To filter or not to filter - That is the question

William Echikson

"Pope Francis endorses Donald Trump"

During the 2016 US presidential election, this headline spread rapidly through Facebook, provoking a wave of Tweets and YouTube videos. It looked true. Yet it was false. How should policymakers and social media platforms fight such "fake news"?

In Europe, governments in Europe are demanding that Facebook, Google, YouTube and Twitter identify and delete hate speech, terrorist propaganda and other forms of problematic expression. The European Commission has signed a memorandum of understanding that obliges social platforms to speed up their takedowns. Germany has adopted a law that imposes large fines on networks that fail to remove unlawful speech within 24 hours of notification.

On a practical basis, this crackdown represents quite a challenge. The amount of material uploaded is astounding – and growing exponentially. Only a few years ago, some six hours of video were going up every minute on YouTube. Today, it is 300 hours of video per minute. In June 2017, Facebook counted 2.01 billion monthly active users worldwide. Every 60 seconds: 510,000 comments are posted, 293,000 statuses are updated and 136,000 photos are put online.

The only possible way to monitor such a huge amount of content is by using machines. Under pressure from policymakers, Google and Facebook are coming up with algorithms, ranging from keyword filters to AI learning software, to moderate objectionable content. These tools work by matching patterns of behaviour and previously identified illegal content with new uploads or web browsing.

But these automatic tools represent a danger to free speech. According to Emma Llansó, Director of the Center for Democracy and Technology's (CDT) Free Expression Project, who was speaking at a recent CEPS Digital Forum event, much "real" news ends up being removed, along with the fake news like the Pope's endorsement of Trump. All too often, machines find it difficult to distinguish not only between fake and real news, but also between what is appropriate and what is not.

In our discussion, we considered grisly ISIS beheadings. When I worked at Google, YouTube struggled to deal with such content. Were the actions illegal? Of course, yes. But were they newsworthy? Again, yes. Was the content grotesque and did it violate the video platform's own rules against violent content? Yes. The terrorist group was moving to a new level of atrocious behaviour and an answer to these questions was always going to be subjective. No machine could make a simple decision, at least not yet. In the end, YouTube banned most of the videos as its community guidelines offered a half-dozen different grounds for removing them, with specific rules against content that incites violence, crime or hatred, depicts gratuitous violence, or is "intended to shock or disgust". It also has a policy of deactivating accounts held by representatives of organisations designated terrorist groups by the US State Department.

But YouTube doesn't, and couldn't, pre-screen content, relying on users to flag violations. It makes exceptions for videos that demonstrate "documentary or news value", often by adding context or commentary. Even absent such added value, it will often err on the side of letting content speak for itself. That's why the "Innocence of Muslims" video that incited deadly riots in the Arab world in 2012 was still online until a court ordered it taken down on copyright grounds.

Market pressures force the platform to pay attention. Earlier this year, advertisers launched a concerted attack to guarantee that advertising spending won't end up going to the likes of far-right groups. In response, YouTube has promised to hire "significant numbers of people", on top of the thousands who already do the work, to review questionable content.

Should policymakers force additional changes? Perhaps. Konrad Nicklewicz, a former correspondent for Gazeta Wyborcza and now a representative of the Civic Institute in Warsaw, has written a fascinating new report on the future of the news industry. He calls on social media to be regulated like traditional media. While this would not impose a filtering requirement, it would allow readers to sue the platforms for defamation.

Faced with this challenge, the European Commission is charting a cautious balancing act. Christel Mercadé Piqueras, of the European Commission's Fundamental Rights Policy Unit, helped negotiate the Memorandum of Understanding between the Commission and Social Media to combat hate and extremist online speech. In her view, such voluntary measures represent a better way of dealing with this difficult problem than restrictive legislation.

The Internet serves as a bastion for freedom. It takes away the power of the elites (and of governments) to control the flow of information. Today, any of us can find out almost anything about anyone with a few clicks of the keyboard. Each of us can post our opinions to the entire world, free of charge. But this freedom also allows all of us to spread lies and hate. Without shutting down the internet, we will never be able to eliminate all extremism. The best we can do is keep on trying to find the correct balance between freedom and responsibility.

Today, any of us can post our opinions to the entire world, free of charge, but this freedom also allows us to spread lies and hate. We must find the correct balance between freedom and responsibility.

Without shutting down the internet, we will never be able to eliminate all extremism. But we must keep trying to find the correct balance between freedom and responsibility.

CEPS