

# Decoding Hate

Using Experimental Text Analysis to Classify Terrorist Content

By Abdullah Alrhoun, Shiraz Maher, Charlie Winter



Global Network  
on Extremism & Technology

The research team developed an algorithm capable of automated, nuanced and rapid analysis of jihadi content to demonstrate that content moderation can be fine-tuned to better account for context.

An archive containing 214 GB of official Islamic State (IS) content was mined by using a range of experimental text-processing techniques.

Some 6,290 items (including photo-reports, videos, radio bulletins and magazines) covering the years from 2015 to 2019 were processed by the algorithm.

While the algorithm could not process one fifth of the items, its analysis of the remaining four fifths was consistent with that of the human coders approximately 91 percent of the time.



To check the algorithm's reliability, it was tested against a team of three human coders and triangulated with extant qualitative research on IS media operations.

The analysis showed that the archive was a partial—albeit large and indicative—sample of IS's total propaganda output during the years in question.

The corpus reflected three key trends that have characterised IS media in recent years: (i) a decline in productivity; (ii) a tack away from civilian life; and (iii) a diffusion in geographic focus.

The algorithm, while imperfect, demonstrates that an additional layer of automated analysis could be leveraged alongside human moderation teams to better triage and resource content disruption efforts.