

Content moderators

BACKGROUND

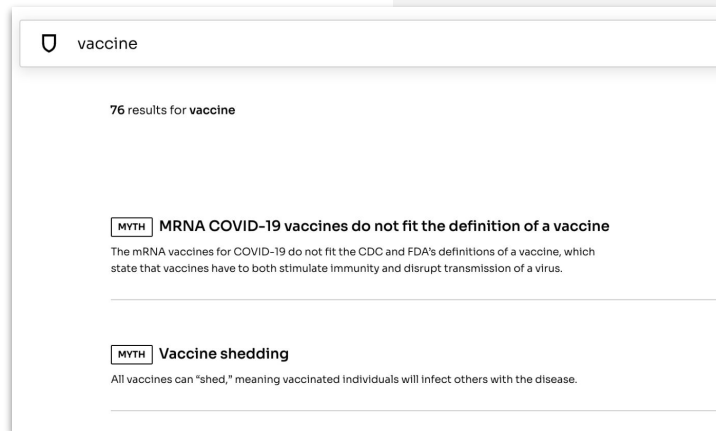
From major social media platforms to smaller online forums, content moderation teams are often overwhelmed by a large amount of continuously updated information, having to sift through posts, reactions, and comments from thousands of users and exercise judgement in order to safeguard communities on the internet.

PROBLEM

When platforms host user-generated content, falsehoods inevitably proliferate, resulting in toxic online environments and PR nightmares. But misinformation is nuanced and constantly evolving, making it difficult for artificial intelligence – and even savvy humans – to spot.

NEWSGUARD'S SOLUTION

- Content moderators rely on NewsGuard's two data sets – the Reliability Ratings and the Misinformation Fingerprints™ – as the most transparent, rigorous, and exhaustive data to help them understand the most important misinformation claims and sources impacting their platforms.
- Using NewsGuard's Insights Dashboard, content moderation teams can search and filter for different site ratings and false claims. For example, a search for “vaccines” calls up all of NewsGuard's Fingerprints related to vaccine myths.



“NewsGuard allows our small Trust and Safety team to stay informed without spending the majority of our time doing research.

The site rating extension helps us quickly identify sources of misinformation and provides plenty of information to inform our enforcement decisions. The Misinformation Fingerprint Library lays out all of the relevant information we need to evaluate content types and accurately enforce our Community Guidelines.

The service keeps us more informed than we would be if we were conducting research on our own, saves so much time, and helps us scale up our operations.”

– **Director of Trust and Safety, major global technology platform**