fandom serves, each with different experiences related to the platform:

- Users seeking information
- Superfan editors looking to create content
- Advertisers trying to appeal to fans

Hundreds of millions of visitors engage with Fandom every month—which means hundreds of millions of possibilities that those users might upload new content to enhance the pages that interest them. While most images are uploaded with good intentions, about 0.5 percent of the 2.2 million images uploaded monthly are malicious and violate Fandom's terms of service. Given the significant scale, this small percentage results in high costs for manually reviewing each image.

With tens of millions of images uploaded per year, it became challenging for the Fandom Trust and Safety team to personally moderate each image. Fandom hired contractors to assist with the manual image moderation process, which took about 500 hours per week—a costly expense added to an already taxing process.



## About Fandom

Fandom is the world's largest fan platform, immersing over 350 million monthly visitors in entertainment and gaming. Hosting more than 300,000 wikis, Fandom is the #1 source for in-depth information on pop culture, gaming, TV, and film, where fans learn and celebrate their favorite fandoms. Its Gaming division manages the online video game retailer Fanatical. Fandom Productions enhances the fan experience with curated editorial coverage and branded content from Gamespot, TV Guide, Metacritic, and Emmy-nominated Honest Trailers.

## Benefits

- 50% cost savings
- 500 manual labeling hours per week reduced by 74%
- 2.2 million image uploads per month, with 90% automatically reviewed
- Solution delivered in 6 weeks

The team needed a solution that would automate visual content moderation, to not only ensure Fandom's continued user trust, but also community safety advertising opportunities.

## **Solution | Managing Mischief and Combatting Trolls with Coactive AI**

The Fandom team had heard about Coactive AI and did a deep dive to learn more about its capabilities. Traditional search tools are limited to understanding video and image content based on the metadata provided, whereas Coactive uses multimodal AI to understand the actual content of a visual asset in the same way a human would. Fandom kicked off a proof of concept (POC) with Coactive to see just how good multimodal AI was at understanding the nuances of its extensive visual library.

"We started the POC process in early 2023. The Coactive team delivered us a solution in just six weeks. We saw value quickly," said Florent Blachot, VP of Data at Fandom. "We recognized that the return on investment for this platform was very good and that the potential for the future is not just focusing on user safety, but far beyond that."

For the POC, the Fandom team curated about 20 million problematic images and categorized them into 25 nuanced labels, such as "gore," "nudity," or "pornography." From there, Coactive used a foundation model that recognizes offensive imagery combined with fine-tuning to further match labels to the images that need to be removed. Each image is given a score between 1 and 100 for how well it fits into its label:

- 90 and above—automatically deemed inappropriate
- 10 to 90—manual review by content moderation team
- Below 10—image is allowed on the site

"In the new solution, our image-review microservice calls one of Coactive's APIs for AI classification against a set of Fandom-defined concepts. We're able to set these thresholds manually via our custom UI, which we can adjust as needed," explained Marcin Kostrzewski, Senior Software Engineer at Fandom.

The POC was a success, showing that Coactive AI was able to accurately categorize content and effectively flag what was acceptable, in violation of community terms, or in need of human review. It operationalized the process for end-to-end image moderation.

## Outcome | Empowering Fan Communities and Refining the Process Along the Way

With the new image scoring process in place, image removal is both automatic and near instant. “On average, this process takes less than half a second. It’s significantly faster than the previous system, where some images could remain on the page for 24 to 36 hours before removal. And it elastically scales with traffic and seasonality,” said Blachot.

Automating image moderation has not only saved the Fandom team time, but also improved morale—they no longer spend hours sifting through disturbing imagery. “Coactive enables us to make automatic judgments for about 90 percent of the images uploaded to Fandom. That alone gives us a 50 percent cost reduction and a 74 percent reduction in manual hours spent reviewing images,” said Blachot. “Before, our team and our contractors were spending about 500 hours per week reviewing the queue. Now it’s down to 130 hours maximum. This solution not only reduces cost, but it also has the added mental health benefit, which is priceless.”

“Coactive met our wildest expectations, and year one has been such a success. With year two, we plan to work on our metadata and SEO processes, and we’re looking forward to making them even better,” said Tim Quievryn, Director of Trust, Safety, and Product Support at Fandom. “A good platform is nothing if you don’t have good people representing it. And Coactive has some wonderful people we’ve genuinely enjoyed working with.”

“On average, [the image review process] takes less than half a second. It’s significantly faster than the previous system, where some images could remain on the page for 24 to 36 hours before removal. And it elastically scales with traffic and seasonality.”

Florent Blachot  
VP of Data, Fandom

## About Coactive AI

**Coactive AI** empowers data teams to unlock insights from unstructured image and video data. Coactive makes visual data useful for data practitioners by integrating it with SQL and big data tools they know and love. Using pre-trained models, Coactive helps organizations understand their visual content for use cases like trend analysis, content moderation, search, and mapping.

