When medical expertise matters

Uncovering medical misinformation during the COVID-19 pandemic

Factmata teamed up with Centaur Labs to **shut down misinformation** about the COVID-19 pandemic, swapping content moderators for medical experts.

A conspiracy theory that the spray they used to get rid of spotted lantern flys last summer is what caused the coronavirus outbreak

Last 30 days

Factmata built a model to label medical misinformation in tweets with the help of Centaur Labs' network of experts
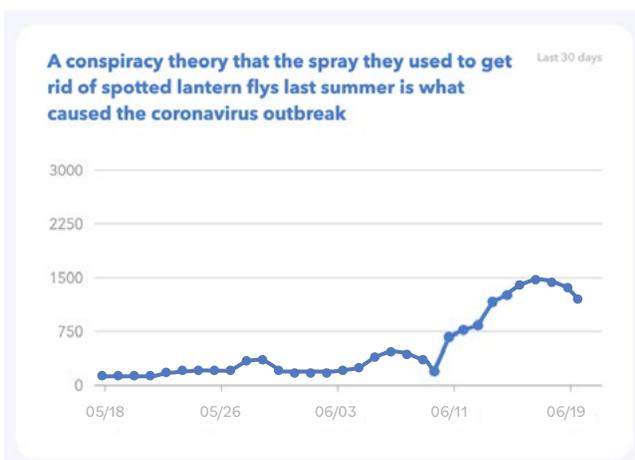
BACKGROUND

Factmata is an artificial intelligence company helping businesses better understand, interpret and explain online content.

Backed by high-profile investors including Twitter cofounder Biz Stone and Craigslist founder Craig Newmark, the London-based startup is applying its patent-pending AI to provide actionable moderation services for online platforms and product insights for brands.

One of its products, Factmata Moderation, identifies content and websites that contain Hateful, Propagandist, Deceptive and/or Fake news. It does this through a hybrid of human and machine where humans are experts who label content used for the training the AI.

Involving human labelers is a key step given that "fake news and propaganda are inherently nuanced and subjective," says Factmata manager Anna Julia Hedges. "It does require expertise to understand the nature of the content and tag it appropriately." Once trained, the AI can analyze an infinite number of new pieces of content without human involvement.

CHALLENGE

With COVID-19 sweeping the globe in the winter of 2020, the Factmata team landed in a unique position to fight the spread of coronavirus-related medical misinformation that proliferated on social networks like Twitter.

However, COVID-19 presented a new challenge because Factmata needed labelers who were not only skilled in the nuances of natural language, but also had medical background to evaluate information that could be harmful to readers' health.

ISSUE #1
MANAGING EXPERT VOLUNTEERS

Initially, Factmata tried to recruit paid medical experts and volunteers on their own but this "turned out to be nightmare," said Claudio Ceruti, an NLP Engineer at Factmata. The pandemic meant most everyone in the healthcare industry was working unprecedented hours and did not have the time to get involved in anything but actually dealing with the crisis.

"Recruiting and then managing the experts required a significant amount of bandwidth that we simply did not have," Ceruti recalls.
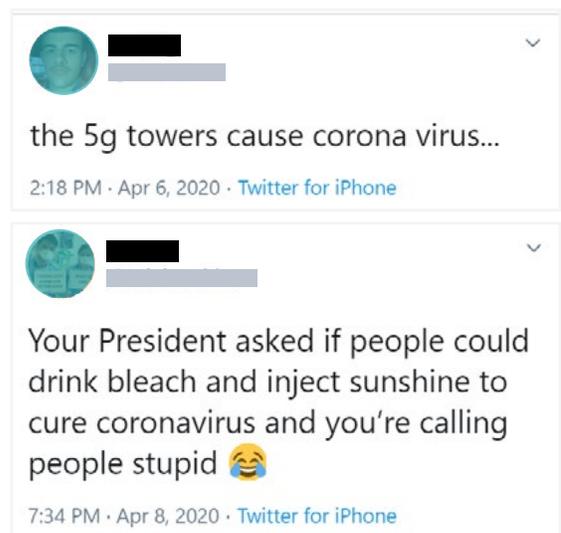
ISSUE #2
CLASSIFYING CONTROVERSIAL TWEETS

Traditional labeling vendors return a singular label for each data sample. However, the Factmata team wanted insight into whether controversial tweets collected multiple, conflicting misinformation labels that should be potentially investigated further.

ISSUE #3
DEFINING MEDICAL MISINFORMATION

Factmata also had to properly define misinformation—and decide how to distinguish medical from general misinformation. How would you classify these tweets?

> the 5g towers cause corona virus...
> 2:18 PM · Apr 6, 2020 · Twitter for iPhone

> Your President asked if people could drink bleach and inject sunshine to cure coronavirus and you're calling people stupid 😂
> 7:34 PM · Apr 8, 2020 · Twitter for iPhone

This is misinformation, but could it be medically harmful? Some tweets contain sarcasm, which can be difficult to classify. Even after creating a working definition, it proved difficult to educate an army of labelers on the definitions.

SOLUTION

As the pandemic continued to accelerate, Factmata was on a mission to find a suitable partner who could offer highly accurate medical labels in a matter of days instead of months. Factmata found Centaur Labs and the two teams quickly got to work on the project.

In less than one week, 4,000 tweets were analyzed for medical and general misinformation using 58,000 opinions gathered from Centaur Labs' network of healthcare experts.

Going beyond a single label per tweet allowed Factmata to identify highly controversial tweets and properly tease out sarcasm.
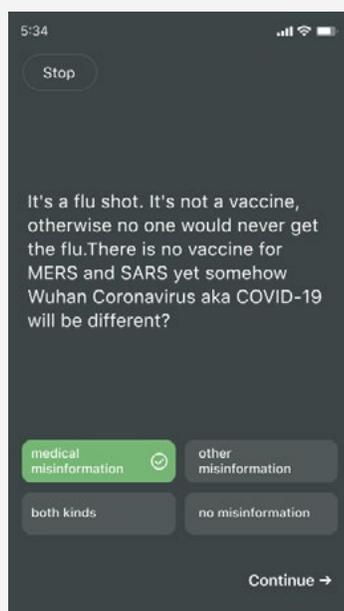
Factmata believed that having a greater number of expert opinions per tweet helped refine their algorithm across the full spectrum of misinformative tweets.
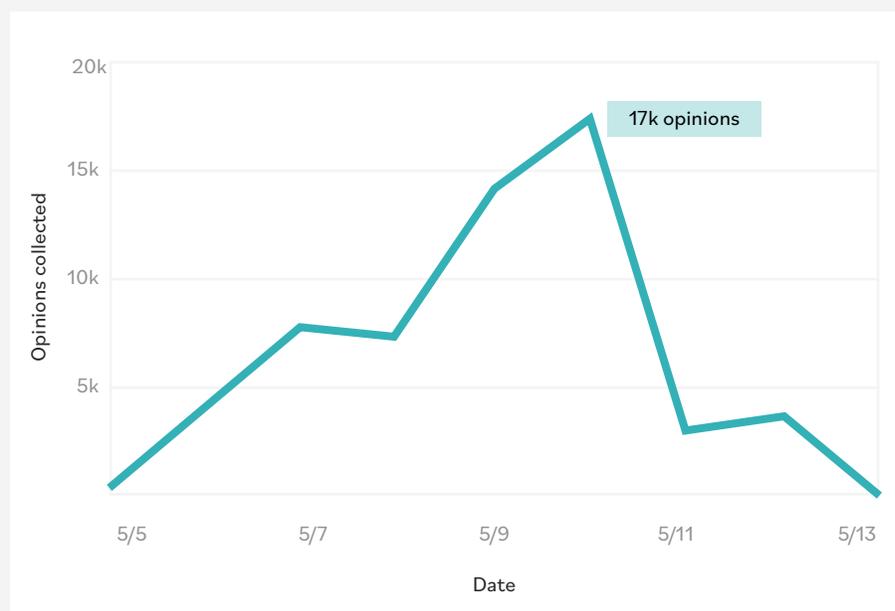
58,177
skilled opinions collected

4,167
tweets labeled

14
opinions / tweet

346
opinions / hour

Centaur Labs labeling interface

Expert opinions collected per day on Factmata tweets using the Centaur Labs platform

BENEFITS

In addition to the rapid turnaround time, the following features stood out to Factmata:

TURN-KEY SOLUTION

Centaur offers a full service solution that ingests raw tweets and returns highly accurate labeled images managed through a simple platform. Meanwhile, hiring, vetting, and quality assurance typically associated with labeling services is abstracted away.

ACCURACY AT THE MARGINS

By offering 10+ opinions per tweet, Centaur captures a more nuanced view into the underlying meaning of the tweets. A single medical expert might miss the nuance in some tweets, but aggregating opinions from multiple experts allows Centaur to determine the overall sentiment. This information is lost by competitors who return a single value for each tweet.

QA BUILT IN

Instead of training and retraining labelers and throwing out potentially faulty results, the Centaur Labs process automatically flags controversial data. Factmata used the Centaur Labs dashboard to monitor and update tweets that the network found difficult to label, in real-time. These updates were automatically reflected to Centaur's labeling network, giving immediate feedback on the labeling task to expert labelers.

PRICING THAT SCALES AFFORDABLY

Factmata needed to work quickly and with a limited budget in order to get high quality labels. As an alternative to paying physicians (or waiting to find volunteers), the Factmata team was amazed at the Centaur Labs price point—which was in the ballpark of the unskilled crowd of workers on Amazon's Mechanical Turk.
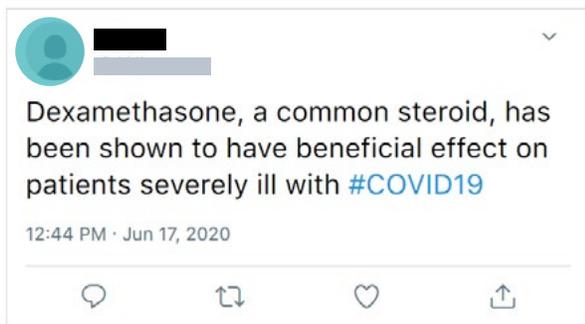
❝

I get pitched on annotation tools all the time and **Centaur is simply the best** ... they offer you the accuracy and sophistication medical experts at the price and speed of Mechanical Turk."

Dhruv Gulati   CEO of Factmata

RESULTS

Factmata used the labeled tweets to create their Artificial Intelligence model, composed of a non-linear SVM classifier applied to sentences embedded using BERT. This model is able to reach **68% accuracy** in classifying general and medical misinformation on an internal annotated dataset.

It is also showing promising results when analyzing new data. As an example, the model pointed out this tweet as potential medical misinformation:

Dexamethasone, a common steroid, has been shown to have beneficial effect on patients severely ill with #COVID19

12:44 PM · Jun 17, 2020

The tweet is indeed true, but the model is able to select among thousands of tweets and pass it to medical experts in order to have it validated.

After the validation, the model will be retrained and will learn that Dexamethasone is known for having beneficial effects in treating COVID-19, and it won't propose this tweet to medical experts unless it is used in misleading context. This procedure eases the task of annotation by medical experts providing only the subset of novel and relevant tweets that requires validation.

With the help of Centaur Labs, Factmata was able to build out and refine an AI misinformation model for a novel pandemic when it mattered.
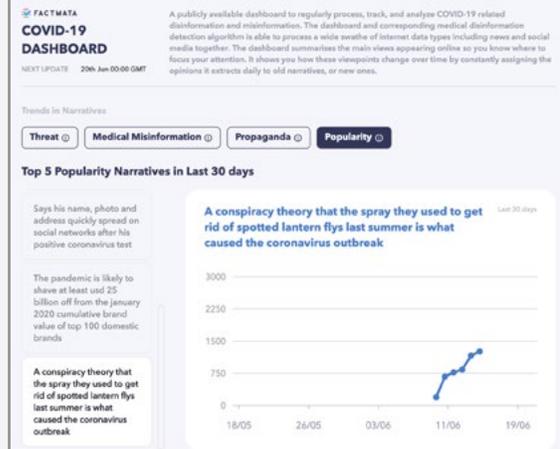
FACTMATA COVID-19 MISINFORMATION DASHBOARD

Interested in learning more about Factmata? Email **info@factmata.com** for more information.