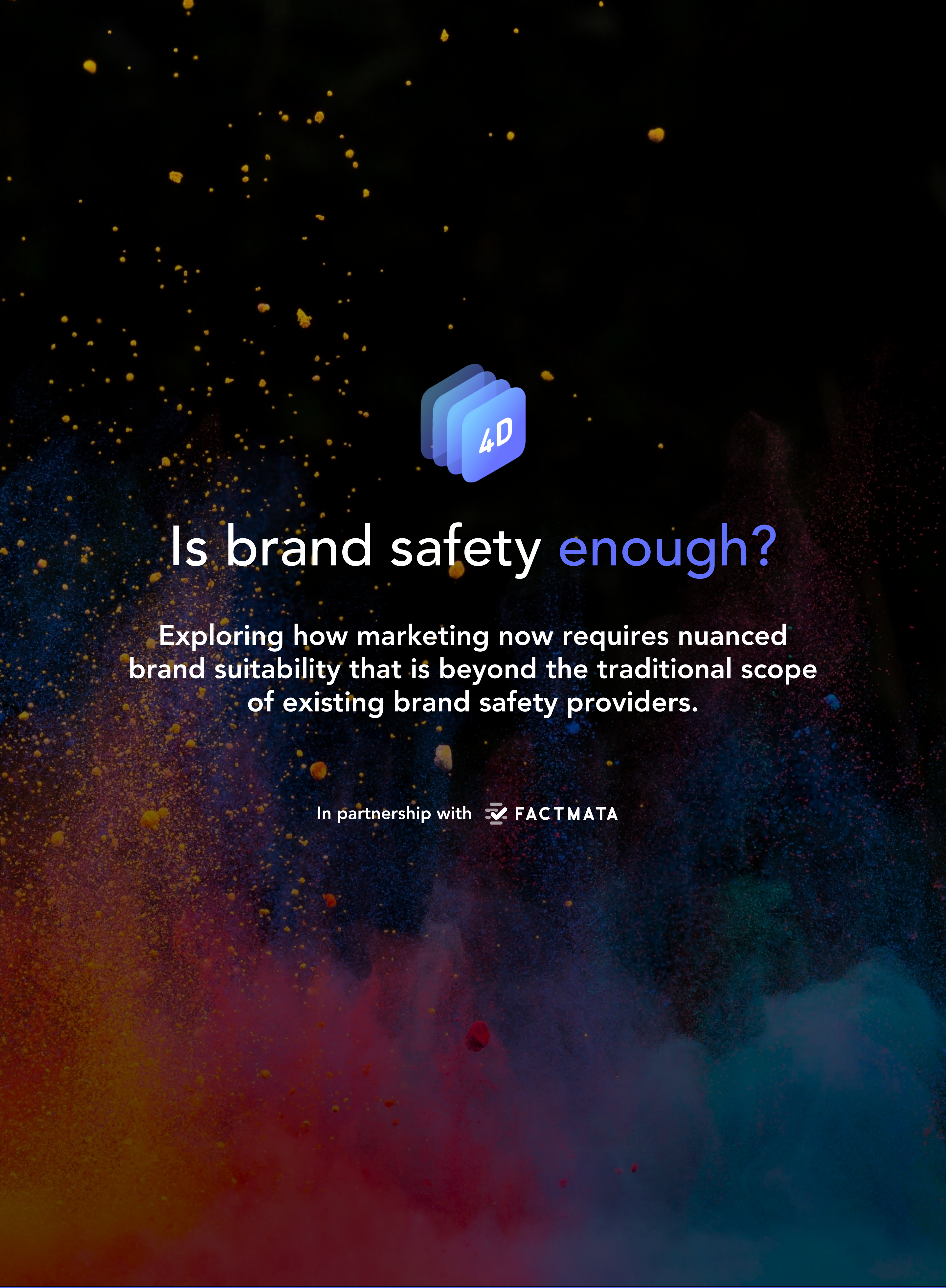


Is brand safety enough?

Exploring how marketing now requires nuanced brand suitability that is beyond the traditional scope of existing brand safety providers.

In partnership with FACTMATA

A 4D Research Study

Table of contents.

Executive summary.

The open web is a difficult place to navigate, and brand safety has fast become the number one focus for marketers moving forward in this politically and economically volatile environment.

Brand safety technologies exist to help advertisers avoid placing ads on inappropriate sites, or next to concerning content. Many of the leading brand safety tools that currently serve these needs do a great job at protecting brands across the web.

Yet, as with many challenges that reside within the (m)adtech landscape, brand safety technologies face their own issues, and given the high risk associated with failure, these tools often find themselves held to the highest scrutiny.

Because the vast majority of these traditional brand safety solutions were built at the dawn of programmatic advertising, many have not kept up to speed with the nuances in how harmful online content is framed, worded, and produced. Dated algorithms rely on keywords, which alone do not protect a brand entirely. And, with restrictions of the walled gardens making it even more difficult to manage brand safety, businesses are searching for new solutions.

What we've learned:

4D and Factmata uncover in this paper, the use case for a blended approach to brand safety and suitability. As we step into the new marketing age, where third-party cookies fade away, and contextual intelligence will surge, brands can look to adopt a variety of tools to safeguard them against future threats.

Key take-aways include:

One size does not fit all.

Brands looking for safe and suitable solutions will want to align themselves with advanced solutions beyond traditional methods that move far beyond the scope of traditional brand safety methods.

Unsuitable content, unsuitable spend.

Unsuitable content flagged by Factmata, but missed by existing brand safety vendors, equates to 0.71% of total spend. With global programmatic spend in 2020 reaching $126.5bn, advertisers might have spent up to $898m on content considered unsuitable or unsafe for brands.

Nuance is key.

Modern marketers who want to thrive in the new marketing age need to be aware of nuance and true context to identify the right moment and the right environment in order to align with the right message.

$898 million

of global programmatic spend is still being spent on unsuitable content.

Introduction.

In 2020, the unforeseen global pandemic brought brand safety and suitability challenges into sharp focus for many marketers.

As publishers across the globe unveiled the impact of Covid-19 to their readers, they simultaneously began to lose revenue from a broad spectrum of advertisers who were blocking pages inadvertently.

Brand safety

Helps advertisers avoid placement or context that could potentially harm the brand or its reputation.

Brand suitability

Considers the meaning, context, and potential implications of content, specific to an actual brand's needs.

Why? Well, the fundamental methodology of many brand safety tools relies simply on the blocking of keywords.

If an advertiser, for example, decides to add Covid-19 to their blocklist in order to avoid appearing next to negative news stories, they would effectively be blocking almost all of the leading, global news sites (as well as many others). With a recent Ofcom[1] report revealing that nearly nine out of ten adult internet users turn to traditional media as a source of Covid-19 information, advertisers could be missing out on an abundance of audience engagement opportunities.

The message is clear. Old ways of working are not enough on their own.

According to a study[2] commissioned by CHEQ and Digiday, nearly two-thirds of advertisers say brand safety tools are not fit for purpose.

92% of the marketing respondents stated that they would forgo the use of brand safety tools if they were not achieving adequate reach, while 99% are seeking more customized tools to ensure safety, without sacrificing reach. So, the question is: is brand safety on it's own enough? Or, do we need to take a more sophisticated approach that blends reach with safety?

99% Of marketers, agencies and brands are seeking more customized tools to ensure brand safety, without impacting reach. *CHEQ 2019*

This whitepaper will explore what the future of brand safety looks like, and how brands can future proof their brand safety and suitability strategies in a world beyond keywords.

Marco Godina
SVP Product,
4D - A division of Silverbullet

Source: 1) Ongoing Ofcom Covid-19 Report, 2) CHEQ 2019 Study

Introducing FACTMATA

Factmata is a London-based AI-company with the core goal of making the internet a better place. They help brands, publishers, and platforms focus on nuanced brand safe environments by giving them a deep understanding of the quality, safety and credibility of any piece of content on the web.

Unlike traditional solutions which rely on keyword algorithms, Factmata's machine learning tool accurately scores online content against eight different signals, giving each page a rating to determine how much of that content matches each score. These signals include:

- Racism
- Hyper partisanship
- Fake news
- Sexism
- Personal insult
- Threatening language
- Toxicity
- Obscenity

> "Moving beyond keywords and labelling of standard brand safety violations, Factmata have built a patent-pending technology that combines AI with feedback from communities, journalists, advocacy groups, and expert knowledge to intelligently and accurately score online content beyond just a top-level understanding."
>
> Dhruv Ghulati
> CEO Factmata

More details of each signal are available on their website, as well others coming soon: https://factmata.com/signals.html and a website demo is available here. including a website demo: https://try.factmata.com

Methodology.

To create this paper, 4D's Media Activation team ran a series of A/B tests to compare the effectiveness of traditional brand safety solutions at identifying and blocking complex suitability signals.

A test campaign was set up on The Trade Desk, with a significant spend budgeted for two line items. Ad units in both line items contained a privacy compliant JavaScript tag that allowed for the full path URL of all impressions to be collected and scored by Factmata.

A (Brand safety absent)
This line item had no brand safety filters applied, and instead relied on standard inventory filtering offered by the DSP at the supply-side level.

B (Brand safety applied)
This line item had traditional brand safety filters applied on top of standard inventory filtering. Brand safety was powered by industry leading brand safe vendors, and focused on the 11 "sensitive topics" as identified in the IAB & GARM 2.2 context taxonomy[3] *released at the end of 2020. These include Online Piracy, Adult Content, Terrorism* and *Obscenity/Profanity.*

All other targeting, bidding and media variables were identical. The campaigns ran for three full days in order to gather enough impression-level data to ensure the results were statistically significant.

Following the activation, 4D's Data Science team and the Factmata team analysed the impression-winning URLs, scoring these pages against Factmata's brand safety signals, to determine a Factmata *Trust Score* per URL (that combines all the algorithms into one - see below).

The Trust Score is Factmata's custom score which combines all signals together in a single metric. The higher the score, the more likely any given offending signal is found within the URL.

Signal	Unlikely	Possible	Very Likely
Hate Speech	< 0.4	0.4 - 0.7	> 0.7
Hyper Partisanship	< 0.5	0.5 - 0.9	> 0.9
Racism	< 0.6	0.6 - 0.8	> 0.8
Sexism	< 0.65	0.65 - 0.8	> 0.8
Personal Insult	< 0.4	0.4 - 0.7	> 0.7
Threatening Language	< 0.4	0.4 - 0.7	> 0.7
Toxic Language	< 0.4	0.4 - 0.7	> 0.7
Obscene Language	< 0.4	0.4 - 0.7	> 0.7

These thresholds are determined by Factmata's advanced machine learning algorithms, and denote the likelihood of each URL containing certain content. For this test we identified the overall Trust Score as "possible" if at least one signal is flagged on the page as Possible, and so on.

Source: 3) https://iabtechlab.com/press-releases/tech-lab-releases-for-comment-content-taxonomy-to-improve-brand-safety-support-brand-suitability/

The Results.

Are brand safety solutions doing enough?

The test results indicated that whether existing available brand safety filters are applied to campaigns or not, Factmata would consistently block 4% to 5% of the total budget, as this is being spent on Hate Speech, Propaganda, Sexism and Racist Content unsuitable to brands and advertisers.

$6.3B

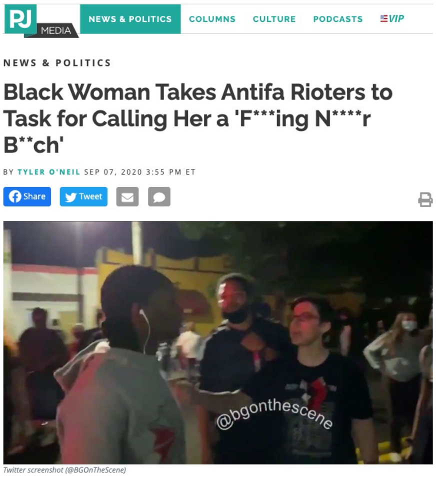
Programmatic spend globally throughout 2020 was expected to reach **$126.5 billion**, revealing the 5% of spend on Propaganda, Racism, Sexism or Hate Speech could equate to **$6.3 billion.**

However, when it comes to issues such as Racism, Obscenity and Toxic Language, the overwhelming majority of this content is successfully filtered out at both the DSP level and through traditional brand safety solutions. These filters successfully use keyword-based solutions to blocklist and block the long-tail of media impressions available in a campaign, but struggle when violations require more nuance to be identified.

When brand safety was applied from existing Brand Safety vendors, 0.71% of the spend was still directed towards content that Factmata flagged as potentially unsuitable. Specifically, 0.42% of this spend violated the Hate Speech score on URLs, and 0.1% against the Hyper Partisanship score.

$898M

Spent on unsuitable content flagged by Factmata (but not traditional brand safety solutions) as part of global programmatic spend.



NEWS & POLITICS

Black Woman Takes Antifa Rioters to Task for Calling Her a 'F***ing N****r B**ch'

BY TYLER O'NEIL SEP 07, 2020 3:55 PM ET

Twitter screenshot (@BGOnTheScene)

https://pjmedia.com/news-and-politics/tyler-o-neil/2020/09/07/watch-black-women-confront-antifa-rioters-for-tearing-up-and-trashing-portland-n902083

Source: Statista Programmatic Spend 2020

To the left is an example of content that Brand Safety vendors missed. This article from a conservative media publication, discusses the Portland Riots and Antifa in the wake of the BLM Protests.

- Factmata picked up strong signals against Hyper Partisanship (0.78), Toxic Language (0.67) and Hate Speech (0.63)
- Although the article itself tries to be a balanced piece, it still discusses a highly contentious topic - and one which most brands would prefer to avoid advertising against
- It is also published on a website which is self-described as providing specifically one side of an argument, hence the high Hyper Partisanship score. However, the article frames Racism issues as being pitted against Anti-Fascist protests, and thus scores relatively lowly on Racism signals (0.01)

Improved reliability and accuracy with Factmata.

Since the 4D test scored all impression-winning URLs against eight Factmata signals, we also researched how different types of brand-unsafe content fared against traditional brand safety vendors.

The results reveal that content which requires more nuance to be understood is more likely to be missed by traditional brand safety solutions, resulting in a higher Factmata score. Remember, a higher score means a higher likelihood for offending content to be found.

An example URL picked up by Factmata flagged higher score-wise than what a traditional brand safety vendor found:

Hyper Partisanship (38% increase)
Threatening Language (41% increase)
Sexism (28% increase)
Personal Insult (6% increase)

However, content with more explicit and obvious brand safety violations were successfully filtered out by existing brand safety vendors.

Examples of this included content containing Hate Speech and Racism themes.

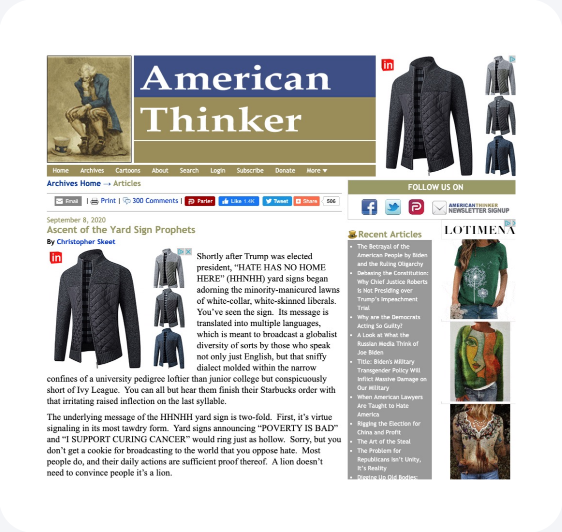
To summarise, when brand safety filters were applied, impression-winning content scored higher for more nuanced signals.

Although it may seem counterintuitive to suggest more unsafe content was flagged after brand safety filters were applied, it highlights how this filtering focuses on obvious content violations, and misses content which only subtly violates brand safety standards.

A blended approach to brand safety is therefore needed.

Harmful content flagged by Factmata alone.

The test highlighted examples of unsafe content which was blocked by Factmata, but not identified by leading brand safety tools. These included:



September 8, 2020

Ascent of the Yard Sign Prophets

By Christopher Skeet

Shortly after Trump was elected president, "HATE HAS NO HOME HERE" (HHNHH) yard signs began adorning the minority-manicured lawns of white-collar, white-skinned liberals. You've seen the sign. Its message is translated into multiple languages, which is meant to broadcast a globalist diversity of sorts by those who speak not only just English, but that sniffy dialect molded within the narrow confines of a university pedigree loftier than junior college but conspicuously short of Ivy League. You can all but hear them finish their Starbucks order with that irritating raised inflection on the last syllable.

The underlying message of the HHNHH yard sign is two-fold. First, it's virtue signaling in its most tawdry form. Yard signs announcing "POVERTY IS BAD" and "I SUPPORT CURING CANCER" would ring just as hollow. Sorry, but you don't get a cookie for broadcasting to the world that you oppose hate. Most people do, and their daily actions are sufficient proof thereof. A lion doesn't need to convince people it's a lion.

- The Betrayal of the American People by Biden and the Ruling Oligarchy
- Debasing the Constitution: Why Chief Justice Roberts is Not Presiding over Trump's Impeachment Trial
- Why are the Democrats Acting So Guilty?
- A Look at What the Russian Media Think of Joe Biden
- Title: Biden's Military Transgender Policy Will Inflict Massive Damage on Our Military
- When American Lawyers Are Taught to Hate America
- Rigging the Election for China and Profit
- The Art of the Steal
- The Problem for Republicans Isn't Unity, It's Reality
- Digging Up Old Bodies:

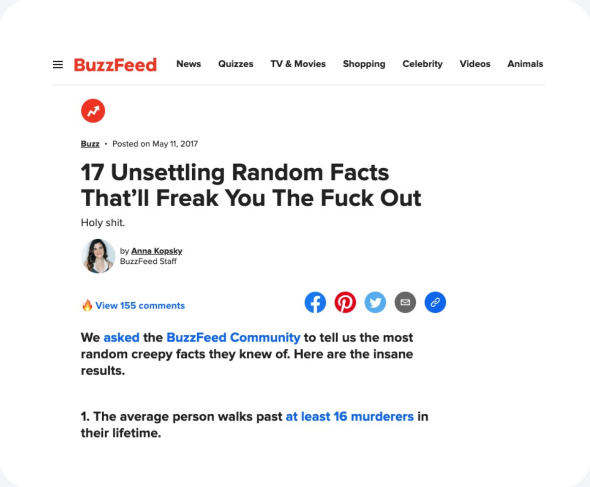
https://www.americanthinker.com/articles/2020/09/ascent_of_the_yard_sign_prophets.html

Article countering woke/progressivism ideology.

- Partisan content (0.88) and Hate Speech (0.68) both score very highly on each signal
- The content - although in itself not advocating anything serious - presents a one-sided viewpoint and can thus be described as being partisan
- The article discusses different slogans used by left-wing activists, most of which revolve around Hate Speech, hence its relatively high score
- Brands and advertisers would want to avoid this content. Although in itself harmless, it can convey association between the opinions given and the brand advertising

Buzzfeed article titled "17 Unsettling Random Facts That'll Freak You The F*ck Out"

- This article ranked especially highly on Toxic (0.89) and Obscene (0.54) signals
- Interestingly, traditional brand safety filters **failed to notice the profanity in the subject title**
- Again, an innocent enough piece of content to some, but far from being the sort of imagery you would want associated with your brand



Buzz • Posted on May 11, 2017

17 Unsettling Random Facts That'll Freak You The Fuck Out

Holy shit.

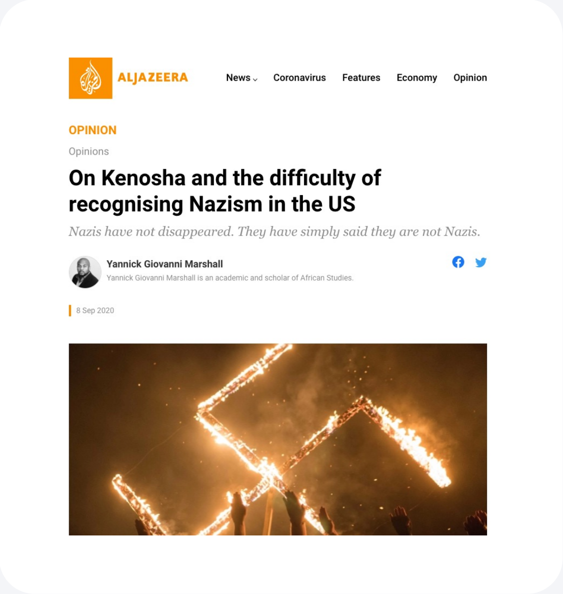
by Anna Kopsky
BuzzFeed Staff

We asked the BuzzFeed Community to tell us the most random creepy facts they knew of. Here are the insane results.

1. The average person walks past at least 16 murderers in their lifetime.

https://www.buzzfeed.com/annakopsky/creepy-facts-thatll-freak-you-the-hell-out

Harmful content flagged by Factmata alone.



OPINION

Opinions

On Kenosha and the difficulty of recognising Nazism in the US

Nazis have not disappeared. They have simply said they are not Nazis.

Yannick Giovanni Marshall

Yannick Giovanni Marshall is an academic and scholar of African Studies.

8 Sep 2020

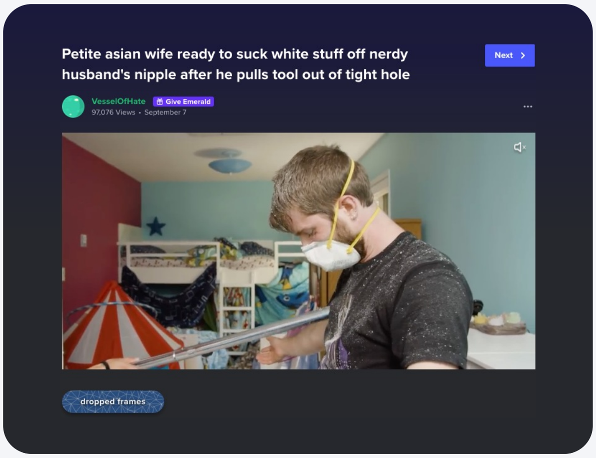
https://www.aljazeera.com/opinions/2020/9/8/on-kenosha-and-the-difficulty-of-recognising-nazism-in-the-us

Al Jazeera article discussing the prevalence of white supremacy nationalists and Nazism in the US

- This article scored particularly highly across Hyper Partisanship (0.84) and Hate (0.74) signals
- Typically, brand safety would flag any content that mentions topics such as Nazism, but this article slipped through the net. Although it's from a reputable publication and is a balanced investigative piece, if brand safety is applied it should be blocked
- Factmata can help your brand identify the few cases which slip through the net such as this

An Imgur.com gif of a well-known youtuber *Linus Tech Tips*, with a suggestive title: "Petite asian wife ready to suck white stuff off nerdy husband's nipple after he pulls tool out of tight hole"

- Although technically a correct label of the gif, Factmata were able to spot the nuance of the text, scoring highly against Toxic (0.83) and Obscene (0.33) signals
- A clear example of brand safety solutions not being built for the modern age of content, and allowing potentially damaging content slip through the net



Petite asian wife ready to suck white stuff off nerdy husband's nipple after he pulls tool out of tight hole

VesselOfHate

97,076 Views • September 7

https://imgur.com/gallery/3ejOmx1

Summary.

The open web is a difficult place to navigate, and brand safety has fast become the number one focus for marketers moving forward in this politically and economically volatile environment.

> "It's no easy feat. Leading brand safety tools are playing a fantastic role in the effort to protect brands, but cannot do it alone. Brands need to explore additional layers of protection to gain 100% confidence in where their ads are being placed."
>
> **Dhruv Ghulati,**
>
> CEO Factmata

A blended approach:

4D and Factmata have uncovered the use case for a blended approach to brand safety and suitability. As we step into the new marketing age, where third-party cookies fade away, and contextual intelligence will surge, brands can look to adopt a variety of tools to safeguard them against future threats.

One size does not fit all.

Brands looking for safe and suitable solutions will want to align themselves with advanced and trusted products that move far beyond the scope of traditional brand safety and targeting methods. As keyword measurement alone proves to show pitfalls, a blended approach to safety/suitability is key.

Unsuitable content, unsuitable spend.

Unsuitable content flagged by Factmata (but not traditional brand safety solutions) is estimated to be £19.6 million ($26.8 million) of **UK programmatic spend**, meaning 5% of budget is landing on unsafe or unsuitable environments. This investment could be utilised elsewhere.

The world is complex. Nuance is key.

Modern marketers who want to thrive in the new marketing age need to be aware of nuance and true context to identify the right moment and the right environment in order to align with the right message. The world we live in is complex, so a more considered approach is vital.

“4D, built by the leading marketing transformation company Silverbullet, is set to pave way for the post-cookie era with our trusted partnership with Factmata. Together, 4D and Facmata’s effective contextual targeting and brand suitability engines can analyse all types of content that exist on a page, to give true 360 degree guidance as to the page's true meaning. This partnership allows marketers to have greater confidence in where their ad is placed, to protect its identity and ethos.”

Marco Godina

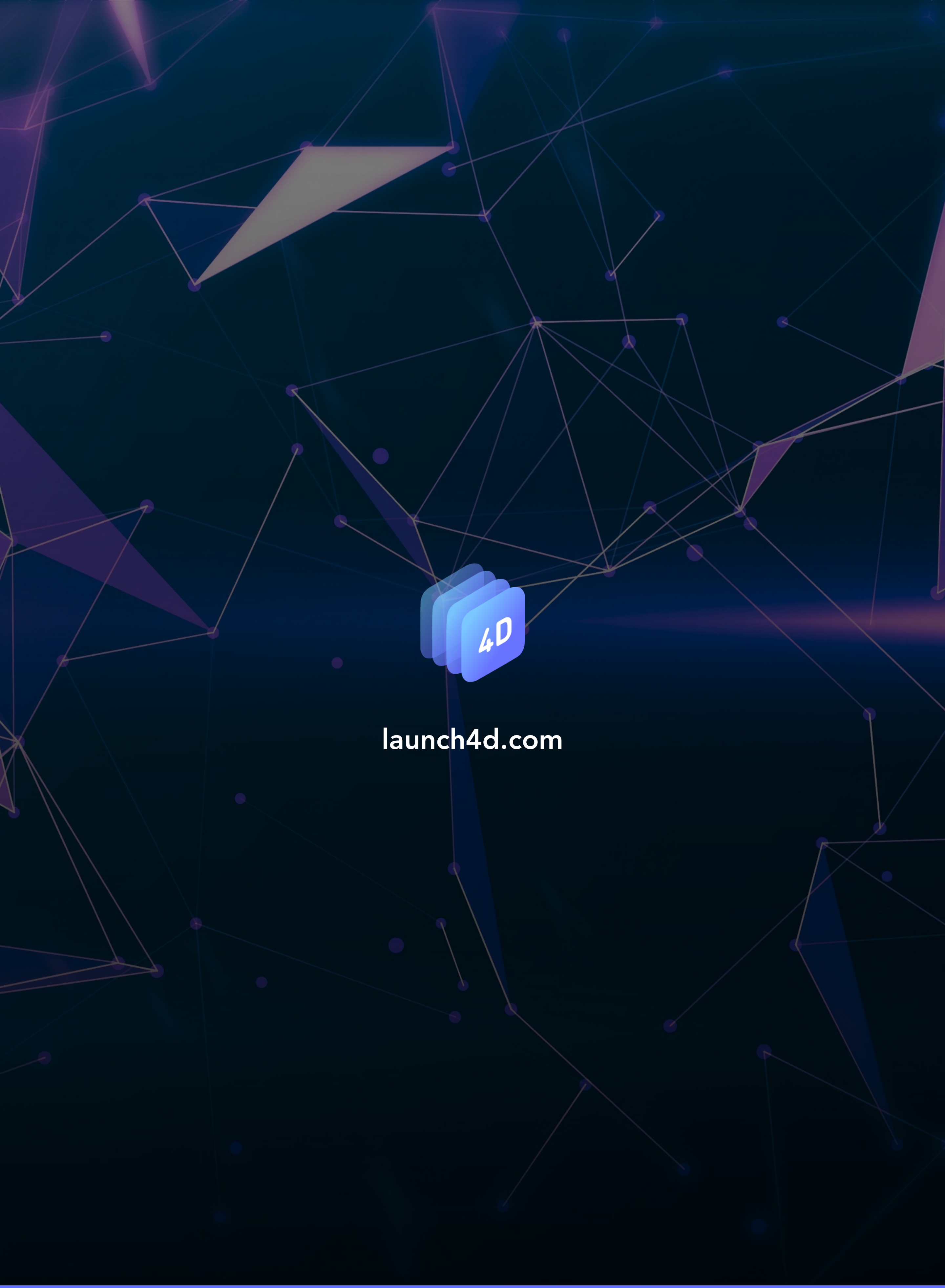
SVP Product, 4D – a division of Silverbullet.

To find out more about how 4D and Factmata can help you, contact us today.

launch4d.com

FACTMATA

factmata.com



launch4d.com