



Sex education against the algorithm: the algorithmically enforced deplatforming of YouTube sex edutainment

Lisa Garwood-Cross, Ben Light, Anna Mary Cooper-Ryan & Cristina Vasilica

To cite this article: Lisa Garwood-Cross, Ben Light, Anna Mary Cooper-Ryan & Cristina Vasilica (08 Jul 2024): Sex education against the algorithm: the algorithmically enforced deplatforming of YouTube sex edutainment, Journal of Gender Studies, DOI: [10.1080/09589236.2024.2374752](https://doi.org/10.1080/09589236.2024.2374752)

To link to this article: <https://doi.org/10.1080/09589236.2024.2374752>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 08 Jul 2024.



Submit your article to this journal [↗](#)



Article views: 84



View related articles [↗](#)



View Crossmark data [↗](#)

RESEARCH ARTICLE



Sex education against the algorithm: the algorithmically enforced deplatforming of YouTube sex edutainment

Lisa Garwood-Cross , Ben Light , Anna Mary Cooper-Ryan  and Cristina Vasilica 

School of Health and Society, University of Salford, Manchester, UK

ABSTRACT

Deplatforming of sexual content has increased across social media, usually operationalized by commercially charged, and algorithmically enforced, platform policies. This paper extends work on the algorithmic deplatforming of sex through the case study of how sex education content, or 'sex edutainment' on YouTube, is impacted by the platform's algorithmic structures. Enrolling actor-network theory, we demonstrate the delegation work YouTube enacts through algorithms by presenting empirical findings from a multi-method study examining the assemblage of YouTube, sex edutainment influencers, and young people. The findings highlight that despite YouTube's curated platform imaginary as an amplifier of voices, algorithmic delegation of platform governance creates significant barriers for influencers creating sex edutainment on YouTube. Although not contravening YouTube's policies, influencers regularly battle demonetization, age-restrictions and algorithmic bias. This undermines the benefits of sex edutainment by limiting access to content, creates precarious financial environments for influencers and risks erosion of audience trust. Meanwhile, algorithms designed to protect users enact this governance without adequately protecting influencers themselves, predominantly women and LGBTQ+ individuals, from harm. Despite YouTube's veneer of democratization, the social discourses and protectionist narratives that have destabilized traditional sex education efforts permeate our digital environments and can be seen in the algorithmic enactment of YouTube's policies.

ARTICLE HISTORY

Received 26 December 2023
Accepted 20 June 2024

SUSTAINABLE DEVELOPMENT GOALS

SDG 3: Good health and well-being; SDG 4: Quality education

KEYWORDS

Algorithms; platform governance; sex education; YouTube; influencers

Introduction

Sex education in various parts of the world has been destabilized and undermined by persistent moral panics and public anxiety over the moral wellbeing and safety of young people (Bialystok & Andersen, 2022; Herdt, 2009). This has led to protectionist discourses and policies that have impacted young people's access to sex education which meets their needs (Bialystok & Andersen, 2022; Ringrose, 2013). Given this, some young people, and those who seek to educate them, have been drawn to social media (Perez (2021), including YouTube (Johnston, 2017), as an alternative means of consuming and sharing information about sex, relationships and sexual health. As YouTube has curated a platform imaginary (van Es & Poell, 2020) of democratization and amplifying user voices, it appears to be well suited for sharing comprehensive, entertaining and intersectional educational content about sex and relationships. As such, YouTube provides a possible avenue for sex edutainment, content that merges educational messaging about sex with entertainment

CONTACT Lisa Garwood-Cross  L.J.Garwood-Cross@salford.ac.uk  School of Health and Society University of Salford, Mary Seacole Building, Frederick Road Campus, Salford MS2.50, UK

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

products such as televisual content, magazines and online video (Johnston, 2017; Lim et al., 2019; McKee, 2017; McKee et al., 2018; Paasonen & Saarenmaa, 2023).

However, existing literature has identified the void between YouTube's claims of giving everyone a voice (YouTube, 2021a) and their competing priorities as a corporate platform (Gillespie, 2010). To manage this disconnect, YouTube have been found to utilize algorithmic moderation as part of their governance strategies (Caplan & Gillespie, 2020; Rodriguez, 2023). As an additional concern, sexually related content is increasingly being regulated and deplatformed on social media (Paasonen et al., 2019; Pilipets & Paasonen, 2022; Are, 2021), which raises concerns about the stability of YouTube as a platform for sex education.

This paper provides a unique contribution to the extensive literature around platformed algorithmic governance, and the discriminatory biases algorithms enact, by providing an empirical case study of their application, in YouTube, in relation to sex education. We do so by exploring the role of algorithmic deplatformization of sex through analysis of findings from a multi-method study rooted in Actor-Network Theory that sought to understand the possibilities and problems of YouTube sex edutainment content made by influencers.

To contextualize this case study, we begin by providing a short overview of the challenges faced by, and limitations of, 'traditional' classroom-based sex education to identify why and how YouTube sex edutainment has come into being. We then engage with what is already known about YouTube's platform governance and the role of algorithms in their management of content, before turning to our case study, which demonstrates how these issues play out in relation to sex edutainment content on YouTube, and the harms and inequalities this creates and perpetuates.

From sex education to YouTube sex edutainment

Sex education is a crucial means of providing early, and sometimes the only, education young people receive regarding issues related to sex, sexuality, gender, and relationship dynamics, making it a key frontier for many of the issues that concern us in gender studies (Grose et al., 2014).

However, in many anglosphere countries, sex education has suffered from being caught in the midst of political and ideological battlegrounds which cause moral panics over the protection of innocence (Bialystok & Andersen, 2022; Herdt, 2009). These protectionist discourses have led sex education curricula to predominantly focus on reforming the sexual behaviour of young people by focusing on risk narratives around reproductive function, the prevention of pregnancy and sexually transmitted infections (Bialystok & Andersen, 2022; Kantor & Lindberg, 2020; Lenskyj, 1990; Ringrose, 2013; Shannon, 2016). This limited view of sexuality neglects the reality of young people's experiences, concerns and questions (Bauer et al., 2020; Kantor & Lindberg, 2020; Pound et al., 2016) and can lead some young people to develop shame about their sexual selves (Irvine, 2009; Shannon, 2016).

A review of 55 studies on young people's perspectives of sex education from the UK, Ireland, USA, Australia, New Zealand, Canada, Japan, Iran, Brazil and Sweden found that school-based sex education is often '*negative, gendered and heterosexist*' (Pound et al., 2016, p. 1), and lacks information young people want about female pleasure and LGBTQ+ topics. However, as progress in reforming sex education in these school-based settings is slow (Bialystok & Andersen, 2022; Pound et al., 2016), internet interventions have been considered as an alternative method of disseminating sex education, due to benefits including anonymity, informality and interactivity (Oosterhoff et al., 2017). The ability to tailor online content may also be an opportunity to fill gaps in sex education provision for marginalized individuals and underrepresented topics including female pleasure and LGBTQ+ identities. Given this, and young people's high rates of social media usage, some have taken to creating informal, entertaining sex education, or sex edutainment, content on social media platforms, including YouTube (Johnston, 2017).

YouTube presents an identity that is based around amplifying the voices of their users, given this platform imaginary it is not surprising that content creators have been drawn to the platform to

share sex education content, particularly to meet the needs of individuals traditionally excluded from sex education.

Johnston (2017) highlights that YouTube sex edutainment videos utilize techniques including informal language and direct address to camera, delivered by 'Friendly and engaging' (p.77) content creators, or influencers, who act like a 'cool older friend' (p.77) discussing sex and relationships without embarrassment. Videos cover a range of topics including discharge, LGBTQ+ sex, consent, vaginal dryness, STIs, intimate partner violence, virginity, sexual identity, contraceptives and pleasure. Influencers in this niche are predominantly women or LGBTQ+ individuals, some have relevant professional training in sexual health while others start as passionate amateurs about sex education. Many create intersectional content linked to LGBTQ+ relationships, disability or cultural/religious beliefs. Some of these content creators, having grown a following, monetize content to remunerate their labour and professionalize on the platform (Johnston, 2017), which follows patterns within the wider influencer industry to make content creation a full or part-time job (Abidin, 2017; Cunningham & Craig, 2017). Johnston (2017) suggests that YouTube sex edutainment provides not only an extension of school sex education but a resource and community to provide new modes of education and belonging around sexuality for young people.

YouTube's platform governance

However, whilst YouTube utilizes participatory and co-creative rhetoric about their platform (Burgess & Green, 2009), it is primarily a for-profit business venture (Beer, 2017; Caplan & Gillespie, 2020). As such, YouTube semantically positions itself using the rhetoric of 'platform' to straddle its competing priorities between audiences and advertisers (Gillespie, 2010), with the seemingly democratic aspects of the platform in tension with YouTube's commercial interests (Rodriguez, 2023). Whilst YouTube may allow the bypassing of traditional media or institutional governance, YouTube itself has become institutionalized in shaping celebrity and content to meet its commercial aims through its own forms of governance; algorithmic curation, community guidelines and terms of service (Hou, 2019; Ørmen & Gregersen, 2023). Therefore, whilst YouTube supports users to create content, the platform's 'patronage' comes with conditions and controls (Burgess et al., 2020; Caplan & Gillespie, 2020). Ultimately, although a rhetoric of empowerment and democratization are commonly touted by platforms, there is a more complex balance of power at play, as power structures are often hidden, acting from within the frameworks of our software within the algorithms, making them challenging to unmask (Beer, 2009).

Algorithms, although largely imperceptible, have become increasingly ingrained in our lives (Beer, 2017; Willson, 2017). Algorithms are often closely guarded by platforms, with processes such as machine learning and relational databases allowing them to be changed rapidly and invisibly in continuous evolution (Gillespie, 2014). Yet, this speed of change, alongside platforms intentional cloaking and modification of algorithmic factors to discourage users from manipulating them, make studying their inner workings challenging (Petre et al., 2019). This leads to influencers being caught in a 'visibility game' (Cotter, 2019), trying to negotiate how to make their content visible to appease social media algorithms through engaging in algorithmic gossip (Bishop, 2019), and creates increased precarity for content creators (Glatt, 2021, 2022).

Algorithms are often portrayed by platforms as impartial or objective; however, technologies are rarely apolitical and algorithms are inscribed with assumptions about what is important, valuable or unacceptable (Gillespie, 2014). Academic debate about the impact of algorithmic sorting and computational logics has highlighted concerns about inequality, discrimination and knowledge dissemination as we are divided into calculated publics (Graham, 2004; Gillespie, 2014; Noble, 2018). As our information gathering becomes increasingly internet-driven, we are often exposed only to what algorithms deem appropriate to the silent groupings they sort us into. With different groups exposed to different information, this can affect not only what we find but who we become, in an algorithmic self-fulfilling prophecy (Gillespie, 2014; Mittelstadt et al., 2016; Willson, 2017). Thus, algorithms are not only computational processes but have

social impacts driven by human and institutional choices entwined with censorship and governance.

YouTube has already come under fire due to algorithmic mismanagement of content. In 2017, the platform received a significant backlash when content tagged with LGBTQ+ terms was auto-flagged by algorithms and age restricted into 'restricted mode' (Abidin, 2019; Caplan & Gillespie, 2020; Rodriguez, 2023). For YouTube's LGBTQ+ content creators, this became an unacceptable form of censorship and #YouTubelsOverParty began to trend on twitter. In response YouTube blamed the algorithm: *'Our system sometimes makes mistakes in understanding context and nuances when it assesses which videos to make available in Restricted Mode'* and promised to *'better train our systems'* (YouTube, 2017). Despite this, further controversy later arose around algorithmic demonetization of LGBTQ+ YouTube content as unsuitable for most advertisers (Rodriguez, 2023). This example highlights a lack of human discretion in YouTube's algorithmically enacted governance and raise concerns for how these processes may impact sex edutainment content on the platform (Rodriguez, 2023). It has been noted that social media platforms often conflate sexual content with risk and enact governance and flagging without nuanced consideration to context (Paasonen et al., 2019) and sexual health content may be mistakenly flagged as pornography (Perez, 2021). This raises concerns about what the possibilities and problems might be for sex edutainment in the platformed environment of YouTube. Therefore, our study presents a case study which asks: does YouTube amplify the voices of those sharing educational content about sex and relationships, or do the underlying algorithmic processes shape the journey of sex edutainment to its intended audience in alternative ways?

Methods

Theoretical lens

Burgess and Green (2009) have highlighted that YouTube is a co-creative culture whereby the platform of YouTube (which includes the framework, infrastructure and architecture of the service), content creators and users all interact. Given this, Actor-Network Theory (Latour, 2005) was selected as the underpinning theoretical lens for this study, as it encourages the interrogation of ways that humans and technologies (or non-humans) are mutually shaping, and provided the opportunity to consider how each of the actors in the assemblage of YouTube sex edutainment (YouTube, sex edutainment influencers, and young people) interact to generate a sociology of associations.

Therefore, each phase of the three-phase study focused on one of these key actors.

Within ANT, Callon's (1986) sociology of translation, and the concept of delegation, also provides us with a useful concept to consider how mechanisms by which seemingly protective elements of YouTube's service (such as policies around sexual content and algorithmic moderation) are combined in ways that create and reinforce imbalances of power. Callon proposes four moments of translation, which can assist in understanding how control can be enacted on some actors within an assemblage by other actors: problematization, interessement, enrolment and mobilization. We will expand upon these moments in our discussion, using examples from the data to explore how, through this process of translation, delegation to algorithms is achieved on YouTube in relation to sex education content. Delegation refers to the 'Strategy, process or act of allocating a social control function to a material artefact' (Sørensen, 2002, p. 122). Before exploring the ways, this delegation leads to the deplatforming of sex education.

Study design

This paper draws on findings from across all three phases of a larger study into the possibilities and problems of YouTube Sex Edutainment. Below we provide an overview of the methods involved:

Phase one: YouTube

The walkthrough method (Light et al., 2018), a method rooted in ANT, was used to explore YouTube in app and website form in April 2020 and repeated in July 2021. The walkthrough was used to understand how YouTube's platform architecture, identity, policy and governance might impact sex edutainment, and included comparisons of features and content available between adult users and users aged under 18. Data were analysed using reflexive thematic analysis (Braun & Clarke, 2019).

Phase two: sex edutainment influencers

In total, 60,070 comments from 22 YouTube sex edutainment videos, created by eight influencers, were scraped, to understand responses to their sex edutainment content, using a comment scraper on the 26th April 2020. Content analysis was performed on this comment data using a stepped approach for analysing large social media datasets (Vasilica et al., 2021) that combines elements of framework and content analysis and the *Big Content Machine*, a lightweight open-source software tool for analysis of large-scale conversational data. Sex edutainment influencer accounts were identified for inclusion by scoping various sources including YouTube searches of terms related to sex education, searches of influencer agency search-engines, articles compiling lists of the best channels that teach sex education, and enquiries within the researchers' professional networks. For inclusion influencers needed 40,000+ subscribers, to create content in English, and have a body of content related to sex and relationships. Organization accounts were excluded. The eight influencers selected reflect that YouTube sex edutainment is a small niche. However, the influencers represented a range of intersectional sex education topics including general sex education alongside disabled, LGBTQ+, and religious perspectives, and the 22 videos selected represented a range of intersectional topics. Videos included had been posted onto YouTube between October 2012 and March 2020, providing both recent and older content.

All eight influencers above were also approached for a 10-question email interview to provide their perspectives. Email interviews were selected for flexibility to increase influencer engagement. Of the eight influencers, only one responded, therefore five additional influencers who provided frequent sex edutainment content with fewer than 40,000 followers or in non-English language were also approached. In total, three influencers participated. All interview data were analysed using reflexive thematic analysis.

Phase three: young people

An online mixed-methods survey was conducted with 85 British young people, aged 13–18 years old ($n = 50$), and 19–24 years old ($n = 35$). This survey asked their opinions on influencers in general, their experiences of sex education and how they seek and share information about sex and relationships. Analysis was conducted using reflexive thematic analysis and descriptive statistics.

Findings

While the wider study found evidence of multiple opportunities presented by YouTube sex edutainment (including self-directed learning, peer-support amongst commenters, content creators acting as role-models and health influencers, and accessible and easily consumable content), these possibilities were impeded by platform governance practices enacted by YouTube through algorithmic delegation.

Policy vs practice: algorithms, demonetization and age restriction

The walkthrough noted YouTube construct their identity around the platform as a democratized space, stating: '*Our mission is to give everyone a voice*' (YouTube, 2020). Formulating their identity

around the democratic language of four ‘freedoms’ (Freedom of Expression, Freedom of Information, Freedom of Opportunity, and Freedom to Belong). The platform suggest that in order for this freedom to be safe from harmful content, community guidelines need to be in place and upheld through moderation. YouTube emphasize they ‘*apply [the guidelines] to everyone equally – regardless of the subject or the creator’s background, political viewpoint, position or affiliation*’ (YouTube, 2021a). These guidelines are enforced using a combination of user flagging practices, human reviewers and machine learning (YouTube, 2021b). Although YouTube emphasize the human element, this approach leans heavily on algorithmic machine learning as of the 6,229,882 videos removed from YouTube between July 2021 and September 2021 only 328,641 were not removed by automated algorithmic flagging (Google, 2021). Our data noted YouTube highlighting the importance of machine learning in this process by emphasizing that over 500 hours of video are uploaded to the platform every minute, requiring ‘*the power of advanced machine learning systems*’ as a solution (YouTube, 2021a). However, they remain vague in explaining the mechanisms behind their recommendation and moderation algorithms. Although, in the creator academy guide, they provide content creators with details of what influences their algorithms, emphasizing that the algorithm does not penalize creators, instead focusing on known data about the individual user based on watching habits and engagement.

The walkthrough identified that YouTube has no specific policy for sex education/edutainment content although sex education was eligible for advertising according to YouTube’s advertiser-friendly guidelines. Sex edutainment content does not appear to contravene the ‘nudity and sexual content’ or ‘child safety’ policies, and the sexual content policy draws a distinct line between content designed to arouse and content designed to educate. For example, nudity is permitted on YouTube if used for educational purposes, however content intended for sexual gratification is not. Despite this, one of our survey respondents expressed concerns about YouTube censoring educational content, stating ‘*YouTube blocks out a lot of things even if it’s educational*’. This was corroborated by the observation that all 22 videos used in this study had been moved into restricted mode, where they were age-restricted from viewers under the age of 18, despite the majority of these videos having been age-appropriate for under 18s.

From the email interviews with influencers, it became clear that these were not isolated incidents:

YouTube, in particular, heavily restrict sex-education based content, both making it impossible to monetise and make a living from, and also restricting the potential reach of the content. Sex ed based content specifically targeted at teenagers for instance often won’t reach them as it will be marked as 18+ – Influencer A

Since I talk about sexuality in general, almost all videos are demonetized, so I make very little money. – Influencer B

A lot of my videos have been age restricted by YouTube and there’s a constant battle for me to keep trying to prevent this from happening. Also I can’t monetise any of the videos (ie have ads) because of the content. [. . .] It’s happened so many times that I’ve lost count. Also content has just been silently deleted without telling me.

Therefore, despite sex edutainment content not contravening policy, it appears YouTube’s algorithms lack the nuance to distinguish between content to arouse and content to educate, thus limiting valuable educational content, creating barriers to the creation of the content and economic instability for the influencers involved.

The walkthrough highlighted that in some cases, sex edutainment content was restricted while other content related to the same sexual topics was not. When using a profile imitating a 15-year-old to search the question ‘is masturbation wrong?’ the first recommended video titled ‘is masturbation a sin?’ was by a YouTube sex edutainment influencer and was designed for young people. However, clicking on the video triggered an age restriction rendering the content unviewable. Meanwhile, other videos on the topic aimed at religious adults had not been age-restricted. Therefore, the only content designed for young people had been restricted from them. Whether this restriction was initiated by automatic moderation algorithms or a user flagging the video as inappropriate is unknown, as this information is not provided. It was noted that the

most common form of resistance to sex edutainment content observed in this study was from commenters who felt the topics were inappropriate or rude, examples of this included stating that an influencer was obsessed with sex, suggesting influencers were wrong to use the words 'penis' and 'vagina' in a video, or that the content was an example of how degenerative society had become. Therefore, YouTube's enrolment of users to report content they felt broke community standards alongside algorithms may lead to personal politics conflating content with rule-violations when the content does conform to policy. Videos must then remain on restricted mode until they can be reviewed by a human. Regardless, the restriction of this content may mean young people under 18 are unaware of content designed to educate them. As can be seen in the answer from one survey participant;

The vids I saw were very poor and lacked any depth, they were sterile and bland (I suppose because all ages can see them). I don't have an account to see more "detailed" age-appropriate advice

This raises concerns that when young people seek information, they may not be reaching content designed to engage them, although some YouTube comments discussed using disconnection practices to circumvent this such as using a fake date of birth.

Demonetization and the erosion of audience trust

Age restriction on sex edutainment content impacts not only who can view videos but how the influencers creating them are remunerated for their labour. Influencers cannot monetize via YouTube advertising on age-restricted videos, even when being shown to users over 18. The walkthrough identified that YouTube has previously emphasized a 'family-friendly' agenda for advertising, suggesting content creators avoid topics that might put off advertisers (Creator Blog, 2013), but YouTube now state sex education is suitable for advertising under their advertiser-friendly content guidelines. Despite this, all the videos in this study had been demonetized due to age restriction, a regular occurrence according to the influencers quoted earlier in this paper.

Whilst YouTube regularly state the role of human moderators in their system, algorithms appear to conflate sex edutainment with sexual content and apply age restrictions, leaving influencers unable to monetize most of their content. This creates a dilemma for YouTube sex edutainment, as high-quality content creation is labour-intensive, and influencers are entitled to seek remuneration for their work. Without monetization through YouTube, influencers are left to find sponsorship or income to fund content creation. This study noted funding comes primarily through advertising partnerships with businesses (e.g. condom or sex toy brands), crowdfunding using external sites such as patreon where followers pay a small monthly payment to access additional content, or the selling of merchandise; practices previously discussed by Johnston (2017) and Cunningham and Craig (2017). However, young people in this study identified that advertising affected their trust of influencers. For young people who did not follow any influencers at all, the reasons given often related to distrust of their motives:

Don't give two hoots about influencers as everything they promote is to increase their image or promote products they have been given for free or a fee in which they would probably not have looked twice at before

– Survey respondent

For young people who did follow influencers in general, when asked what made an influencer untrustworthy, the most common answers related to advertising contexts:

'Just selling/advertising stuff in every post' - Survey respondent

'They're being paid to flog something' - Survey respondent

'Being a sell out. Doing paid ads all the time without engaging with their following normally'. - Survey respondent

They are more worried about their image and success and making the bucks than being in the real world' -
Survey respondent

Too much advertising, or promoting products unaligned with their values, was seen as untrustworthy, with young people questioning if advertisements created bias in the validity of information shared. Resistance or distrust against influencers often linked with assumptions influencers were greedy or unethical in seeking money. These findings were corroborated by the comment analysis. While most comments were positive, some questioned influencer authenticity, accusing influencers of being *'money grabbing'* or *'only interested in making money'*. One comment asked how the influencer could be a reliable, objective source of information while using their platform to sell merchandise and products.

These findings align with existing literature on the delicate balance influencers have between promotional discourse and retaining authenticity in the eyes of their audience (Arriagada & Bishop, 2021; Cunningham & Craig, 2017). In the case of this study, however, sex edutainment influencers are often forced into external forms of monetizing their content as algorithms conflate their videos with sexual content. Given that young people aged 13–24 who participated in this study indicated that excessive monetization of content reduced their trust in social media influencers, the demonetization of sex edutainment may lead to the erosion of audience trust in the influencers who create this content. This is a concern as one of the opportunities we identified around YouTube sex edutainment content was that these influencers can act as role models and sexual health influencers; however, this requires a parasocial trust relationship.

Protectionist algorithms?

Content moderation is not inherently negative and is essential to the management of safe virtual communities (de Gregorio, 2020). The walkthrough identified that rhetoric of safety and protection was regularly used in relation to moderation practices, such as using automated flagging systems to prevent harmful content reaching users. However, analysis of YouTube comments in this study called into question how successful algorithms are at protecting, particularly for sex edutainment influencers themselves.

Although most responses to the influencers and their content were positive, sex edutainment influencers were regularly harassed or trolled on their videos. Female influencers received more of these comments than their male counterparts, but harassing/trolling content varied dependent on influencer characteristics. For example, sexually harassing comments were noted towards every female sex edutainment influencer in the dataset. These often centred on sexualizing the women, making objectifying comments about their bodies, and asking if they liked specific sexual practices. Meanwhile, LGBTQ+ male, transgender and non-binary influencers had homophobic and transphobic comments directed towards them. None of the influencers were heterosexual cis-gender males. One female influencer who publicly identified as feminist received many trolling and aggressive anti-feminist comments that included those telling her to *'go kill yourself'* or suggesting she should be burned. This raises concerns that YouTube's moderation practices not only algorithmically restrict sex edutainment content from reaching the audiences it is designed for but do so while not adequately protecting content creators themselves.

Discussion & conclusion

The walkthrough confirmed that YouTube position themselves as a platform from which to speak (Caplan & Gillespie, 2020; Gillespie, 2010) with the intention to amplify all voices, utilizing the rhetoric of democracy and equality. However, although YouTube centre their mission around the rhetoric of empowerment and democratization with the language of *'freedoms'*, those freedoms are not equally

extended to sex edutainment influencers, or audiences seeking ‘freedom of information’ related to sex.

Instead, Returning to Callon’s four moments of translation, we can see how YouTube establish delegation work to algorithms, and encourage users to participate in the process:

- *Problematization* is where a situation is portrayed as a problem requiring an intermediary actor within an assemblage to become an indispensable and obligatory passage point within the narrative of providing a solution, e.g. users are told that maintaining YouTube community standards at scale is a problem, as 500 hours of video are uploaded per minute, and that this necessitates machine learning algorithms to manage.
- During *Interessement* actions are made to encourage other actors to conform to the problematization so they become invested in the solution, e.g. YouTube creates community guidelines that employ rhetoric of risk and the need for protection to invest users in the need for algorithmic governance.
- *Enrolment* is the process of negotiation to persuade actors to perform the role that is being set for them in relation to the problem, e.g. YouTube set out the penalties for violation of community standards and how their algorithms will enact them through content removal and bans.
- Finally, mobilization encourages actors to speak on behalf of the problem and champion the solution, e.g. YouTube emphasizes that human reviewers are used to confirm algorithmic content restriction, which further trains their systems. Users are encouraged to join them in supporting the deployment of algorithms to solve problems of risk by engaging in user-flagging of content.

This process pays lip service to participatory narratives while delegating power to the algorithm through a process of translation. However, in the context of sex education content, this delegation work to algorithms is problematic.

Firstly, sex education frequently requires the use of language that is closely linked, or the same, as that used in sexual content (e.g. sex, masturbation, virginity) and algorithms lack the discretion to distinguish between the content that is designed to arouse and to educate. This parallels similar challenges prevention campaigns for terrorist extremism have experienced from YouTube’s algorithmic misunderstanding of the nuance between promotion and prevention content (Schmitt et al., 2018).

In addition, the process of mobilization, which enrolls users to flag content they believe to be inappropriate, can be problematic. Given that sex education for young people remains a contested topic, personal politics may impact user’s decision to flag a video. As YouTube then requires humans to review flagged content, this can leave wrongly flagged content which does not contravene YouTube’s policies waiting in restricted mode until they can be reviewed.

However, by utilizing this process of translation to enable the delegation of algorithms, when it comes to sex education content, YouTube continue to politically position themselves betwixt their progressive platform imaginary and the quietly conservative ‘family friendly’ agenda that drives their advertising revenue. Further demonstrating the dissonance between the public and private faces of YouTube, which profess partnership, diversity, and social progress whilst quietly enacting discrimination and harm on marginalized content creators (Rodriguez, 2023).

The individuals creating sex edutainment content on YouTube are largely women and LGBTQ+ individuals seeking to meet the intersectional needs of young people who may have been marginalized by traditional sex education. Their content gives voice to topics that rarely grace classrooms and has valuable potential due to the trust relationships sex edutainment influencers create with their audiences. However, as the influencers interviewed in this study highlighted, they consistently experience the deplatforming of their content, which in turn

leads some to be disheartened with creating content at all. This further destabilizes the already precarious labour and working conditions of content creators (Caplan & Gillespie, 2020; Glatt, 2021) who we know are often at the mercy of algorithms (Glatt, 2022). The practices of algorithmic demonetization and age restriction force sex edutainment influencers further into the complex negotiation between serving audiences and advertisers (Abidin & Ots, 2015; Arriagada & Bishop, 2021; Cunningham & Craig, 2017), and risk the potential erosion of trust between the influencers and their audiences. Thus, in the case of sex education on YouTube, algorithms may be silent non-human actors but they can exert power with human impacts.

However, we must be critical of attempts to simply blame algorithms without accountability for the human decisions that underlie them (Gillespie, 2014). Algorithms do not operate in a social vacuum, they learn from the human patterns they observe and the training they receive from platform governance (Noble, 2018) and human flagging practices, reflecting wider perceptions in our society. In the case of YouTube sex edutainment, the same social discourses and protectionist narratives that have caused destabilization of sex education in schools permeate our digital environments and can be seen in the deplatforming of sex edutainment. As such, while YouTube states they are committed to amplifying voices and providing freedom of information and the freedom of opportunity for content creators, tracing the connections between the actors in the assemblage of YouTube sex edutainment demonstrates that the platform continue to reinforce social norms about sex and sexuality.

Furthermore, it has long been argued that protectionist narratives around sex education do not necessarily safeguard children and young people and instead limit their agency (Levine, 2002). Likewise, YouTube's algorithmic protectionism does young people a disservice by limiting their freedom to locate age-appropriate information about sex and relationships and other benefits. Likewise, while YouTube may intend to shield users from harmful content through policy and algorithmic structures, they fail to protect influencers from harassment, furthering Tarvin and Stanfill (2022) concerns about 'governance-washing' on the platform.

Whether YouTube's quiet deplatforming of sex education is intentional or is not unclear, it remains to be seen if YouTube will address the tensions within its service to reduce these issues. Although in 2021 YouTube announced the formation of their health partnerships team and intention to occupy space as a legitimate broker of health knowledge, perhaps this may act as a catalyst for the platform to review their algorithmic moderation to utilize the strength of their existing sex edutainment content, as this content is already being used successfully and is largely valued by the audiences who can engage with it. Given this, we echo Perez (2021) in calling for YouTube to review appropriate and nuanced content moderation policies for educational sexual health content.

Finally, what can be said for Johnston's (2017) suggestion that YouTube sex edutainment provides resources for young people beyond school sex education? Whilst we identified benefits of YouTube sex edutainment, these advantages are destabilized by the platform itself. Whilst YouTube may give sex edutainment influencers a voice, those voices will not reach all ears and those who choose to speak about sex education risk restriction, demonetization, economic precarity, and harassment on the platform. YouTube delegate the enforcement of their policies to algorithms, subtly contradicting the public social imaginary they curate around their platform and reinforcing the deplatforming of sex education content on social media.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Lisa Garwood-Cross is University Fellow in Digital Health and Society at the University of Salford. Her research interests include social media, influencers and their intersections with health and sexuality, alongside the use of digital methods.

Ben Light is Professor of Digital Society at the University of Salford. His research interests include how digital technologies impact society, particularly in relation to dating apps, social networking sites, gender and sexuality. Ben works frequently in digital methods research.

Anna Mary Cooper-Ryan is Head of Public Health at the University of Salford. Her work focuses on behaviour change interventions, using digital data collection tools within interventions and during their evaluation.

Cristina Vasilica is Reader in Digital Health & Head of Digital Engagement at The University of Salford. Cristina's work focuses on digital health, digital methods, engagement and digitization of learning.

ORCID

Lisa Garwood-Cross  <http://orcid.org/0000-0002-4354-3350>

Ben Light  <http://orcid.org/0000-0002-0038-8979>

Anna Mary Cooper-Ryan  <http://orcid.org/0000-0002-8305-8587>

Cristina Vasilica  <http://orcid.org/0000-0002-3012-1087>

References

- Abidin, C. (2017). Influencer extravaganza: Commercial "lifestyle" microcelebrities in singapore. In L. Hjorth, H. Horst, A. Galloway, & G. Bell (Eds.), *The Routledge companion to digital ethnography* (pp. 184–194). Routledge.
- Abidin, C. (2019). Yes homo: Gay influencers, homonormativity, and queerbaiting on YouTube. *Continuum: Lifelong Learning in Neurology*, 33(5), 614–629. <https://doi.org/10.1080/10304312.2019.1644806>
- Abidin, C., & Ots, M. (2015, August 6–9). The influencer's dilemma: The shaping of new brand professions between credibility and commerce. *Paper presented at the AEJMC 2015, annual conference, San Francisco, CA*.
- Are, C. (2021). The shadowban cycle: An autoethnography of pole dancing, nudity and censorship on instagram. *Feminist Media Studies*, 22(8), 2002–2019. <https://doi.org/10.1080/14680777.2021.1928259>
- Arriagada, A., & Bishop, S. (2021). Between commerciality and authenticity: The imaginary of social media influencers in the platform economy. *Communication, Culture and Critique*, 14(4), 568–586. <https://doi.org/10.1093/ccc/tcab050>
- Bauer, M., Hammerli, S., & Leeners, B. (2020). Unmet needs in sex education—what adolescents aim to understand about sexuality of the other sex. *Journal of Adolescent Health*, 67(2), 245–252. <https://doi.org/10.1016/j.jadohealth.2020.02.015>
- Beer, D. (2009). Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society*, 11(6), 985–1002. <https://doi.org/10.1177/1461444809336551>
- Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, 20(1), 1–13. <https://doi.org/10.1080/1369118X.2016.1216147>
- Bialystok, L., & Andersen, L. M. (2022). *Touchy subject: The history and philosophy of sex education*. University of Chicago Press.
- Bishop, S. (2019). Managing visibility on YouTube through algorithmic gossip. *New Media & Society*, 21(11–12), 2589–2606. <https://doi.org/10.1177/1461444819854731>
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
- Burgess, J., & Green, J. (2009). *YouTube: Online video and participatory culture*. Polity Press.
- Burgess, J., Green, J., & Rebane, G. (2020). Agency and controversy in the YouTube community. In H. Friese, M. Nolden, G. Rebane & M. Schreier (Eds.), *Handbuch Soziale Praktiken und Digitale Alltagswelten* (pp. 105–116). Springer VS.
- Callon, M. (1986). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of st brieuc bay. In J. Law (Ed.), *Power, action and belief: A new sociology of knowledge?* (pp. 196–223). Routledge.
- Caplan, R., & Gillespie, T. (2020). Tiered governance and demonetization: The shifting terms of labor and compensation in the platform economy. *Social Media+ society*, 6(2), 2056305120936636. <https://doi.org/10.1177/2056305120936636>
- Cotter, K. (2019). Playing the visibility game: How digital influencers and algorithms negotiate influence on instagram. *New Media & Society*, 21(4), 895–913. <https://doi.org/10.1177/1461444818815684>
- Creator Blog, Y. (2013). 'A friendly reminder and monetization advice' [accessed 12/12/2020]. <https://youtube-creators.googleblog.com/2013/02/a-friendly-reminder-and-monetization.html>
- Cunningham, S., & Craig, D. (2017). Being 'really real' on YouTube: Authenticity, community and brand culture in social media entertainment. *Media International Australia*, 164(1), 71–81. <https://doi.org/10.1177/1329878X17709098>

- de Gregorio, G. (2020). Democratising online content moderation: A constitutional framework. *Computer Law & Security Review*, 36, 105374. <https://doi.org/10.1016/j.clsr.2019.105374>
- Gillespie, T. (2010). The politics of 'platforms'. *New Media & Society*, 12(3), 347–364. <https://doi.org/10.1177/1461444809342738>
- Gillespie, T. (2014). The relevance of algorithms. *Media Technologies: Essays on Communication, Materiality, and Society*, 167(2014), 167.
- Glatt, Z. (2021). We're all told not to put our eggs in one basket: Uncertainty, precarity and cross-platform labor in the online video influencer industry. *International Journal of Communication*, 16(2022), 1–19. <https://ijoc.org/index.php/ijoc/article/view/15761>
- Glatt, Z. (2022). Precarity, discrimination and (in)visibility: An ethnography of "the algorithm" in the YouTube influencer industry. In E. Costa, P. Lange, N. Haynes, & J. Sinanan (Eds.), *The routledge companion to media anthropology* (pp. 546–559). Routledge.
- Google. (2021). *Transparency report: YouTube community guidelines enforcement*. Retrieved March, 2020, from https://transparencyreport.google.com/youtube-policy/removals?hl=en&total_removed_videos=period:2021Q3;exclude_automated:all&lu=total_removed_videos
- Graham, S. (2004). The Software-sorted City: Rethinking the "Digital Divide". In S. Graham (Ed.), *The Cybercities Reader* (pp. 324–332). Routledge.
- Grose, R. G., Grabe, S., & Kohfeldt, D. (2014). Sexual education, gender ideology, and youth sexual empowerment. *The Journal of Sex Research*, 51(7), 742–753. <https://doi.org/10.1080/00224499.2013.809511>
- Herd, G. (2009). *Moral panics, sex panics*. NYU Press.
- Hou, M. (2019). Social media celebrity and the institutionalization of YouTube. *Convergence*, 25(3), 534–553. <https://doi.org/10.1177/1354856517750368>
- Irvine, J. M. (2009). Shame comes out of the closet. *Sexuality Research & Social Policy*, 6(1), 70. <https://doi.org/10.1525/srsp.2009.6.1.70>
- Johnston, J. (2017). Subscribing to sex edutainment: Sex education, online video, and the YouTube star. *Television & New Media*, 18(1), 76–92. <https://doi.org/10.1177/1527476416644977>
- Kantor, L. M., & Lindberg, L. (2020). Pleasure and sex education: The need for broadening both content and measurement. *American Journal of Public Health*, 110(2), 145–148. <https://doi.org/10.2105/AJPH.2019.305320>
- Latour, B. (2005). *Reassembling the social: An introduction to actor network theory*. Oxford University Press.
- Lenskyj, H. (1990). Beyond plumbing and prevention: Feminist approaches to sex education. *Gender and Education*, 2(2), 2. 217–230. <https://doi.org/10.1080/0954025900020206>
- Levine, J. (2002). *Harmful to minors*. Minneapolis; University of Minnesota Press.
- Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, 20(3), 881–900.
- Lim, R. B. T., Tham, D. K. T., Cheung, O. N., Adaikan, P. G., & Wong, M. L. (2019). A public health communication intervention using edutainment and communication technology to promote safer sex among heterosexual men patronizing entertainment establishments. *Journal of Health Communication*, 24(1), 47–64. <https://doi.org/10.1080/10810730.2019.1572839>
- McKee, A. (2017). Learning from commercial entertainment producers in order to create entertainment sex education. *Sex Education*, 17(1), 26–40. <https://doi.org/10.1080/14681811.2016.1228528>
- McKee, A., Albury, K., Burgess, J., Light, B., Osman, K., & Walsh, A. (2018). Locked down apps versus the social media ecology: Why do young people and educators disagree on the best delivery platform for digital sexual health entertainment education? *New Media & Society*, 20(12), 4571–4589. <https://doi.org/10.1177/1461444818778255>
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 205395171667967. <https://doi.org/10.1177/2053951716679679>
- Noble, S. U. (2018). *Algorithms of oppression*. New York University Press.
- Oosterhoff, P., Müller, C., & Shephard, K. (2017). Sex education in the digital era. *IDS bulletin*, 48(1), 1. <https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/12818>
- Ormen, J., & Gregersen, A. (2023). Institutional polymorphism: Diversification of content and monetization strategies on YouTube. *Television & New Media*, 24(4), 432–451. <https://doi.org/10.1177/15274764221110198>
- Paasonen, S., Jarrett, K., & Light, B. (2019). *NSFW: Sex, humor, and risk in social media*. MIT Press.
- Paasonen, S., & Saarenmaa, L. E. (2023). Short-lived play: Trans-European travels in print sex edutainment. *Media History*, 29(2), 240–254. <https://doi.org/10.1080/13688804.2022.2054410>
- Perez, P. B. (2021). Facebook doesn't like sexual health or sexual pleasure: Big tech's ambiguous content moderation policies and their impact on the sexual and reproductive health of the youth. *International Journal of Sexual Health*, 33(4), 1–5. <https://doi.org/10.1080/19317611.2021.2005732>
- Petre, C., Duffy, B. E., & Hund, E. (2019). "Gaming the system": Platform paternalism and the politics of algorithmic visibility. *Social Media+society*, 5(4), 205630511987999. <https://doi.org/10.1177/2056305119879995>
- Pilipets, E., & Paasonen, S. (2022). Nipples, memes, and algorithmic failure: NSFW critique of Tumblr censorship. *New Media & Society*, 24(6), 1459–1480.

- Pound, P., Langford, R., & Campbell, R. (2016). What do young people think about their school-based sex and relationship education? A qualitative synthesis of young people's views and experiences. *British Medical Journal Open*, 6(9), e011329. <https://doi.org/10.1136/bmjopen-2016-011329>.
- Ringrose, J. (2013). *Postfeminist education?*. Routledge.
- Rodríguez, J. A. (2023). LGBTQ incorporated: YouTube and the management of diversity. *Journal of Homosexuality*, 70(9), 1807–1828. <https://doi.org/10.1080/00918369.2022.2042664>
- Schmitt, J. B., Rieger, D., Rutkowski, O., & Ernst, J. (2018). Counter-messages as prevention or promotion of extremism?! The potential role of YouTube: Recommendation algorithms. *Journal of Communication*, 68(4), 780–808. <https://doi.org/10.1093/joc/jqy029>
- Shannon, B. (2016). Comprehensive for who? Neoliberal directives in Australian 'comprehensive' sexuality education and the erasure of GLBTIQ identity. *Sex Education*, 16(6), 573–585. <https://doi.org/10.1080/14681811.2016.1141090>
- Sørensen, K. H. (2002). Social shaping on the move? on the policy relevance of the social shaping of technology perspective. In K. H. Sørensen & R. Williams. (Eds.), *Shaping technology, guiding policy: concepts, spaces and tools* (pp. 19–35). Edward Elgar.
- Tarvin, E., & Stanfill, M. (2022). "YouTube's predator problem": Platform moderation as governance-washing, and user resistance. *Convergence*, 28(3), 822–837. <https://doi.org/10.1177/13548565211066490>
- van Es, K., & Poell, T. (2020). Platform imaginaries and dutch public service media. *Social Media+ Society*, 6(2), 2056305120933289. <https://doi.org/10.1177/2056305120933289>
- Vasilica, C. M., Oates, T., Clausner, C., Ormandy, P., Barratt, J., & Graham-Brown, M. (2021). Identifying information needs of patients with IgA nephropathy, using an innovative social media stepped analytical approach. *Kidney International Reports*, 6(5), 1317–1325. <https://doi.org/10.1016/j.ekir.2021.02.030>
- Willson, M. (2017). Algorithms (and the) everyday. *Information, Communication & Society*, 20(1), 137–150. <https://doi.org/10.1080/1369118X.2016.1200645>
- YouTube. (2017). *Restricted mode*. <https://blog.youtube/news-and-events/restricted-mode-how-it-works-and-what/> [accessed 12/12/2020].
- YouTube. (2020). *About - press*. [accessed 12/12/2020]. <https://www.youtube.com/about/press/>
- YouTube. (2021a). *Community guidelines* [accessed 12/12/2020]. <https://www.youtube.com/howyoutubeworks/policies/community-guidelines/>
- YouTube. (2021b). *Progress on managing harmful content* [accessed 12/12/2020]. https://www.youtube.com/intl/ALL_uk/howyoutubeworks/progress-impact/responsibility/#detection-source