wirtualna polska

Partnership

Wirtualna Polska's mission is to be the first-choice news partner of Poles and people living in Poland, providing engaging information, entertainment and services that inspire everyday decisions. In March this year, 21.9 million people used their online products [gemius.pl], according to a Gemius and Polish Internet Research Mediapanel survey.

Wirtualna Polska treats the problem of hate speech with significant care. The publisher strongly believes that a safe platform for open discussion is crucial for a healthy society and democracy, and they are committed to providing such a platform. After in-depth internal workshops, Wirtualna Polska decided to build an automated AI-based comments moderation engine to fight toxic content on their sites.

Overview

Wirtualna Polska operates media, advertising and e-commerce businesses. As one of Poland's leading nationwide media corporations, Wirtualna Polska always tries to stay ahead of trends and motivates its employees to develop innovative products that the company is often first to introduce to the Polish market.

In addition to WP's home page, Wirtualna Polska operates specialized websites covering such topics as news and economy, the automotive industry, technology, entertainment, culture, lifestyle and tourism. Wirtualna Polska also broadcasts digital radio and television.
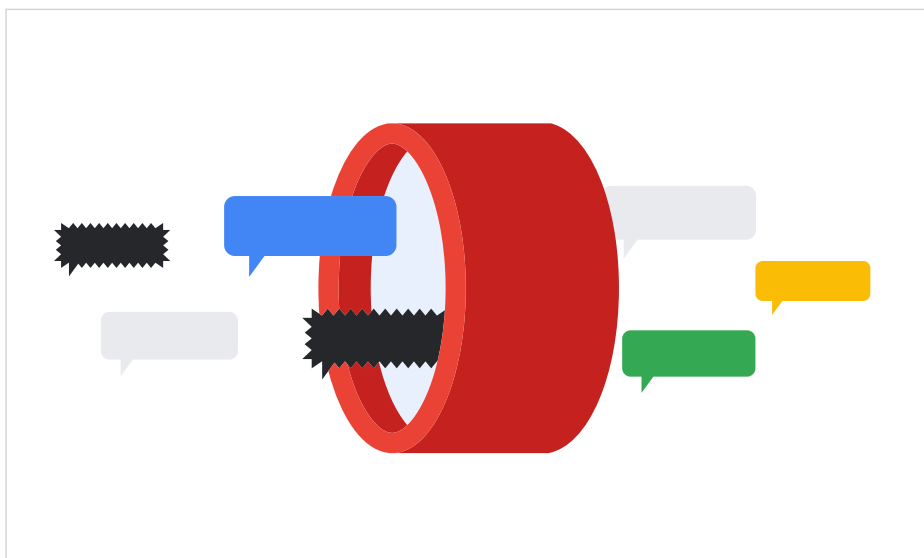
AUTHORS:

- **Damir Agović**
  News & Publishers, Central & Eastern Europe

- **Dorota Cyndecka**
  Head of PM's and Product Owner of Moderation System and Home Page WP.pl

How Wirtualna Polska uses AI to tackle toxic content on their platform

In Poland, WP built a moderation engine to provide a safe platform for open discussion

The Project

In the past, Wirtualna Polska had built their own internal live content moderation team and used third-party systems to moderate the incoming stream of comments. However, the changing legal and political environment in Europe, as well as growing extremism and polarisation in society, meant that these tools are often inadequate for modern journalism. Many publishers have chosen to close their comments sections altogether, as moderation is increasingly an economic burden they can't handle.

WP built a complex and automatic comments moderation engine based on AI to fight toxic content

"The tragic death of the Mayor of Gdańsk Paweł Adamowicz in 2019 was the impulse for us to start the project of developing an automated moderation system at WP," explains Dorota Cyndecka from WP.pl.

"This situation showed us at WP what scale of emotions can pour out from users and how quickly we should respond to the flood of especially negative comments.

We were – and continue to be – also motivated by the ongoing observation of the public debate in Poland and the willingness to play a part in improving it. Hence the launch of the #stophejt (stop hate) campaign during the 2019 parliamentary elections in Poland.

The awareness that readers wanted to express their emotions and opinions, especially those who do not have space for it on a daily basis, is a positive external factor that has a huge impact on the sense of mission and responsibility of WP."

Speaking on the partnership, Cyndecka comments, "Our cooperation with the Google News Initiative is very fruitful for two reasons. First of all, it set us up for the design process stages and mobilized us to present and share the effects at regular intervals. Secondly, the Google News Initiative team was helpful with substantive advice thanks to AI experts."

Because Wirtualna Polska are committed to providing a safe space for open discussion they decided against closing their reader comments. Instead, they used Google Cloud tools to create an automated, AI-based system and ease the burden on content moderators. Wirtualna Polska created a Comments Management Platform (CMP) using Google Cloud tools which allowed them to analyze, calculate and convert higher volumes of data at much higher speeds. Wirtualna Polska had not previously used AI technology or machine learning models in this way or on such a scale.

The Results

"Readers are more likely to speak up in a safe environment that we achieve through better content moderation," adds Cyndecka. "Thanks to models adapted to the type of content, people have a better experience amongst others and on the site itself.

Cyndecka concludes, "As the largest information portal in the country, we believe that we can set trends. We share our knowledge from the project at industry conferences and engage socially by creating campaigns such as #stophejt. We believe that these activities will help reach others with good practices."

Wirtualna Polska plans to develop their project further with the Google News Initiative. The next step will be a system that checks the accuracy of the information provided in the comments — something like a watchdog system. The publisher's aim is to build an open and secure discussion system that is also reliably factual.

40% fewer accepted comments

66% more deleted comments

> " The awareness that readers wanted to express their emotions and opinions, especially those who do not have space for it on a daily basis, is a positive external factor that has a huge impact on the sense of mission and responsibility of WP. "
>
> **DOROTA CYNDECKA, HEAD OF PROJECT & PRODUCT MANAGEMENT, RESPONSIBLE FOR THE MODERATION SYSTEM FOR WP.PL**

7% fewer reported comments

More removed and fewer accepted comments by the new AI model gives our team more time to assess comments that are submitted for manual moderation and we have more time to better train the AI models.

On WP websites, users can report comments that violate the regulations and the netiquette. When we started using the new AI model, the number of user-reported comments was 7 percent points less.

Learn more www.wp.pl