Surveilling young people online: An investigation into TikTok's data processing practices

Reset. AUSTRALIA

Dylan Williams
Rys Farthing
Alex McIntosh

July 2021

Contents

01. Introduction

THIS SUMMARY OUTLINES RESET AUSTRALIA'S INVESTIGATION INTO SOME OF TIKTOK'S DATA PROCESSING PRACTICES, AS A CASE STUDY OF HOW SOCIAL MEDIA PLATFORMS COLLECT AND USE CHILDREN'S DATA.

In many ways, TikTok is not unique among social media platforms in how it collects and uses children's data. Data underpins the attention economy and the business model of almost all social media platforms. The longer a user is engaged on a social media platform, the more opportunities the platform has to profit by serving them personalised advertising. This is the primary business model of most social media platforms[1], from Facebook to Snapchat to Twitter and TikTok. And it is true for both adults and children.

TikTok is of particular interest because of its popularity with younger people. Gen Alpha (born between the early 2010s to now) and Gen Z (born in the mid to late 90s to early 2010s) make up TikTok's two biggest user groups in Australia by age[2]. This means TikTok is collecting a lot of data about young Australians.

How children's data is collected and used warrants particular attention in part because there is so much of it. Gen Z and Gen Alpha are 'datafied' before they take their first breath; from commercial pregnancy apps, heartbeat monitors to ultrasounds shared on social media, data was extracted and harvested from them in utero. This data collection continues throughout their childhood, from AI enabled baby monitors to connected toys. One estimate suggests over 72 million data points are collected about children by tech companies by the time they reach 13[3]. And at 13, they are able to join social media platforms; the real engines of the attention economy. Between 13 and 18 young people will make an estimated 70,000 social media posts[4], all of which are analysed for content, location tagging, interactions with friends... The amount of data that is now held about Australia's younger generations is truly staggering.

And this data hungry business environment presents unique risks for the realisation of children's rights. The recently published UN General Comment about children's rights in the digital world outlines that processing children's personal data can:

1 Ranking Digital Rights 2020 'It's the Business Model' https://rankingdigitalrights.org/wp-content/uploads/2020/07/Its-the-Business-Model-Executive-Summary-Recommendations.pdf

2 Roy Morgan 2020 'Nearly 2.5 million Australian's using TikTok' http://www.roymorgan.com/findings/8538-launch-of-tiktok-in-australia-june-2020-202010120023

3 In Donell Holloway 2019 "Surveillance Capitalism and Children's Data: The Internet of Toys and Things for Children." Media International Australia, Incorporating Culture and Policy 170(1), pp. 27-36

4 Children's Commissioner of England and Wales 2018 Who Knows What About Me? https://www.childrenscommissioner.gov.uk/digital/who-knows-what-about-me/

'result in violations or abuses of children's rights, including through advertising design features that anticipate and guide a child's actions towards more extreme content, automated notifications that can interrupt sleep or the use of a child's personal information or location to target potentially harmful commercially driven content.[5]'

The business model that drives these risks rests on the uncomfortable collection and use of young people's data.

WHY IS UNREGULATED DATA RISKY FOR CHILDREN?

Data can pose real risks to the realisation of children's rights.

Data can potentially be used to shape and limit future opportunities for children in ways that are concerning. Data has a problem with permanence; unlike paper it does not fade away unless it is actively deleted. This poses a particular problem for children who short of tragedy have a long time to live with all this data. Already, older generation's significant life events are shaped by data; large employers use AI to scan CVs and banks use as much data as they can to run credit checks and approve mortgages and loans. Carrying around a whole lifetime's worth of data could mean that events in your past continue to shape your future in ways we cannot even begin to imagine. Offensive social media posts made when you were 13 could limit job opportunities in your 50s, or data about a mental health crisis at 15 could be used to deny access to medical insurance in your 60s. A whole childhood of data could be used to determine or limit eligibility for future opportunities.

Aside from damaging future opportunities, here-and-now data can also be used to shape childhoods in ways that aren't fair or amplify existing bias. Data and algorithms are already used to determine a range of childhood experiences in Australia. For example, young people completing school in NSW are awarded a HSC score which is automatically converted into an ATAR[6] score for university entry by an algorithm. Decisions around child support payments in Australia can be legally made by computer programmes[7]. While both these systems may be functioning perfectly safely, there are international examples where these exact systems have failed children in unfair and biased ways. Last year in the UK, a 'mutant algorithm' converted high school grades into university entrance criteria, in a way that systematically downgraded the scores of children from low-income areas[8]. Earlier this year in the Netherlands, data about children's ethnicity was used to automatically 'red-flag' child support payments, leading to Black and dual national families automatically having payments stopped in error[9].

continued on next page →

5 UN Committee on the Rights of the Child 2021 'General Comment 25 on Children's Rights in Relation to the Digital Environment' https://www.ohchr.org/EN/HRBodies/CRC/Pages/GCChildrensRightsRelationDigitalEnvironment.aspx

6 UAC 2015 Calculating the Australian Tertiary Rank in New South Wales https://www.uac.edu.au/assets/documents/atar/atar-technical-report.pdf

7 Child Support (Assessment) Act 1989 (Cth) Section 12A https://www.legislation.gov.au/Details/C2016C00954

8 Sean Coughlan 2020 'A Levels and GCSEs: Boris Johnson blames mutant algorithm for exam fiasco' https://www.bbc.com/news/education-53923279

9 European Parliament Anti Racism & Diversity Group 2021 'Condemning the Dutch Child Benefit Scandal' https://www.ardi-ep.eu/wp-content/uploads/2021/02/2021_02_01-ARDI-letter-to-EC_the-Netherlands_child-benefits-scandal_final.pdf

As we move towards more data driven practices, the experience of growing up as an Australian will be increasingly shaped by data. Data is a notorious agent of bias[10], and it is not always guaranteed that these data-driven practices will be fair.

But data can also affect children's freedom of thought, and threaten their right to be 'let alone', especially from aggressive and persuasive advertising. A recent Unicef report outlined how personalised and behavioural advertising can be 'used to manipulate children's consumption patterns and behaviours'[11]. This manipulation can take troubling forms. A recent investigation by Reset[12] found that Facebook used the data it collected from tracking children's online behaviour, such as what posts they liked and who they interacted with, to identify children who were interested in alcohol, weight loss and gambling. This sort of 'microtargeting' was made available to advertisers with nefarious interests. This relentless personalised targeting can be risky, and impact on children's freedom of choice and expression.

Ultimately, this can harm society as well as children. Excessive data collection and use of children inculcates 'surveillance' culture in the next generations, and paves the way for even more surveillance that can be used in unanticipated and harmful ways[13].

10 Ruha Benjamin 2019 Race After Technology Polity Press Cambridge UK

11 Unicef 2020 'The Case for better governance of children's data: A manifesto' https://www.unicef.org/globalinsight/media/1741/file/UNICEF%20Global%20Insight%20Data%20Governance%20Manifesto.pdf

12 Dylan Williams Alex McIntosh Rys Farthing 2021 Profiling Children for Advertising https://au.reset.tech/news/profiling-children-for-advertising-facebooks-monetisation-of-young-peoples-personal-data/

13 Unicef 2020 'The Case for better governance of children's data: A manifesto' https://www.unicef.org/globalinsight/media/1741/file/UNICEF%20Global%20Insight%20Data%20Governance%20Manifesto.pdf

THIS INVESTIGATION USES TIKTOK AS A CASE STUDY TO EXPLORE THE PROBLEMATIC PROCESSING OF YOUNG PEOPLE'S DATA WITHIN THE BROADER PROBLEMS OF THE ATTENTION ECONOMY.

Previous research by Reset Australia has interrogated many popular social media platforms, and raised multiple issues[14]. Unfettered use of children's data is risky business, and TikTok is one part of this broader issue.

The report outlines two key investigations:

- **TikTok's data collection practice.** Young people are meant to consent to having their data collected by agreeing to TikTok's terms and conditions.

  This report used an experimental method to analyse how TikTok obtains consent from young people, and surveyed 238 16 and 17 year old TikTok users. It finds that it is not always clear if young people have meaningfully consented to the data collection practices employed by TikTok.

- **TikTok's data use practices.** Young people's data is used, among other things, to create personalised content recommendations.

This report used an experimental method to audit the algorithm driving TikTok's recommender system, which is created by and trained on young people's data. This report analyses the capacity of the recommender algorithm to deliver questionable content to young people. It finds that TikTok's algorithm can quickly and easily learn to serve young people content that reinforces negative ethnic stereotypes and content that perpetuates negative views about women.

It also finds, in this instance, TikTok's algorithm could not be trained to recommend content that promoted dieting and weight loss, nor Covid-19 misinformation to young people. This could either be a limitation of this experiment, or it could suggest that TikTik may be actively directing their algorithm to moderate this type of content.

14 Dylan Williams Alex McIntosh Rys Farthing 2021 *Profiling Children for Advertising* https://au.reset.tech/news/profiling-children-for-advertising-facebooks-monetisation-of-young-peoples-personal-data/ and Dylan Williams, Alex McIntosh, Rys Farthing 2021 Did We Consent to That? https://au.reset.tech/news/did-we-really-consent-to-this-terms-and-conditions-young-people-s-data/

02. About TikTok

1 WHAT IS TIKTOK?

TikTok is a social media platform. Specifically, it is a User Generated Content (UGC) video streaming platform that allows users to both post and watch short form video clips.

2 WHO USES TIKTOK?

TikTok has been downloaded over 2 billion times globally, and is extremely popular in Australia with 2.5 million active monthly users[15]. In the first half of 2020 TikTok grew its Australian user base by 52.4% making it the fastest growing social media platform in the country[16].

The platform is dominated by users under the age of 30, with more than 70% of its Australian users belonging to Generation Alpha or Generation Z[17].

3 HOW DO PEOPLE USE IT?

Users are able to both watch and create content on TikTok[18].

For the creation of content, TikTok offers a number of features that are common across other UGC video platforms (like filters and editing tools), and each creator will have a channel that holds all their video posts together.

For the consumption of content, users are served up a stream of videos that TikTok's algorithm predicts they will be interested in. Users also have the ability to interact with videos (through likes, comments or 'following' a video's creator), which helps to 'personalise' their stream. They can also share videos with their friends on TikTok or other popular social media platforms like Facebook and Twitter.

15 Roy Morgan 2020 'Nearly 2.5 million Australian's using TikTok' http://www.roymorgan.com/findings/8538-launch-of-tiktok-in-australia-june-2020-202010120023

16 Roy Morgan 2020 'Nearly 2.5 million Australian's using TikTok' http://www.roymorgan.com/findings/8538-launch-of-tiktok-in-australia-june-2020-202010120023

17 Roy Morgan 2020 'Nearly 2.5 million Australian's using TikTok' http://www.roymorgan.com/findings/8538-launch-of-tiktok-in-australia-june-2020-202010120023

18 Although it is worth pointing out that users can watch and create content to different extents, some users are avid consumers who do not create content or create very little.

4 HOW DO PEOPLE FIND VIDEOS ON TIKTOK?

The masses of video content created and hosted on TikTok can be organised in a number of ways, including by the content creator who made the video, or by tagging each video with relevant hashtags. Users have the ability to search keywords, with either creators names or hashtags associated with a video, to locate content that is of interest to them.

But users don't have to specifically 'search and select' content to be served up videos. Each user is provided with three main personalised channels[19] ('streams') through which they are recommended video content, including content from creators and advertisers:

- The 'Search' stream. If a user searches for a word or phrase, a selection of videos that TikTok thinks match this phrase are recommended. How videos are chosen and selected to appear in this stream is driven by an algorithm created by TikTok.
- A 'Following' stream. It serves a stream of video content from creators a user has chosen to follow. After a user has selected which creators are included on their Following stream, how these videos are chosen and selected to appear in this stream is also driven by an algorithm created by TikTok.
- A 'For You Page' stream (FYP). It streams video content that TikTok predicts users may like. How videos are chosen and selected to appear in this stream is also driven by an algorithm coded by TikTok.

The FYP stream is the main vehicle users use to discover content and watch TikTok. This recommender system, and the algorithm underpinning it, is the subject of this research.

I still have no idea what TikTok is or how it works at all

TikTok can be understood through comparison with terrestrial television. Terrestrial television has a limited number of channels, and each channel has an editorial team responsible for deciding what content is aired (or for this comparison, recommending what its viewers watch on their Friday nights). The content selected by each channel's editorial team is created by professional production crews, to varying levels of quality and popularity. Very few professionals are involved in television production, or making editorial choices about what to air, whereas many people are TV viewers.

On TikTok however, every user is potentially both a viewer and a content creator. TikTok allows every registered user to create and post ('air') content on their channel without needing a TV editor's approval. Needless to say, this also produces varying levels of content quality and popularity.

As a viewer, each TIkTok user is provided with three channels (or streams); a search channel; a follower channel, and; a 'For you' channel. What is posted on each channel is decided by TikTok's recommender system. This recommender system is not an editorial team like at a TV station, it is an algorithm created by TikTok. This algorithm is trained on, and uses data about its users, to decide what to recommend and post on each channel.

19 For clarity, users could also view a stream of videos by looking at a content creator's page. This would allow them to stream all the videos posted by that viewer. This video stream is not governed by an algorithm, rather is a chronological stream of videos so it's not explored in this report.

03. Data collection practices & consent

Terms and Conditions (T&Cs) create the 'rules of engagement' that shape the relationship between a platform and its users. They are important because they outline what users, including young users, consent to when they join a service. This includes consenting to data collection and use. The report draws together existing and new analysis of TikTok's T&Cs and explores if they are presented in ways that maximise meaningful, informed consent for young users.

METHODS

This research involved four steps.

1. **Establishing a new TikTok account on a mobile phone,** documenting the signup process and what the Terms and Conditions are, and how they are presented.
2. **Analysing the language of TikTok's terms and conditions.** To determine if a young person would be able to reasonably understand these T&Cs, we analysed them using Readable.com.
3. **Establishing a new TikTok account on a mobile phone,** documenting the signup process and what the Terms and Conditions are, and how they are presented.

   To evaluate positive design features, we assessed each platform against five techniques identified in previous research as maximising understanding[20]:

   - Are key terms presented as frequently asked questions?
   - Are icons used to illustrate key terms?
   - Are the T&Cs shown directly on the sign up page?
   - Is the information provided in short chunks?
   - Do they use comics or illustrations?

   To evaluate negative features, we reviewed the signup process to identify dark patterns or persuasive design features. These are design techniques that may encourage young users to hand over more data than they need to.
4. **Conducted an online survey with 238 Australian 16 and 17 year old TikTok users.** We ran an online survey to capture young people's thoughts and opinions about TikTok's data collection and use. This was a self-selected sample of 16 and 17 year olds, not a representative sample.

20 Behavioural Insights Team, 2019 Best practice guide for Improving consumer understanding of contractual terms and privacy policies https://www.bi.team/wp-content/uploads/2019/07/Final-TCs-Best-Practice-Guide-July-2019-compressed.pdf and Marietjie Botes, 2017. 'Using Comics to Communicate Legal Contract Cancellation'. The Comics Grid: Journal of Comics Scholarship 7(0) p. 14

FINDINGS

TikTok's terms and conditions are not presented in ways that maximise meaningful, informed consent.

ANALYSING THE LANGUAGE OF TIKTOK'S TERMS AND CONDITIONS

TikTok's terms and conditions are spread over 8 documents, many of which are difficult to find. Two of these are presented for review when a user joins via a mobile phone, and 6 are 'hidden' on joining (see figure one).

PROMPTED TO REVIEW ON JOINING	ESTIMATED COST OF REACHING 1000 PEOPLE FROM THIS AUDIENCE
◆ Terms of Service ◆ Privacy Policy	◆ TikTok Platform Cookies Policy ◆ Privacy Policy for Younger Users ◆ Open Source Software Notices ◆ Virtual Items Policy ◆ Law Enforcement Data Request Guidelines ◆ Intellectual Property Policy

Figure 1: TikTok's terms and conditions documents and their location when you join via a mobile phone.

Once a user has found them, TikTok's T&Cs are difficult to read. They scored 41.73 on the Flesch Reading Ease scale, which suggests they take a Tertiary Level degree to comprehend. They are also extremely long. Combined, they comprise 73,101 words which would take an estimated 5 hours and 24 min to read (see Appendix 1 for a full list). It is worth noting that among the most popular 10 apps in Australia, TikTok has the lengthiest T&Cs by far. Instagram's T&Cs are less than a third as long (at 21,480 words), and Snapchat's are almost one fifth (at 15,378 words).

ANALYSING THE LANGUAGE OF TIKTOK'S TERMS AND CONDITIONS

Beyond sheer length and complexity, how T&Cs are presented also matters. There are design tools and techniques that maximise comprehension and understanding, but there are also designs that can undermine understanding. Looking at ways that might improve young people's understanding of T&Cs, experimental research outlines five design features that significantly increased the comprehension of T&Cs[21]:

- Displaying key terms as frequently asked questions
- Using icons to illustrate key terms
- Showing customers your terms in a scrollable text box instead of requiring a click to view them
- Providing information in short chunks at the right time
- Using illustrations and comics

TikTok does not use any of these five known techniques in presenting their T&Cs.

21 Behavioural Insights Team, 2019 Best practice guide for Improving consumer understanding of contractual terms and privacy policies https://www.bi.team/wp-content/uploads/2019/07/Final-TCs-Best-Practice-Guide-July-2019-compressed.pdf and Marietjie Botes, 2017. 'Using Comics to Communicate Legal Contract Cancellation'. The Comics Grid: Journal of Comics Scholarship 7(0) p. 1

On top of this, we noted the use of design techniques that may actively work against understanding, or use 'dark patterns' in the presentation of their T&Cs[22]. In their signup process, TikTok inferred consent to the T&Cs when users choose how they wanted to signup – a seemingly unrelated action. Whereas other platforms explicitly ask new users to agree to the T&Cs, TikTok's process obfuscates that users are consenting to anything (see figure two).

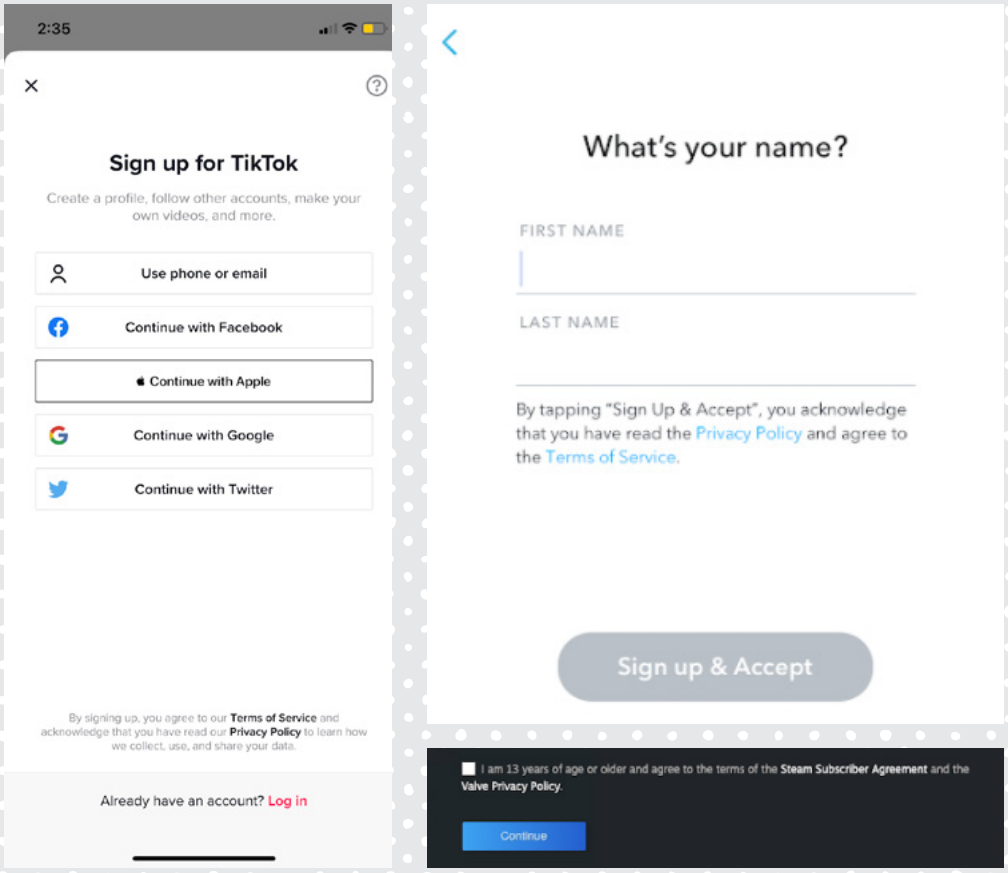


Figure two: TikTok's sign up screen, compared to other social media platforms. For example, Snapchat's sign up screen makes it clear that by signing up you are accepting T&Cs, and Steam's sign up screen which asks users to also click a separate box to indicate that users are aware that they are accepting T&Cs.

TikTok's terms and conditions are so problematic that Reset has previously awarded them 0 out of 5 stars, making them one of the worst performers among ten of Australia's most popular apps for young people[23].

22 'Dark patterns', are designs or features deployed to nudge users away from actions that align with their best interests and towards actions that are in the platform's interest. See for example, Arunesh Mathur et al 2019 'Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites' Proceedings of the ACM on Human-Computer Interaction November, pp. 81

23 Dylan Williams, Alex McIntosh, Rys Farthing 2021 Did We Consent to That? https://au.reset.tech/news/did-we-really-consent-to-this-terms-and-conditions-young-people-s-data/

DO YOUNG TIKTOK USERS FEEL THEY HAVE MEANINGFULLY CONSENTED TO THE WAY THEIR DATA IS COLLECTED AND USED?

We surveyed 238 16 and 17 year old Australian TikTok users to get a sense of how effective their T&Cs are at generating meaningful understanding and informed consent about TikTok's data collection process.

We asked survey participants if they were aware that TikTok was collecting certain types of data about them as outlined in TikTok's Privacy Policy (and summarised in Appendix 1). Young people who have offered informed, meaningful consent should be aware that TikTok is able to collect and use this data, or at least fathom a solid guess that they are. However, this is not what we found (see figure three).

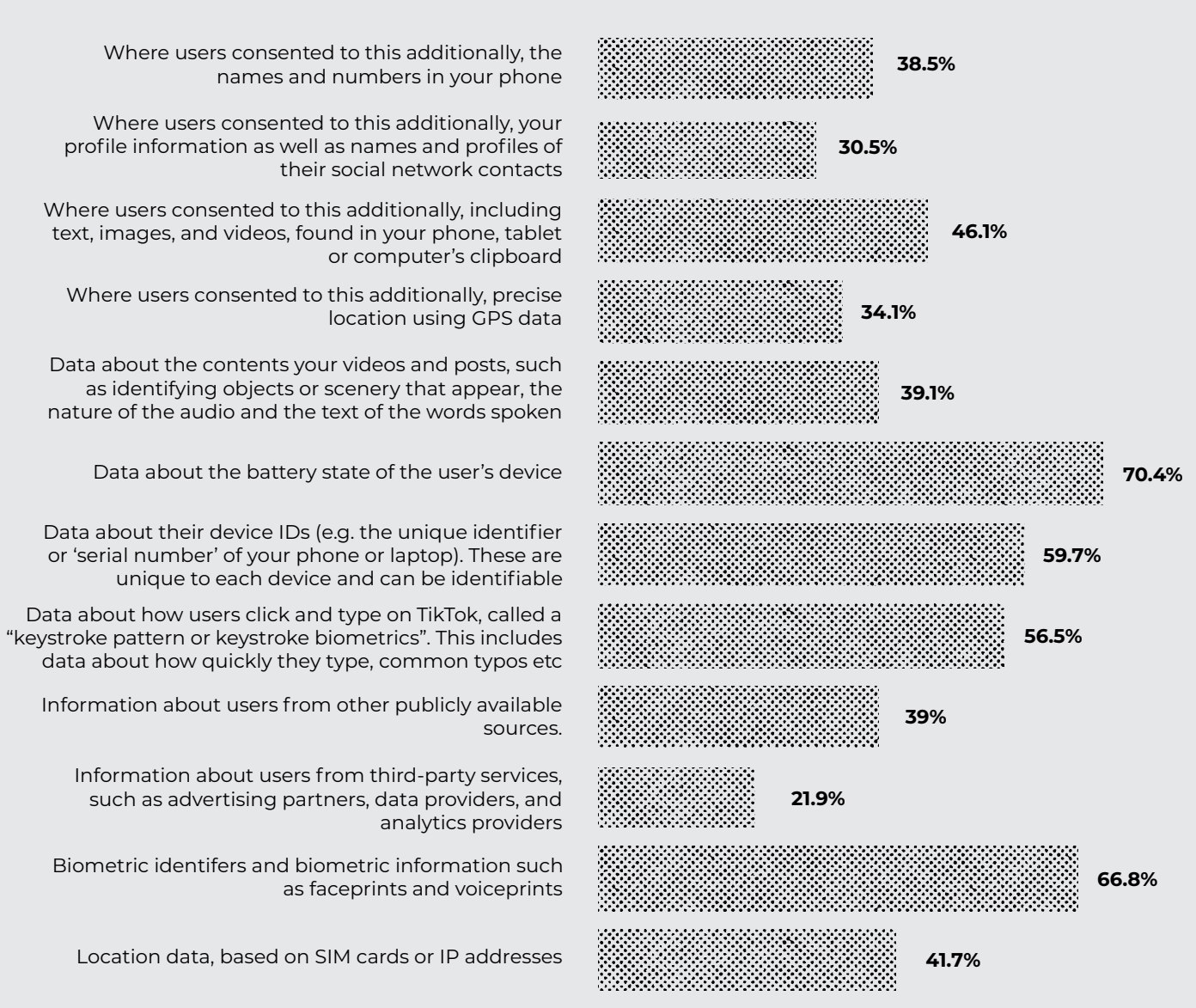


Figure 3: The percentage of young TikTok users surveyed who were unaware that TikTok was able to collect this data about users.

In total, when we asked young TikTok users if they thought they had meaningfully consented to all this data collection, an alarming 67.8% of respondents suggested that they did not think that they had offered meaningful consent.

We also found that many young TikTok users were a bit worried about TikTok's data collection practice (see figure four).

Figure 4: The percentage of young TikTok users surveyed who were worried about TikTok being able to collect this data about its users.

QUOTES FROM YOUNG PEOPLE

“WHAT THE F*”**

“THEY ARE AWESOME AT TRACKING”

“IT’S SCARY ”

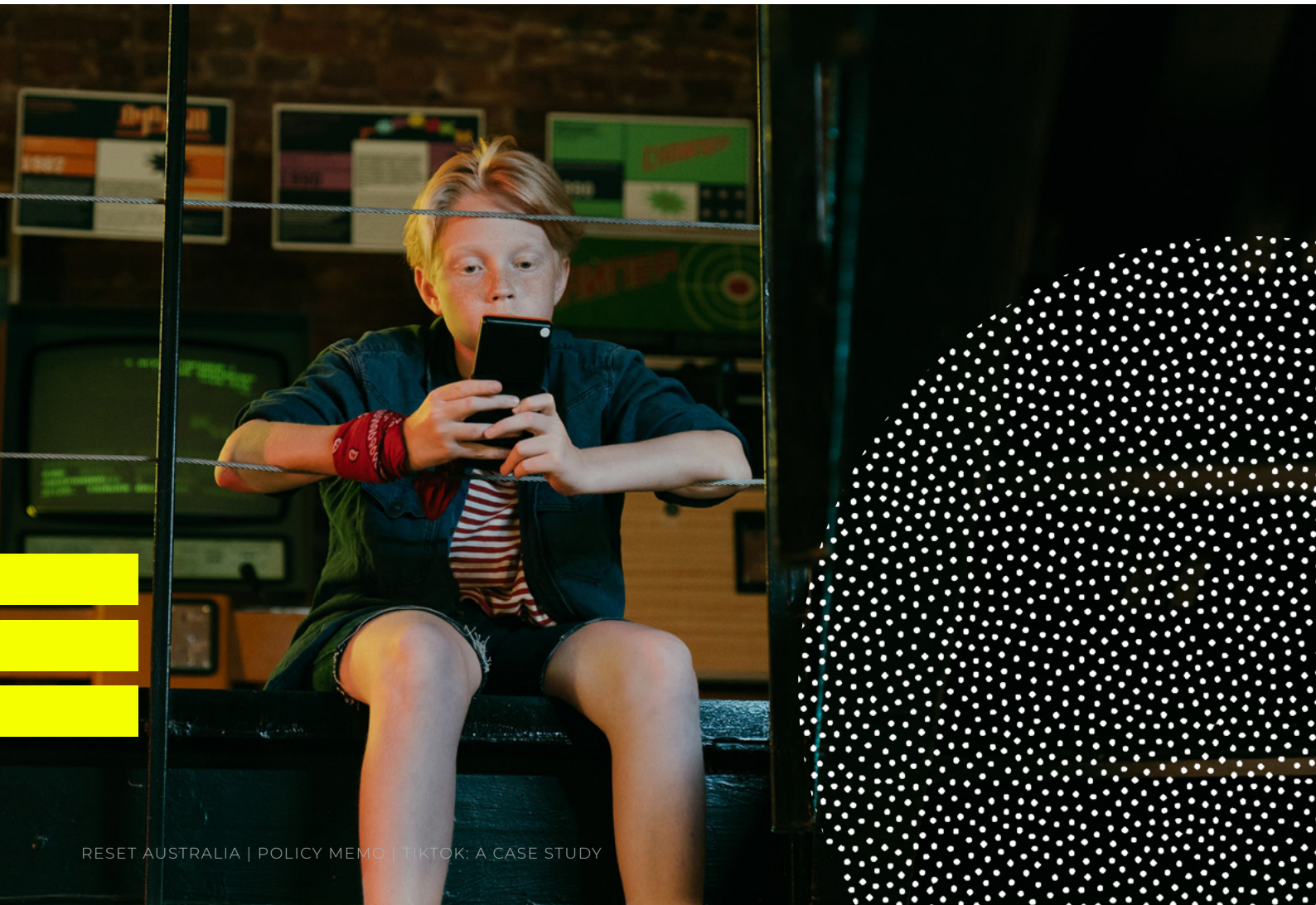
“WHY CAN THEY GET AWAY WITH THAT?”

“OVERALL I FEEL NEGATIVE ABOUT IT”

“It’s scary how much information apps can collect. Especially when there’s children as young as 9 on the platform. I no longer use Tik Tok as it’s not my thing, however I have family members that are glued to the app”

“Honestly as children we can not truely consent to our data being used in such a way, and unless legislation covers these new ways to gain information, these apps and websites will be able to continue to abuse their power over people who are unaware of the effects and problems that arise from such breaches in confidentiality”

“Everyone should assume that all of the above information is known to advertisers/ companies as soon as they own a phone or computer, or use any social media. It would be hypocritical to say I was worried about using TikTok specifically, this is a problem with the tech industry as a whole and not in any way just Tiktok. It’s okay to do this to an extent. If more people knew about this it would be a bit better”



AN ANALYSIS OF TIKTOK'S DATA FLOWS

Working with Rufposten.de we undertook a technical data flow analysis to explore what data and data types TikTok collects and sends as it is being used. To accomplish this we installed our own certificate on a rooted Android and 'tricked' the TikTok app to accept this. This allowed us to intercept and decrypt the traffic between the app and the TikTok servers on a Linux computer that was setup as a proxy.

IS YOUNG PEOPLE'S DATA USED ANY DIFFERENTLY?

While TikTok was being used in this test, the amount of data flows showed no significant difference between a test account for a 13 year old and a 30 year old's account.

WHAT DATA DOES TIKTOK SHARE?

During this test, we were able to see the TikTok app collect and send data to two third party vendors:

- Appsflyer, a marketing and analytics platform that is known for capturing attribution data (that is, tracking when people click on ads to install or open apps, if the app is already installed on the device). It also tracks how long a user keeps an app on their device.
- Facebook ads, which appeared only to be used for internal app analytics and marketing, for example to exclude existing TikTok users from receiving ads about Tiktok on Facebook. To be clear, this means that Facebook 'knows' if you're using TikTok or not.

This data can be used from Facebook[24] and Appsflyer[25] for services to their customers. In this instance, this data was transferred with the unique interoberable ID (the Android Advertising ID, or AAID) which means third party vendors can build up a personalized data profile of your online habits.

During our use, the App regularly transferred the unique Android Advertiser ID. As these IDs are unique to the device (as unique as a phone number), and they are the most commonly used and easy to compare digital identifiers, this means that every data point TikTok collected can be aggregated with data points from other apps. Again, this allows third party vendors to build up a personalised data profile of your online habits including your behaviours on TikTok.

The Android Advertiser ID, and likewise iPhone device numbers, are the main data point used by data brokerage companies. While users can reset the AAID on their phone, many users are aware of this option and even where they are reset, the old and new ID can be matched easily by unchanged IDs like an installation ID.

DOES TIKTOK WAIT FOR CONSENT BEFORE THEY COLLECT IDENTIFYING INFORMATION?

The Android Advertising ID was collected and used instantaneously. This means TikTok always knows who is using the website or app, who is sharing and viewing content, even if you do not create an account. There is no anonymous TikTok user.

24 Facebook 2020 'Business Tools Terms' https://www.facebook.com/legal/technology_terms
25 Appsflyer 2020 'Terms of Use' https://www.appsflyer.com/terms-of-use/

In our test, this meant that TikTok was collecting and using personal, identifying data before consent was obtained. (Or as previous research shows[26], even if consent is never obtained).

WERE THERE ANY VULNERABLE POINTS?

In our test, we found vulnerabilities in the way TikTok link sharing works.
If you share a link to a Tiktok video, the link always contains the User ID of who is sharing it. This means the link is not neutral to the video, as on Twitter and YouTube, but differs between different two people sharing the same video.
This means that shared content can always be traced back to who shared it, and potentially the person first sharing the video outside of the app. This poses some data protection problems and it means that an individual TikTok user's behaviours and actions can be tracked 'off the app'.

WHAT ABOUT LOCATION DATA?

From Reset's work with young people, we are aware that young people are especially concerned about how their geolocation data is collected and used so we wanted to explicitly address this. In our test, we had turned off the permission to access GPS location. We could not see GPS data collected nor shared during this time. Data about a user's time zone was collected, but this is a very broad global indicator of where in the world you are located.

HOW DOES THIS COMPARE WITH OTHER SOCIAL MEDIA PLATFORMS?

While TikTok was the focus of this test, the amount of data flows showed no significant difference from similar levels of use on Snapchat and Instagram. As the report highlights, data collection and use is an issue across the entirety of the social media business model.
A study by Citizen Lab similarly found that TikTok and Facebook collect similar amounts of data about both the user's device, and how a user interacts with the app[27].

This technical analysis was completed in partnership with Matthias Ebral from Rufposten.de.

26 Riccardo Coluccini 2021 'TikTok Is Watching You – Even If You Don't Have an Account' Vice https://www.vice.com/en/article/jgqbmk/tiktok-data-collection
27 The Citizen Lab 2021 'TikTok and Douyin Explained' https://citizenlab.ca/2021/03/tiktok-and-douyin-explained/

04. Data use practices & algorithms

EACH TIKTOK USER IS PROVIDED WITH THREE PERSONALISED CHANNELS THAT RECOMMEND VIDEO CONTENT, BUT THE FYP STREAM IS THE MAIN VEHICLE USERS USE TO DISCOVER CONTENT AND WATCH TIKTOK. WHAT APPEARS ON THIS STREAM IS DRIVEN BY A RECOMMENDER SYSTEM UNDERPINNED BY AN ALGORITHM. THIS IS THE SUBJECT OF THIS RESEARCH.

WHAT DATA ABOUT USERS DOES TIKTOK USE TO TRAIN ITS ALGORITHM AND PERSONALISE THE FYP?

Like most social media platforms, TikTok collects a trove of data about its users (see Appendix 1 for a complete overview). This includes information about what each user watches, searches and likes, in order to develop a personalised algorithm on the FYP.

To help train its recommender system, TikTok collects and uses data about[28]:

- Interactions with previous videos, such as which videos they like or share, how long they watch each video, whether they swipe to finish a video early, the accounts they follow, comments they post, etc
- The nature of any content they create themselves
- Device and account settings like language preferences, country settings, and device types
- Video information, such as details about captions, sounds, and hashtags

Videos are then ranked to determine the likelihood of a user's interest in a piece of content, and delivered to each unique 'For You Page' stream. As a result, no two FYPs will be the same. Each stream is tailored to best match the user based on the data that TikTok has collected.

28 TikTok 2020 'How TikTok recommends videos #ForYou' https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you

Like all recommender systems on social media, masses of user's personal data is collected and used to train TikTok's algorithm to serve personalised recommendations that will engage the user. Recommender systems use individual users data to do this, but also data about users they think are 'like them', to make predictive recommendations.

Recommender systems normally aim to deliver content that will maximise how long a user will stay on a platform[29]. This aim is central to the general business model of the attention economy, the longer a platform can keep a user engaged, the more opportunities they have to profit through personalised advertising. This is TikTok's primary business model. As an advertising platform, TikTok is open about this – they promote their ability to maximise reach and brand engagement via advertising in users feed[30].

So far, all of this may sound great. We've got an algorithm to predict exactly what each person will be interested in watching, with no filler nor fodder. But there's a potential dark side to this personalisation. Where users have potentially risky interests, or where the system thinks users 'like them' have potentially risky interests, the algorithm that drives their FYP could push users down a harmful rabbit hole.

But even innocuous content can lead to a questionable rabbit hole, if the recommender system thinks enough people with risky interests also like this[31].

Previous research has shown that YouTube's recommender system is capable of 'radicalising' users[32], and an investigation by Vice in 2018 found that it was possible to train TikTok's algorithm to deliver white supremacy content[33].

This research set out to see to what extent TikTok's algorithm is prepared to push young people down questionable rabbit holes. We investigate four types of harms, to see if it was possible to establish a FYP for a young person that recommended content perpetuating:

- Negative or damaging ethnic stereotypes
- Negative or damaging gender stereotypes
- COVID-19 or vaccine misinformation
- Weight loss and dieting

METHODS

This research involved three steps.

1. SET UP AND ESTABLISH FIVE NEW TIKTOK ACCOUNTS

Given the amount of data TikTok can collect from devices and linked apps, five reset phones using new SIM cards were set up (one for each risk, and one 'control' phone). TikTok was downloaded on each phone and a new account was created identifying as a 13 year old female user. Each account was asked to indicate some interests, and we selected from a range of Comedy, Gaming and either Dancing or Auto for our accounts.

29 Algorithms are proprietorial, so there is often limited public data available about the inner workings of and aims of a recommender system. However, a research paper published by Google, YouTube's parent company, outlined that their recommendation system was trained to increase 'watch time'. Paul Covington et al. 'Deep Neural Networks for YouTube Recommendations' 2016 Proceedings of the 10th ACM Conference on Recommendation Systems, ACM, New York, NY, USA

30 TikTok 2020 'The Power of TikTok' https://www.tiktok.com/business/library/the-power-of-tiktok.pdf

31 Jonas Kaiser and Adrian Rauchfleisch 2020 ' Birds of a Feather Get Recommended Together: Algorithmic Homophily in YouTube's Channel Recommendations in the United States and Germany' Social Media + Society October-December, pp 1-15

32 Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, Wagner Meira 2020 'Auditing Radicalization Pathways on YouTube' Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency Pages 131–141

33 Joseph Cox 2018 'TikTok has a Nazi Problem' https://www.vice.com/en/article/yw74gy/tiktok-neo-nazis-white-supremacy

While location or WLAN data may have indicated to TikTok that these phones were connected, or used by an organisation at a business address, this was the most effective and practical way of exploring what TikTok's recommender system would deliver to new, young users.

For safeguarding reasons, each account was created using an organisational email and were publicly labelled as research account, no avatars nor bios were included in the account, nor did we create any videos or comments. We did not contact nor were contacted by any users from these accounts. Account activity was recorded, each account deleted, and we requested a full data deletion at the end of the experiment.

2. EXPLORE THE TIME IT TOOK, AND LEVEL TO WHICH, THE FYP WOULD RECOMMEND QUESTIONABLE CONTENT TO EACH ACCOUNT

Each of the five accounts followed a similar process to assess the speed and extent to which the algorithm would deliver specific, questionable content to the account's For You Page (FYP):

- Watched 30-35 videos (roughly 15 min) of content from the FYP without any interaction, just to establish a baseline and see if any questionable content was recommended. Each video was watched to the end, and where it was questionable (or borderline) it was 'liked' and allowed to play for a second time, but otherwise, it was swiped at the end to move on to the next video.
- Over the 70-65 videos (roughly another hour of use), each account did three specific hashtag searches relevant to each harm, for example #Weightloss was searched on the account running the Weightloss experiment.
- Ten or 11 videos (depending on length) were watched from the selection recommended by each search. Where a video in the search recommendations was questionable or borderline, it was 'liked' and allowed to play for a second time, but otherwise at the end of each video we returned to the search stream. Between each of the three search cycles, we returned to the FYP and watched between 10-30 videos. We considered steps A-C to be the 'training' phase.
- After a total of 100 videos were watched by each account, we began 'swiping up' to skip videos that were not relevant to more accurately mimic real user's behaviour
- One additional search loop was undertaken at around 260 videos in, with ten or 11 videos (depending on length) watched from the selection recommended on the search channel. Where a video in the search recommendation was questionable or borderline, it was 'liked' and allowed to play for a second time, but otherwise at the end end of each video, we returned to the search stream. We then returned to the FYP to see what TikTok's algorithm was recommending.

3. ANALYSE THE CONTENTS OF WHAT THE ALGORITHM RECOMMENDED ON THE FYP OF EACH ACCOUNT.

Each video recommended in the FYP was watched in sequence and coded as either containing questionable content, borderline content or not[34]. This experiment explores TikTok's algorithm rather than the content itself, and aims to explore the dynamics and speed at which TikTok's algorithm will recommend questionable content. The nature of the experiment meant that videos had to be coded as questionable or not quickly; TikTok videos are short, and rewatching them for clarity trains the algorithm.

This introduces inevitable error and coding bias. To reduce this, the same researchers coded each test account to reduce the risk of researchers coding fewer questionable videos at the start of the experiment than the end. The assumption is that the remaining coding error is random, so the dynamics and speed of the algorithm can be inferred.

The focus of this research is not to evaluate the merits of this content, nor the context or intent of the video creators. This research aims to unpack the rate and speed at which TikTok's recommender system will deliver questionable content.

For clarity, questionable videos watched on the Search channel are not included in the figures below nor was 'borderline' content.

34 It is worth noting that this method does not evaluate co-located harms. A number of videos presented in the feed exploring negative ethnic stereotypes also presented negative gender stereotypes, and vice versa

05. Findings

HOW THE RECOMMENDER SYSTEM FARED IN PERPETUATING NEGATIVE OR DAMAGING ETHNIC STEREOTYPES

TikTok's terms and conditions are spread over 8 documents, many of which are difficult to find. Two of these are presented for review when a user joins via a mobile phone, and 6 are 'hidden' on joining (see figure one)

TikTok's algorithm is fast to 'learn' that a young user is interested in content that perpetuates negative or damaging ethnic stereotypes. After the initial three search cycles in the training phase the algorithm recommends around one in ten videos that perpetuate negative or damaging ethnic stereotypes. With one more search at 260 videos in, this rises to around one in four videos (23.1%) after watching 600 videos on the FYP stream (see figure five).

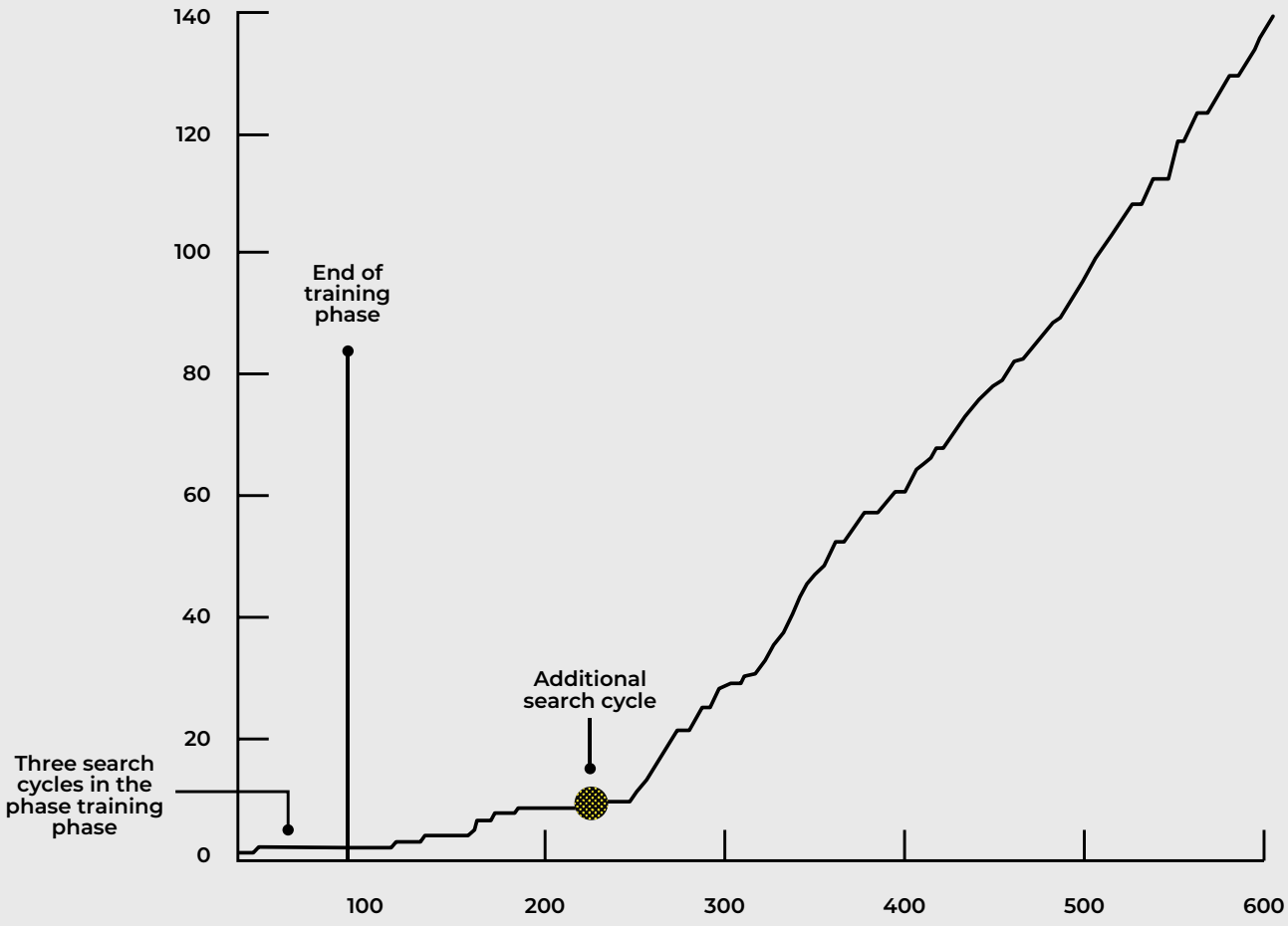


Figure 5: The count of videos that perpetuates negative or damaging ethnic stereotypes by total video count in the For You Page (FYP).

In total, to reach a FYP stream that recommended one questionable video out of every four videos, it took:

- An estimated 4 hours and 41 min
- Viewing 650 videos in total (41 in the Search stream, 609 in the FYP stream)
- Making 200 likes
  - Liking 25 videos in the search stream (For clarity, these are excluded from the harms tally)
  - Liking 175 videos in the FYP stream (of which 146 were questionable videos and 29 were borderline)
- Making four searches for problematic hashtags
- Swiping to 'skip' 352 videos before they were finished (an indication to the algorithm that you are not interested in this content).

It is worth noting that the proportion of questionable videos recommended in this experiment was still increasing after 650 videos (see figure six). After the training phase on the FYP, analysis identified that the proportion of questionable videos increased 0.04% per video watched ($R^2 = 0.98$).

If a user were to fall into a questionable ethnic stereotype rabbit hole, and repeat the same process of liking questionable videos and swiping through videos not related to this harm, without safeguards around the algorithm, it potentially could take around 2500 videos to be served up consistently questionable content.

Research suggests that young people who use TikTok are often using the platform for 80 minutes a day[35]. Based on the length of time this experiment took, this suggests that within four days, a user who is new to the platform could be pushed down a rabbit hole of potentially questionable ethnic stereotype content by the TikTok algorithm.

35 Sarah Perez 2020 'Kids now spend nearly as much time watching TikTok as YouTube in US, UK and Spain', Tech Crunch https://techcrunch.com/2020/06/04/kids-now-spend-nearly-as-much-time-watching-tiktok-as-youtube-in-u-s-u-k-and-spain/

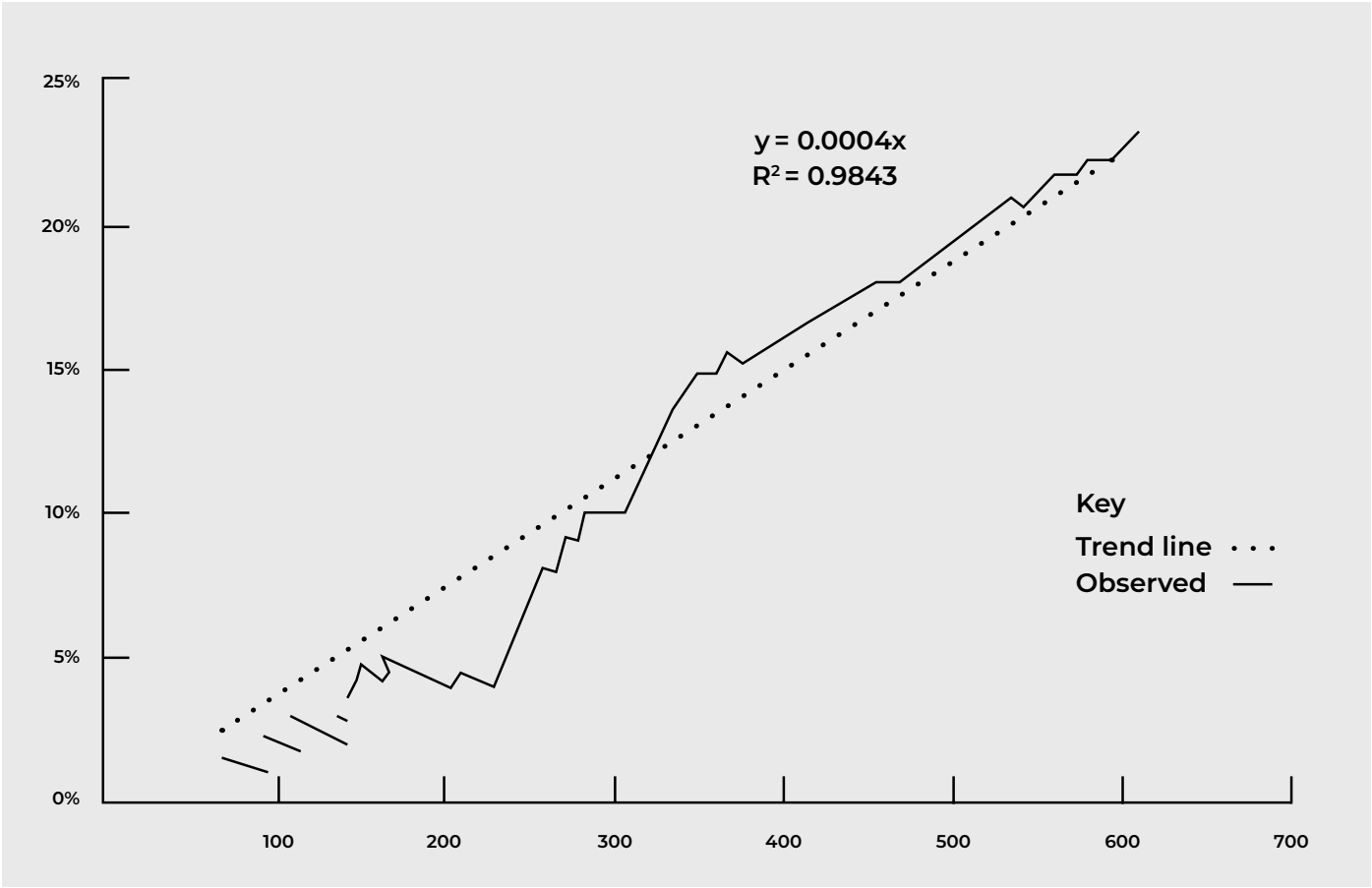


Figure 6: The proportion of videos that perpetuates negative or damaging ethnic stereotypes by total video count in the For You Page.

WHAT COUNTED AS PERPETUATING NEGATIVE OR DAMAGING ETHNIC STEREOTYPES?

This research is concerned with the algorithm driving TikTok's recommender system, rather than the content itself. However, understanding this algorithm and its potential for harm, in this experiment, requires understanding of how we coded content. Videos were counted as perpetuating negative or damaging ethnic stereotypes if they explicitly portrayed a stereotype as the 'main theme' of the video. Many of the videos in this account were short and couched as 'skits or jokes'. Where the butt of the joke was a ethnic stereotype, it was counted as questionable . However, this category more than any other, presented a lot of borderline content. Where it was unclear, it was not counted as a harm but was noted as potentially questionable (and liked to 'train' the algorithm). The focus of this research is not to evaluate the intent nor merits of this content, but the algorithm that recommends it.

The sorts of videos counted as questionable included, for example:

- A disturbing video of a young woman being driven home after a date with a man of a particular ethnicity. The conversation becomes increasingly uncomfortable in ways that reinforce ethnic stereotypes, and ends with the young woman using an ethnic slur against the young man.
- A skit video joking about the propensity for violence among a particular ethnicity. In this video, a young man who is visible and audible caricatured as belonging to a particular ethnicity accidentally walks into an inanimate object . He then proceeds to pick a fight with the object.
- A video parodying how often young men of a particular ethnicity end up in court

Borderline videos were counted as videos that may have perpetuated an ethnic stereotype, but not clearly in a troubling or negative way. For example, one video was titled 'what's in every Italian's garden', which showcased an extensive and verdant vegetable patch. Borderline videos were liked to train the algorithm only, and are not included in the count of questionable videos.

HOW THE RECOMMENDER SYSTEM FARED IN PERPETUATING NEGATIVE OR DAMAGING GENDER STEREOTYPES

TikTok's algorithm is equally fast to learn that a young user is interested in content that perpetuates negative or damaging gender stereotypes. After the initial three search cycles in the training phases, the algorithm recommends around one in twenty videos that perpetuate negative or damaging gender stereotypes. With one more search at 260 videos in, this rises to around one in five videos (21.2%) after watching 600 videos on the For You Page (see figure seven).

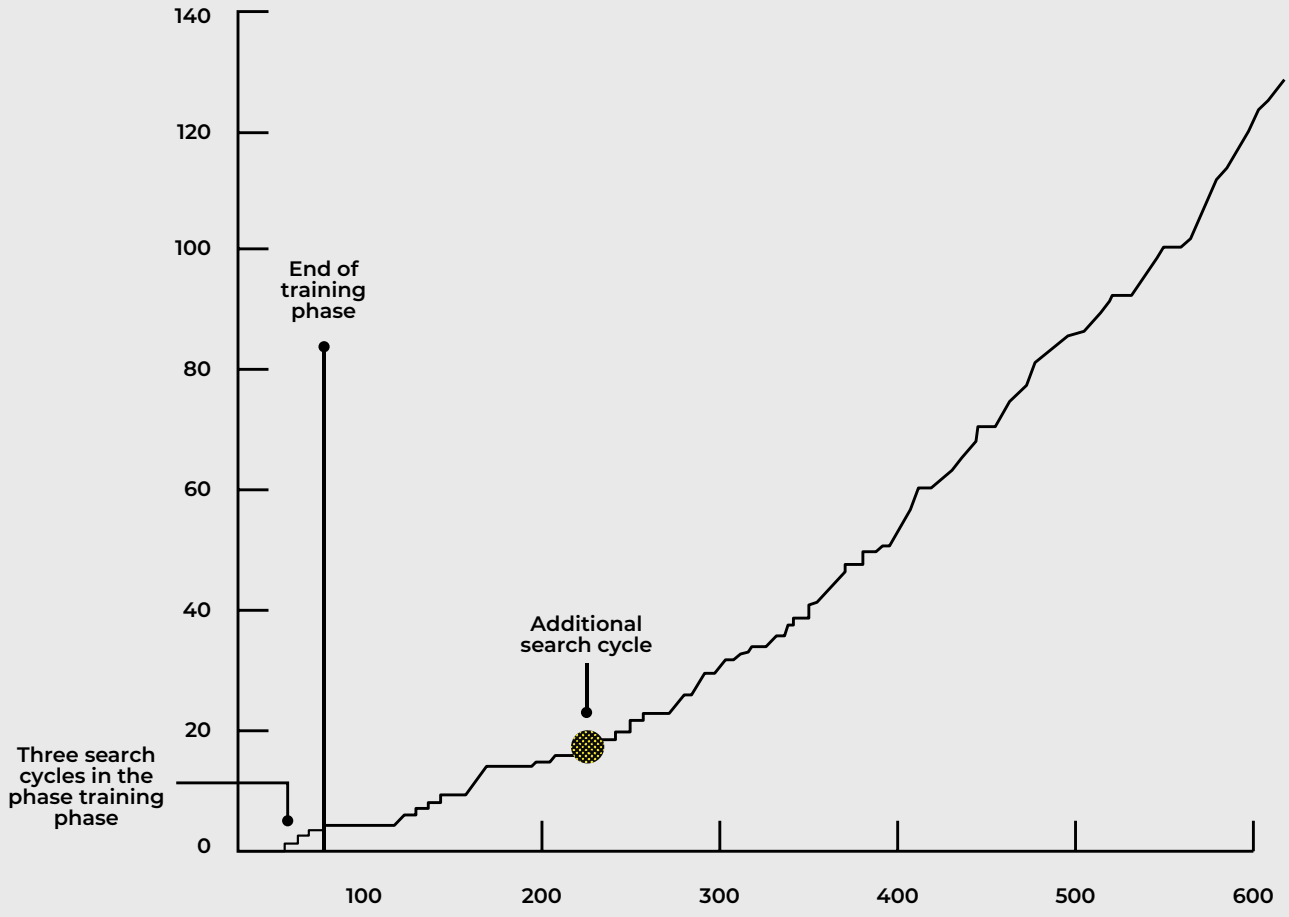


Figure 7: The count of videos that perpetuates negative or damaging gender stereotypes by total video count in the For You Page.

In total, to reach a FYP that recommended one questionable video of five videos, it took:

- An estimated 7 hours and 42 min
- Viewing 650 videos in total (42 in the Search stream, 608 in the FYP)
- Making 240 likes
  - Liking 32 videos in the search stream (For clarity, these are excluded from FY harms tally)
  - Liking 208 videos in the FYP (of which 129 were questionable videos and 79 were borderline)
- Making four searches for problematic hashtags (and watching 42 videos in the resulting recommendations stream)
- Swiping to 'skip' 272 videos before they were finished (an indication to the algorithm that you are not interested in this content).

It is worth noting that the proportion of questionable videos recommended in this experiment was still increasing after 650 videos (see figure eight). After the training phase on the FYP analysis identified that the proportion of questionable videos increased 0.04% per video watched ($R^2 = 0.99$).

If a user were to fall into a negative gender stereotype rabbit hole, and repeat the same process of liking questionable video and swiping through videos not related to this harm, without safeguards around the algorithm, it potentially could take around 2500 videos to be served up consistently questionable content. Based on usage patterns in the US, Spain and the UK[36], based on the length of time this experiment took, this would be between 5-6 days of use.

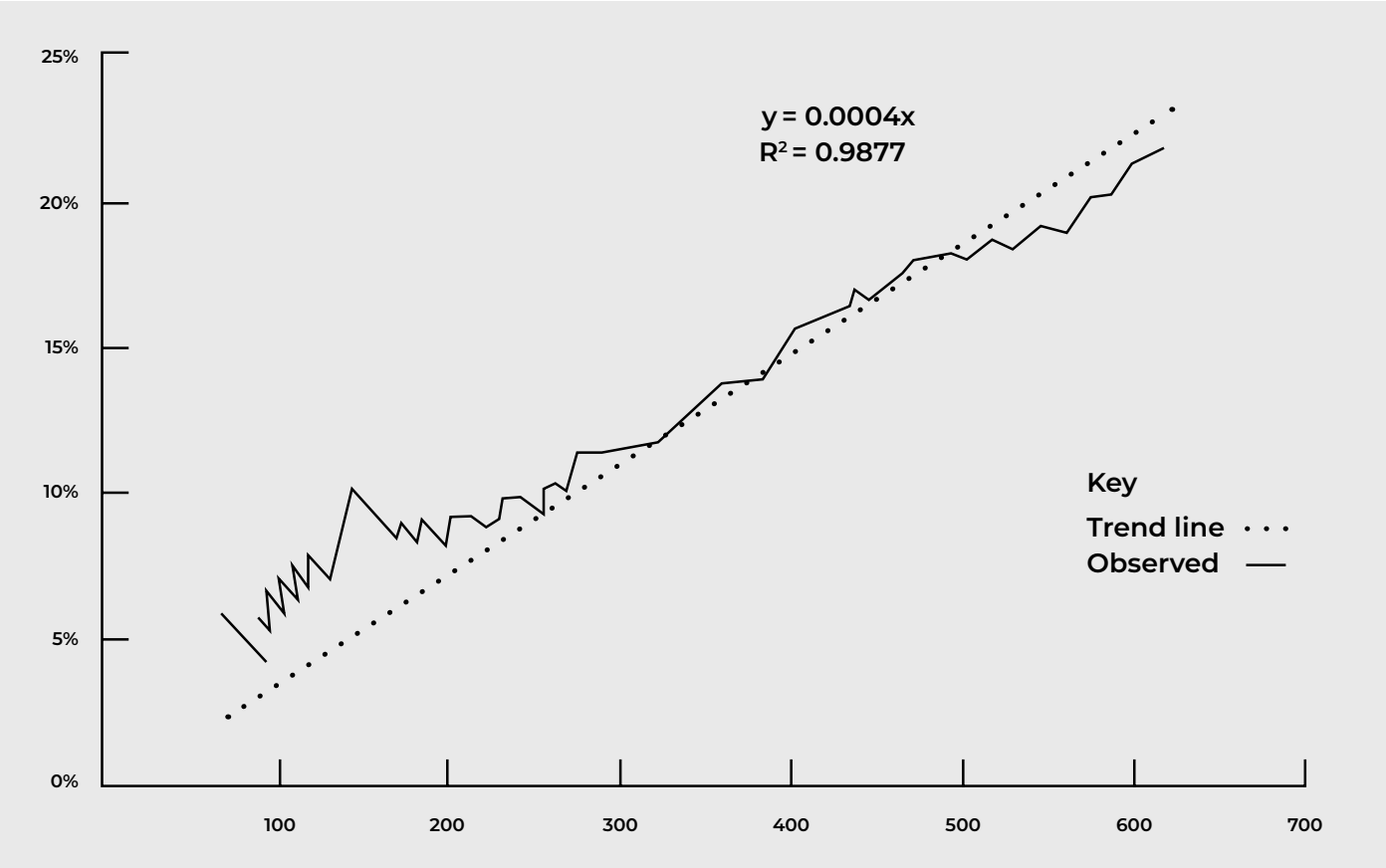


Figure 8: The proportion of videos that perpetuates negative or damaging gender stereotypes by total video count in the For You Page.

WHAT COUNTED AS PERPETUATING NEGATIVE OR DAMAGING GENDER STEREOTYPES?

This research is concerned with the algorithm driving TikTok's recommender system, rather than the content itself. However, understanding this algorithm and its potential for harm, in this experiment, requires understanding how we coded content. Videos were counted as perpetuating negative or damaging gender stereotypes if they explicitly portrayed a gender stereotype as the 'main theme' of the video. Many of the videos in this stream were longer and in the style of monologues, speeches, or snippets from talk shows. Repeated themes included women ruining men's lives (most often financially), justifying violence against women as a means of redressing the perceived double standard and 'naturalistic' or religious arguments about women's rightful place as being beneath men.

36 Sarah Perez 2020 'Kids now spend nearly as much time watching TikTok as YouTube in US, UK and Spain', Tech Crunch https://techcrunch.com/2020/06/04/kids-now-spend-nearly-as-much-time-watching-tiktok-as-youtube-in-u-s-u-k-and-spain/

THE SORTS OF VIDEOS COUNTED AS QUESTIONABLE INCLUDED:

- A clip of a podcast talk show, about why men should never 'invest' time nor money in women because women are illogical and are only after men's money
- A selfie style clip from a young man who was extremely angry that women use the Twitch platform (a game streaming platform). He claims all female Twitch users are only there to show off their bodies and attract male attention, not because they like video games
- A clip from a Ted style talk, where a speaker suggests that modern women have forgotten their subservient place in society

Borderline videos were counted as videos that may perpetuated a gender stereotype, but not clearly in a troubling or negative way. As an example there was a stitch video (or a video stitched together from two different creators) about how different genders shower. The first clip was a boy explaining what boys do in the bathroom (pretend to be aqua man in the shower), and asking if girls do that too. The second stitched clip was a young woman saying girls absolutely do not play aqua man in the shower. Borderline videos were liked to train the algorithm only, and are not included in the count of quest

HOW THE ALGORITHM FARED IN PROMOTING COVID-19 AND VACCINE MISINFORMATION

In this experiment, we could not train TikTok's FYP algorithm to recommend content containing Covid-19 or vaccine misinformation, despite following the same methodology. This could be an experimental failure, or it could be because TikTok's FYP algorithm has parameters coded into it that prevent it from doing so.

We repeated four search cycles, watching 40 videos across various Covid-19 and vaccine misinformation hashtags. 15 of these were tagged as questionable. However, liking and rewatching these videos did not appear to train the algorithm to deliver Covid-19 or vaccine misinformation[37]. This experiment was abandoned after 400 videos. Of the 360 videos watched in the FYP stream, none were counted as promoting Covid-19 and vaccine misinformation.

37 This does not mean that this Covid-19 or vaccine misinformation does not exist on TikTok, indeed we could find it through the Search function and on the search stream. It may also be possible that this content could appear in the algorithm in the 'Followers' stream.

HOW THE ALGORITHM FARED IN PROMOTING DIETING AND WEIGHT LOSS

In this experiment, we could not train TikTok's FYP algorithm to recommend content that promoted dieting or weight loss, despite following the same methodology. This could be an experimental failure, or it could be because TikTok's FYP algorithm has parameters coded into it that prevent it from doing so.

To 'really push it', in this experiment we repeated seven search cycles in total, watching 71 videos across various weight loss hashtags. 65 of these were tagged as questionable. However, liking and rewatching these videos did not appear to train the algorithm to deliver up dieting content[38].

This experiment was abandoned after 404 videos in total. Of the 333 videos watched in the FYP stream, three were counted as promoting dieting and weight loss.

WHAT DO TIKTOK'S OWN POLICIES SAY ABOUT THIS SORT OF CONTENT?

TikTok's community guidelines set out the rules for how content is moderated on the platform. They outline specific responses around removing content that includes:

- Violent extremism, such as content that threatens or encourages violence against individuals or organisations
- Illegal activities and regulated goods, such content that violates laws or depicts criminal activity
- Violent and graphic content, such as content that is gratuitously shocking, graphic, sadistic, or gruesome or that promotes, normalizes, or glorifies extreme violence or suffering
- Suicide, self-harm and dangerous acts, such as content depicting, promoting, normalizing, or glorifying activities that could lead to suicide, self-harm, or eating disorders
- Harassment and bullying, including abusive content or behavior that can cause severe psychological distress
- Adult nudity and sexual activities, including nudity, pornography, or sexually explicit content
- Under 18s (minor) safety such as content, including animation or digitally created or manipulated media, that depicts abuse, exploitation, or nudity of minors

38 This does not mean that this content does not exist on TikTok, indeed we could find it through the Search function and on the search stream. It may also be possible that this content could appear in the algorithm in the 'Followers' stream.

- Integrity and authenticity, such as content or accounts that involve spam or fake engagement, impersonation, misleading information that causes harm. This would include Covid Misinformation
- Platform security, such as content that includes viruses or promotes hacks for TikTok
- Hateful behaviours. TikTok's guidelines stipulate that they will remove any content and suspend any accounts that are associated with hate speech. Specifically, they outline that:

> *"TikTok is a diverse and inclusive community that has no tolerance for discrimination. We do not permit content that contains hate speech or involves hateful behavior and we remove it from our platform. We suspend or ban accounts that engage in hate speech violations or which are associated with hate speech off the TikTok platform.*
>
> *Attacks on the basis of protected attributes.*
>
> *We define hate speech or behavior as content that attacks, threatens, incites violence against, or otherwise dehumanizes an individual or a group on the basis of the following protected attributes:*
>
> - *Race*
> - *Ethnicity*
> - *National origin*
> - *Religion*
> - *Caste*
> - *Sexual orientation*
> - *Sex*
> - *Gender*
> - *Gender identity*
> - *Serious disease*
> - *Disability*
> - *Immigration status"*

Although hate speech on the basis of sex, gender, gender identity, race, ethnicity, and national origin are specifically identified by TikTok in their policy, our experiment suggests that questionable content around gender and ethnicity is both hosted by the platform, and recommended to users by their algorithm.

All videos we counted as questionable were reported to TikTok at the end of this experiment, so they could be reviewed against this hate speech policy.

06. Conclusion

LIKE ALL SOCIAL MEDIA PLATFORMS, TIKTOK COLLECTS AND USES DATA ABOUT ITS USERS IN MULTIPLE WAYS.

As this report outlines, some of these are problematic. It is not always clear if young people have offered meaningful, informed consent to the collection and use of all that data when they clicked through the terms and conditions. TikTok also collects and uses identifying information before a young person has agreed to their terms and conditions.

It is also clear that TikTok will use data about young people's activity on the app to train their recommender system to deliver questionable content, which can be in violation of their own community guidelines.

However, when it comes to curating their algorithm, this research suggests it may be possible that TikTok are actively taking steps to limit recommending Covid-19 misinformation and weight loss tips (or that our experiment failed in these instances).

While limiting the amplification of Covid-19 misinformation or weight loss tips to young people is a positive algorithmic intervention, decisions about what is and isn't amplified are made at TikTok's discretion. These same mechanisms could be easily deployed to limit the amplification of videos about politics or protest movements. Without public transparency into how TikTok codes their algorithm to make these decisions, it is impossible to know what information is being lifted up and pushed down—and what consequences that could have for our society.

It cannot be up to TikTok alone to decide what to and not to amplify, what levels of transparency they want to offer, nor what data can and can't be collected and used. Self regulation simply isn't working. Stronger regulations and independent oversight are needed to ensure children and young people's data is collected and used in ways that are in their best interests.

06. Recommendations

Australia needs a regulatory code governing how children and young people's data can be collected and used. Other countries have implemented or are proposing similar codes already, including the UK's Age Appropriate Design Code, and Ireland's Fundamentals for a Child-Oriented Approach to Data Processing. Australia's young people deserve the same, if not better, protections.

RESET AUSTRALIA IS CALLING FOR A CODE THAT:

1. **Creates the best and safest digital world for young people.**

- Put children's best interest at the heart of decision making about their data. This would prohibit the use of children's data for detrimental reasons, such as using it to train algorithms to deliver harmful content
- Put children and young people in control of their data by requiring:
  - Expressed consent - Only process data when children (and parents) have meaningfully consented, except in their best interests. This would place a requirement on digital service providers to ensure their terms and conditions enabled and facilitated informed consent, and they did not infer consent nor use dark patterns to solicit it.
  - Transparency and accountability - Children (and parents) should know every time their data is processed, except in their best interests
  - Data minimisation and restricted data sharing - Only collect the data you really need, and don't share it, except in their best interests

2. **Is overseen by a strong and enabled regulator**

- Enforced by a regulator well resourced to oversee these new responsibilities
- Can issue meaningful penalties that match the scale of any breach
- For extreme violations, there could be the option of criminal sanctions

3. **Aligns with the Online Safety Bill and Basic Online Safety Expectations**

- Australia has some world leading legislation around eSafety, takedown and moderation. A code must join up seamlessly with this legislation
- Take a similar systemic focus

07. Glossary

RECOMMENDER SYSTEM:

A recommender system, or a recommendation system is a type of information filtering system that seeks to predict the "rating" or "preference" a user would give to an item. There are many ways to develop a recommender system[39]. Traditionally, in print media, the 'system' would be an informed, human, decision-making process from an editorial team, which would decide what went on the front page of a newspaper. On User Generated Content (UGC) social media platforms, recommender systems are usually created from algorithms trained by a user's data.

DARK PATTERN:

'Dark pattern', are designs or features deployed to nudge users away from actions that align with their best interests and towards actions that are in the platform's interest[40].

39 Wikipedia 2021 'Recommender Systems' https://en.wikipedia.org/wiki/Recommender_system

40 Arunesh Mathur et al 2019 'Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites' Proceedings of the ACM on Human-Computer Interaction November, pp. 81

08. Appendix

Platform	Document Name	Flesch Reading ease score	Word count	Estimated read time (Min:Sec)
Tik Tok	Terms of Service	39.8	28880	128:21
Tik Tok	TikTok Platform Cookies Policy	48.83	2017	8:57
Tik Tok	Open sources software notices	39.94	16057	71:21
Tik Tok	Virtual Items Policy	49.15	7209	32:02
Tik Tok	Law Enforcement Data Request Guidelines	36.76	2352	10:27
Tik Tok	Privacy Policy	47.23	14528	4,200
Tik Tok	Intellectual Property Policy	34.46	1302	5:47
Tik Tok	Privacy Policy for Younger Users	37.64	756	3:21

DATA CREATED, COLLECTED AND ABOUT ITS USERS[41]

DATA CREATED, COLLECTED AND USED ABOUT USERS FROM THE TIKTOK PLATFORM ITSELF:

Data collected about users by default

- Location data, such as information based on your SIM card or IP address. If you consented, TikTok can also collect precise location GPS data
- Biometric identifiers and biometric information such as faceprints and voiceprints. In countries covered by up-to-date, relevant privacy laws, TikTok will seek any permissions from users prior to collection
- Registration information, such as age, username and password, language, and email or phone number
- Profile information, such as name, social media account information, and profile image
- Any information provided to verify an account such as proof of identity or age
- Any information sent in correspondence to TikTok
- Any information shared through surveys or participation in challenges, sweepstakes, or contests such as gender, age, likeness, and preferences

Data collected about users where additional consent is obtaine

- Content found in their phone, tablet or computer's clipboard, including text, images, and videos
- A user's phone and social network contacts. TikTok will:
  - access and collect the names a nd phone numbers and match that information against existing TikTok users
  - collect a user's public profile information as well as names and profiles of their social network contacts

DATA CREATED, COLLECTED AND USED ABOUT USERS BY TIKTOK FROM OTHER SOURCES:

- Any other social media and login services used, if a user has linked or signed up using a third-party social network or login service (such as Facebook, Twitter, Instagram, or Google). TikTok may collect information from these services, including a user's network or contact lists on those platforms, and information relating to their use of that platform
- TikTok may collect information about users from third-party services, such as advertising partners, data providers, and analytics providers
- TikTok may have data about users from other TikTok users, if they provided information through customer service inquiries for example
- TikTok may collect information about you from unspecified 'other publicly available sources'

DATA CREATED, COLLECTED AND USED ABOUT HOW USERS USE TIKTOK:

- Data about how users click and type on TikTok, called a "keystroke pattern or keystroke biometrics"[42]. This includes data such as when a user hovers over a video, how quickly they swipe on a video, how quickly they type etc
- Data about the device users use. This includes:
  - Device IDs (e.g. the unique identifier or 'serial number' of your phone or laptop). These are unique to each device and are identifiable
  - IP address of where you logged in, which is a broad geolocation tool

41 TikTok 2021 Privacy Policy https://www.tiktok.com/legal/privacy-policy?lang=en

42 These are considered a class of behavioural biometrics that may in the future be identifiable. (Currently, they only allow analysts to explore if users are right or left handed for example).

- Model of your device (e.g. iPhone, Samsung phone, type of computer)
- The screen resolution of the device
- Mobile carrier (e.g. phone provider like Optus or Telstra)
- The battery state of a device (e.g. is a user on full charge or 10%)
- Identifiers for advertising purposes
- Time zone settings
- User agent (e.g. web browser or email reader used)
- Network type (e.g. did you log in from a local area network (LAN) or personal modem)
- Device system and operating system (e.g. iOS 14 or Android 11)
- App and file names and types (e.g. what apps you log in and connect through.
- Audio settings
- Information about any connected audio devices, such as which headphones are users connected to

◆ If you log-in from multiple devices, TikTok will be able to use this information to identify and track your activity across devices.

◆ Data about which content, web pages and advertisements users click on, and track their use across TikTok (this may use cookies and other tracking technologies like web beacons, flash cookies, pixel tags)

◆ TikTok and their partners may create data by linking a users' contact or account information across all an individual user's devices, using email or log-in or device information. TikTok's partners may in turn use this information to display ads on TikTok and elsewhere online.

◆ Payment information, including payment card numbers or other third-party payment information (such as PayPal)

◆ A user's opt-in choices and communication preferences

DATA CREATED, COLLECTED AND USED ABOUT WHAT USERS POST ON TIKTOK:

◆ Any user-generated content itself. This includes any videos posted, but also all comments, searches, photographs, live streams, audio recordings, virtual item videos and hashtags you add[43]. If users apply a filter or effect to a piece of content, TikTok may collect both the affected content and the original content

◆ Data created by analysing a user's content, such as identifying objects and scenery that appear, the existence and location within an image of face and body features and attributes, the nature of the audio and the text of the words spoken. This is used to enable special video effects but also for content moderation, demographic classification to create personalized content and ad recommendations, and for other non-personally-identifying operations.

◆ Data about messages, and when users compose, send, or receive messages on TikTok. That information includes the content of the message and data about when the message has been sent, received and/or read, and who sent and received it

METADATA CREATED, COLLECTED AND USED BY TIKTOK:

◆ TikTok creates, collects and use metadata that can describe how, when, where, and by whom content was created, collected, modified and how that content is formatted.

43 This is regardless of whether users change their mind about posting content: TikTok collects data pre-loading at the time of creation, import, or upload, regardless of whether a users goes on to upload that content or not.

Reset.
AUSTRALIA

Reset Australia is an independent organisation raising awareness and advocating for better policy to address digital threats to Australian democracy.

hello@au.reset.tech | au.reset.tech

Supported by **Minderoo Foundation** and **Luminate**