

**BRAND SAFETY'S TECHNOLOGICAL CHALLENGE:**

# **HOW KEYWORD BLACKLISTS ARE KILLING REACH AND MONETIZATION**

**A STUDY BY CHEQ'S DEPARTMENT OF DATA SCIENCE**

**September 2019**



**CHEQ**

# BACKGROUND

## Brand Safety: One Of Digital Media's Greatest Challenges

Brand Safety describes the controls that companies in the digital advertising supply chain employ to protect brands against negative impacts to the brand's consumer reputation associated with specific types of content and/or related loss of return on investment (Trustworthy Accountability Group and Brand Safety Institute).

The issue of brand safety has become a growing concern for advertisers. An eMarketer survey (June 2018) listed it as the second greatest challenge of programmatic buying, second only to media transparency.

## Keyword Blacklists Have Become The Solution of Choice

Advertisers, eager to mitigate brand safety risk, turned to ad verification solutions offering keyword blacklists as the primary solution. Advertisers today will upload lists ranging from 2,000 unsafe keywords and up to 4,000 in some cases. These lists are then used by verification partners as a filtration mechanism to detect and block harmful content containing those keywords.

## The Issue: Keywords Are Often Extremely Ambiguous

A company that reports to be using keyword blacklists, cites words including "Shooting", "Dead", "Kill", "Crash" "Attack" and "Explosion" among their most commonly blocked words. The problem is that hundreds of such words could easily be used within a perfectly safe context (e.g. Video "killed" The Radio Star), rendering the practice blunt and inaccurate. The full scale of the failure of keyword lists is laid out in this study.

*"We're frustrated that blacklists aren't managed particularly well. So we're getting blocked for keywords like "shoot." We write about photo shoots constantly."*

James Wildman, CEO, Hearst UK



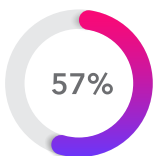
# EXECUTIVE SUMMARY

## MORE THAN HALF (57%) OF NEUTRAL OR POSITIVE STORIES ON MAJOR NEWS SITES ARE BEING INCORRECTLY FLAGGED AS UNSAFE FOR ADVERTISING.

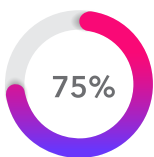
This is severely damaging publishers' ability to monetize premium content. CHEQ studied one day's news cycle (18 July 2019) across 15 top online news sites. The findings, analyzing news sites including CNN, The Guardian, and The New York Times, are alarming.

At best, the crude use of blocklists are severely and unnecessarily limiting reach for ad campaigns (brands) or killing monetization (publishers). However, at worst, the widespread use of keywords is attacking democratic principles, purging key moments of history and making minority news and opinion unviable.

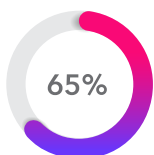
## FINDINGS INCLUDE



57% OF SAFE ARTICLES WERE INCORRECTLY **BLOCKED**



75% OF SAFE HISTORY RELATED NEWS CONTENT IS BEING **BLOCKED**



65% OF ENTERTAINMENT RELATED NEWS CONTENT IS BEING **BLOCKED**

## BREAKDOWN OF SAFE CONTENT ANALYZED

NEWS SITE	SERVED	BLOCKED
<i>The</i> <b>ADVOCATE</b>	27%	73%
<b>PinkNews</b>	27%	73%
<b>B R</b> BLEACHER REPORT	33%	67%
<b>WSJ</b>	33%	67%
<b>VARIETY</b>	33%	67%
<b>FT</b> FINANCIAL TIMES	33%	67%
<i>The</i> New York Times	40%	60%
<b>CNN</b>	40%	60%
<i>The</i> Guardian	47%	53%
<i>the</i> japan times	47%	53%
<i>The</i> Washington Post	47%	53%
<b>news</b> .com.au	47%	53%
<b>COSMOPOLITAN</b>	53%	47%
<b>BUSINESS INSIDER</b>	60%	40%
<b>HUFFPOST</b>	73%	27%

# METHODOLOGY

## TO IMITATE THE KEYWORD-FOCUSED PRACTICE USED IN THE INDUSTRY, CHEQ PROCURED AND DEPLOYED A STANDARD KEYWORD BLOCKLIST OF 2000 WORDS.

CHEQ used this list to analyze the first 15 uncontroversial, or positive, articles on 15 leading news sites – 225 articles in total, on 18 July 2019. This is a highly typical list of words used by advertisers and publishers.

Such a blocklist is used to ensure ads are automatically blocked from any mention of news and current events dealing with "death", "sex", "war" or "assault" to avoid damaging a brand's reputation, by association with violent, controversial or sexual news stories.

The blocklist was applied against the selected articles with the goal of measuring the amount of safe content being incorrectly blocked.



# METHODOLOGY

## A SNAPSHOT OF THE BLOCKLIST

### WE PROCURED FOR THE TEST

A

abortions

abraao jose bueno

abuses abusing

acute radiation syndrome

acute respiratory distress syndrome

admission schemes

admissions scheme

adrian von renteln

adult

aileen carol wuornos

aino nykopp koski

akira nishiguchi

akku yadav

al qaeda

al shabaab

albert fish

ald 52

alejandro m̃jynez

alfred jodl

alfred meyer

allegation

allegations

amfonelic acid

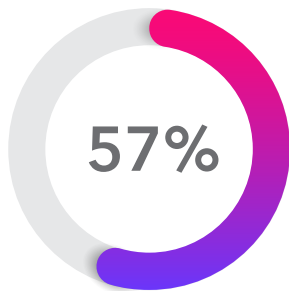
amir qayyum

### Behind the Blocklists

We did not use the most egregious blocklists. These can entirely block brands from appearing next to articles mentioning “gay”, “Muslim”, “fat” and “celebrity”. Equally, we did not block entirely irrelevant words related to old news cycles. One ad verification provider reported that as late as July 2019, it was working with 562 clients who use blocklists which include “Las Vegas”, “Charlottesville”, “Westminster”, “London Bridge”, and “Ariana Grande”, based on past incidents involving these keywords.

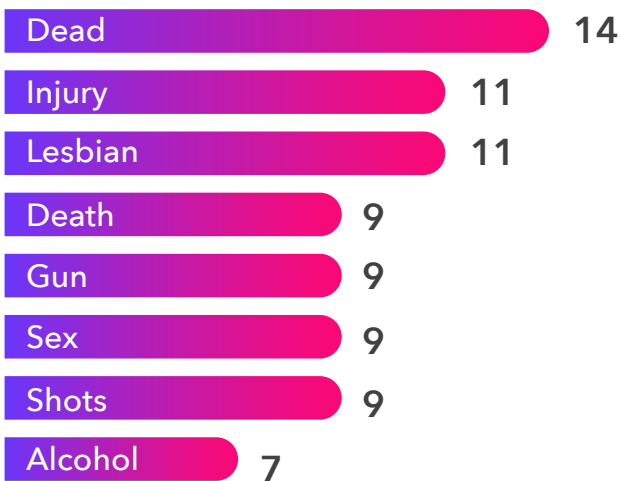
These worst-case examples effectively demonetize every mention of the UK’s third biggest city of 2.7 million, the home of Parliament and a multi-platinum, Grammy Award-winning recording artist. If such all-encompassing blacklists had been used in our study, the percentage of unnecessarily blocked news articles would have been significantly higher. Even using the most conservative keyword list (a sample of which is provided on the left) we see the far-reaching effects of safe content being blocked.

# KEY FINDINGS



57% OF SAFE ARTICLES  
WERE INCORRECTLY  
**BLOCKED**

## MOST COMMON KEYWORDS INCORRECTLY BLOCKED\*



## OTHER COMMONLY BLOCKED WORDS

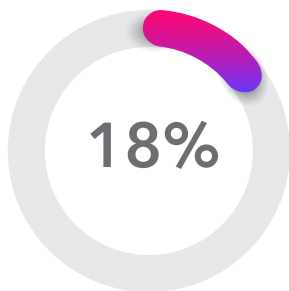
Attack	7	Cancer	5
Racial	7	Conspiracy	5
Sexy	7	Cult	5
Shoot	7	Kill	5
War	7	Killing	5
Army	5	Penthouse	5
Bisexual	5	Shot	5
Adult	5		

\* Number of times an article was blocked on account of that keyword

**FOR THE FULL SCOPE  
OF BLOCKED CONTENT,  
PLEASE CLICK  
THE FOLLOWING LINK**

[www.cheq.ai/full-list-keywords](http://www.cheq.ai/full-list-keywords)

# ADDITIONAL FINDINGS



**1 OUT OF EVERY 5** USES OF THE WORD 'DEATH' WAS IN COMPLETELY SAFE ARTICLES, YET THE CONTENT WAS STILL FLAGGED

## Death by Keywords

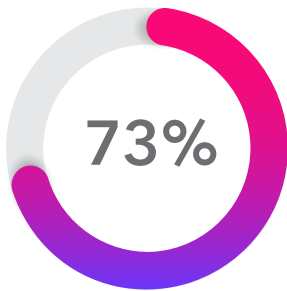
The most common incorrectly blocked keywords on brand safety grounds were variations of "dead", "death" and "died" (18% of stories blocked from receiving revenue in the study). This included Woodstock festival organizers declaring plans to revive the event "dead" (*New York Times*); mention of American rock band The Grateful "Dead" (*Variety*); internet sensation, Grumpy Cat who "died" (*CNN*) and economist-turned-restaurant reviewer Tim Harford describing in the *Financial Times* online, a meal as so good he could "die".

## Sports-based "injuries" "shots", "blocking", and "attacks"

Second-placed "injury" saw blocking of words in stories, primarily relating to sports stories. This including a 1700 word story (*The Bleacher Report*) rendered unsuitable for ads because of one word (mentioning that an Arkansas Razorbacks player picked up an "ankle injury").



# ADDITIONAL FINDINGS



73% OF SAFE LGBTQ  
RELATED NEWS CONTENT  
IS BEING BLOCKED

## NEARLY THREE-QUARTERS OF LGBT STORIES UNMONETIZABLE

Using these industry standard keywords, we analyzed two major news sites for the gay community, *Advocate* and *PinkNews*. Across 30 neutral or positive articles analyzed, 73% of their safe inventory on this day was blocked from serving ads. Though we did not use anything close to the most discriminating keyword lists (where any mention of "gay" is penalized), almost all lists mention "lesbian", "bisexual", "sexual" "drag queens" or "same sex marriage".

In this case, the stories which were penalized included mention of "cult lesbian thriller, *Killing Eve*" ('cult', 'lesbian' and 'killer'), a lesbian version of *Othello*, and - in one of the biggest news stories of 2019 - that the US and Netherlands' female soccer teams "featured five lesbian and bisexual athletes." (flagged for "lesbian and bisexual"). Tellingly one of the few safe stories on LGBT sites on this day was not discussing a gay figure, but an article about a heterosexual mayoral candidate.





# ADDITIONAL FINDINGS

## ADVERTISERS BLOCKED FROM BUYING ADS ON PAGES CONTAINING WORDS LIKE 'GAY' AND 'LESBIAN'

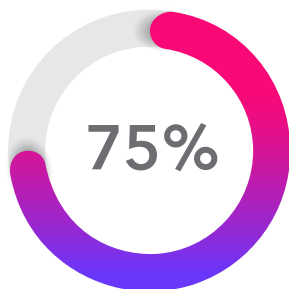
Benjamin Cohen, editor of PinkNews says: "On the open marketplace basically a whole heap of our content gets blocked for no legitimate reason. A lot of ad networks are blocking content for the word 'lesbian' because they lazily think lesbian equals porn. I don't think the issue is the brand, its whoever is administrating their blacklist because often the brands are really surprised when you send them the blacklist." It was not only LGBT sites that suffered. The *Financial Times* also mentioned "lesbian" in one story (mentioning the rights of "gay lesbian, bisexual and transgender politicians in the United States"). Such content is becoming increasingly harder to monetize with the rapid rise of brand safety blacklists.

The long-term damage from this issue has been brought into focus by the closure in July 2019 of *Gay Star News*. Ad exec and LGBTQ+ campaigner Jerry Daykin says: "We'll continue to see minority voices squeezed out of media if advertisers don't actively think about how they spend with them - brand safety measures often actively exclude LGBT audiences with blocked keywords."

*"A lot of ad networks are blocking content for the word 'lesbian' because they lazily think lesbian equals porn."*

Benjamin Cohen, CEO, PinkNews

# ADDITIONAL FINDINGS



75% OF SAFE HISTORY  
RELATED NEWS CONTENT  
IS BEING BLOCKED

## History Purged

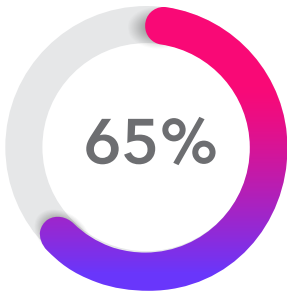
Brute brand safety is also pulling the rug of advertising from stories discussing history. *The Financial Times*, like many sites on this day, celebrated 50 years of the moon landing. The news site included an historic JFK speech which includes the word "weapons". The offending line from JFK, deemed brand unsafe 57 years later: "We have vowed that we shall not see space filled with weapons of mass destruction, but with instruments of knowledge and understanding".

Equally unsafe history was a *Washington Post* article mentioning that 1968 was the year when the Rev. Martin Luther King Jr. and Robert Kennedy were "assassinated". In another example of blocked history, *Japan Times* reported that "tens of thousands of soldiers died at the battle of Waterloo".

Kelly Facer, Vice President of Revenue Operations, Sightline Media, the world's largest publisher of history sites news sites such as History Net, Military Times, and Defense News says fighting blocklists is a "daily battle". She says: "We must work directly with the advertiser and the agency to work to get the blocklist removed. It is not being disclosed to us that there is a keyword blocklist even when we ask. It sounds absolutely crazy that the First World War is not brand safe but when the technology is just scouring the page for specific words rather than context it becomes challenging."



# ADDITIONAL FINDINGS



65% OF ENTERTAINMENT  
RELATED NEWS CONTENT  
IS BEING BLOCKED

Keywords struggle to understand the difference between fact and fiction. Some 65% of content relating to movies and TV was flagged, effectively preventing monetization. Especially hard for keywords was Emmy-nominated *Killing Eve* ("killing") blocked three times, Tom Cruise's return in *Top Gun* ("gun") blocked twice; *The Lion King* (blocked for "death", *Walking Dead* ("dead"), *Dead to Me* ("dead") a new season of *Fargo* ("mafia"); *Game of Thrones* ("arousal" of passion by some fans to the season's climax) and *Fantastic Beasts: The Crimes of Grindelwald* (blocked for "crimes").

## TV & MOVIE INVENTORY

## BLOCKED FROM MONETIZATION



# EXAMPLES OF SAFE CONTENT

## FLAGGED AS UNSAFE

### **Pelicans rookie NBA Summer-League spotlight (*Bleacher Report*)**

I've been working on it a lot, Hayes said in regard to his three-point **shot** versus the Cleveland Cavaliers. "I used to **shoot** them in high school. I was open, so I took it."

### **How a switch to employee ownership is bearing fruit (*Financial Times*)**

Staff have, for example, changed the process to **weed** out toxic leaves that grow among the wild garlic

### **'Game of Thrones' Emmy Love Confounds the Show's Critics (*Variety*)**

That it (*Game of Thrones*) **aroused** such passions from a fanbase that stood by it for a decade, and that it drew those fans even closer as it pushed back against their affection, makes the case.

### **Twitter Is Completely and Totally Obsessing Over Blue Ivy's Cameo in Beyoncé's "Spirit" Video (*Cosmopolitan*)**

"Spirit" is the first single from the album, confirming that a *Lion King* movie with the original number of Nala solos songs (read: zero) basically would have been a **crime** (*Cosmopolitan*).

### **Woodstock festival news coverage (*New York Times*)**

In late April, the festival's financial backer pulled out, declaring the event **dead**.

### **Brussels clears Vodafone's €19 billion Liberty Global Deal (*The Financial Times*)**

The company now has a **war** chest of about € 10.6 billion, having sold its German cable business Unitymedia to Vodafone for 12 times the unit's operating cash flow.

# IMPLICATIONS

## GROSS OVERBLOCKING, HURTING BOTH SUPPLY AND DEMAND

### **Publishers Lose Ability to Monetize**

Publishers are suffering, as keyword practices render decent content as unmonetizable. Some 93% of publishers say brand safety solutions are hurting their revenues through overblocking. One publisher explained: "We have had times in the past when block rates are high and the clients have turned around and said they are not going to rebook, or where we have had to refund them. We recently had a campaign running that flagged 40% of impressions that day as unsafe. We published an article on how the Chernobyl disaster zone is becoming a sanctuary for wildlife, which was being blocked because it had the word 'disaster' in it."

*"What turned out to be a reaction to protect a brand from unsafe things that are mostly user generated content, on sites such as YouTube, ended up hurting media companies focused on producing real journalism".*

Christine Cook, CRO, CNN

### **Advertisers Losing Reach**

Many brands like the idea of erring on the side of caution when it comes to brand safety. But, with the scale of overblocking as demonstrated, it is safe to say that keyword blocklists are significantly limiting campaign reach. A recent survey by CHEQ and Digiday (June 2019) supports this notion, as 92% of marketers said that blacklists and whitelists severely limiting reach, damaging the effectiveness of campaigns.

*"If you're spending precious ad dollars to be around content, you'd better have a system that knows the difference between a shooting and "shooting a movie."*

Seth Rogin, CEO, Nucleus

# TAKEAWAYS

While it is imperative that brands have in place measures to avoid genuinely unsafe material, in almost all cases, brands and publishers will be left baffled by the wide range of inventory being denied advertising dollars.

This issue has wider implications. Publishers are facing a daily battle to understand blacklists and suffer a **direct impact on revenue opportunities**. At the same time, blunt brand safety is often **denying brands an opportunity to target the exact audiences that they want to reach**. Even more damagingly keyword-based brand safety is **hurting minority voices**, such as LGBT communities.

Despite these issues, **the use of blunt technology is only increasing**. One ad verification provider has said that last year alone they saw a 250% increase in blocking through this mechanism.

Clearly a more sophisticated approach to understanding these challenges is now necessary. This needs to adapt to a modern online digital world and the explosion of online content seeking ad revenue. **The use of AI is one key technology that offers the ability to understand**, with a human degree of accuracy, the exact meaning of online content without resort to keywords. In milliseconds, AI can confirm with human-levels of accuracy what a piece of content is – and is not.

Besides technology, brand safety must also have at its heart more transparency and accountability. It is imperative that brand safety provides for greater openness of decision-making, monetization of all safe content, while enabling and championing a thriving online news ecosystem.



# CHEQ

**Research Lead:** Udi Baumatz, Head of Data Science

**Project Director:** Jonathan Marciano, Director of Communications

**Strategy:** Daniel Avital, CSO

**FOR FURTHER INQUIRIES CONTACT**

Priti Shah, Director of Business Development, US (East Coast)

[priti.s@cheq.ai](mailto:priti.s@cheq.ai)