ISD | Institute for Strategic Dialogue

The 'cost of doing politics'?

Gendered abuse and digital platforms' role in undermining democracy

Paula-Charlotte Matlach & Charlotte Drath

About the authors

Paula-Charlotte Matlach
Paula Matlach is an analyst at ISD Germany. She is an expert on online gender-based violence (OGBV) and far-right extremism. Her most recent publications include a series of briefings analysing manifestations of OGBV on TikTok, such as "Off-limits: Sexual Violence on TikTok" and "Recommending Hate: How TikTok's Search Engine Algorithms Reproduce Societal Bias". She further co-authored the ISD report "The German Far Right Online. A Longitudinal Study".

Charlotte Drath
Charlotte Drath is a research consultant at the Institute for Strategic Dialogue. Her current research focuses on online gender-based violence (OGBV) in multiple languages as well as the German-speaking Manosphere. She is the co-author of ISD research reports including "Toxic Tips: Misogynistic Narratives on TikTok in Hungary" and "Recommending Hate: How TikTok's Search Engine Algorithms Reproduce Societal Bias".

Acknowledgements

We would like to express our gratitude to Allison Castillo, Martin Degeling, Eva F Hevesi, Sina Laubenstein, Anna-Katharina Meßmer and Pauline Zaragoza for their contributions to the conceptualisation and analyses.

Contents

Executive Summary

Online Gender-Based Violence (OGBV) has become a pervasive threat in the digital age and space: it undermines democratic processes, silences marginalised voices and perpetuates systemic inequality. Harassment, threats and abuse (both in person and online) are so common, that women and gender minoritised people simply consider them "the cost of doing politics". As a result, 21 percent of women parliamentarians in Europe said that they did not want to pursue another term in office.

This project was designed to examine the different ways in which the harmful stereotypes and beliefs underpinning OGBV manifest on TikTok, an increasingly vital space for political expression especially among younger audiences. ISD's research included an audit of TikTok's search algorithm and the platform's content moderation practices on sexual violence, as well as analyses of the prevalence of misogynistic content in Hungary and the spread of gendered disinformation in Germany. ISD also examined harmful content targeting candidates during the 2024 European Parliamentary elections campaign period on TikTok and the French legislative election campaign.

On TikTok, ISD found:

- Gendered bias and gaps in TikTok's content moderation and search engine,
- Gaps in the platform's response to gendered disinformation,
- Overblocking of non-violent content including reproductive health and sex education,
- Unequal safeguarding of users across locations and languages,
- Normalised misogyny and other wide-spread harmful but legal content.

In relation to governments, regulators, political representatives and other stakeholders, ISD found:

- Gaps in ensuring platform accountability,
- Flawed response mechanisms for gendered disinformation,
- Gaps in holding political representatives accountable,
- Irresponsible reporting by media and other outlets on gendered issues.

ISD's findings demonstrate that addressing OGBV requires a holistic approach – one that not only strengthens content moderation but also tackles the underlying social norms and biases that enable online gender-based violence. They further highlight the urgent need for coordinated action and cross-sector collaboration from policymakers, platforms and civil society organisations to effectively combat OGBV. This includes developing stronger legal protections, ensuring consistent enforcement, and supporting victims through research and advocacy initiatives.

This report summarises ISD's findings in each issue area and provides evidence-based recommendations for creating safer and more inclusive online spaces that uphold these principles. It is part of a series examining online gender-based violence (OGBV) on TikTok in English, German, French and Hungarian. It is part of the project Monitoring Online Gender Based Violence Around the European Parliament Election 2024, funded by the German Federal Foreign Office.

Glossary

Algorithmic bias
Algorithmic bias refers to "instances where the outputs of an algorithm benefit or disadvantage certain individuals or groups more than others without a justified reason for such unequal impacts."

Anti-feminism
Anti-feminism describes the countermovement opposing emancipatory or feminist ideals, that uses misogynist strategies and tactics. Anti-feminists oppose gender equality efforts and the democratic negotiation of gender relations.

Gender
Gender refers to a "system of symbolic meaning that creates social hierarchies based on perceived associations with masculine and feminine characteristics." A person's gender identity refers to "an individual's internal, innate sense of their own gender." Their gender expression refers to how individuals present their gender through appearance and behaviour, incorporating elements of femininity, masculinity, and androgyny that influence others perceptions of their gender. Candidates and politicians who did not explicitly express their gender identity are further referred to as women candidates/politicians or male candidates/politicians depending on whether they presented more feminine or masculine in their online content.

Gender-based violence (GBV)
This term refers to "violence directed against a person because of that person's gender or violence that affects persons of a particular gender disproportionately." Women and the LGBTQ+ community, including transgender and gender-diverse persons, experience disproportionate rates of GBV.

Misogyny
Misogyny is considered as a "system that operates within a patriarchal social order to police and enforce women's subordination and to uphold male dominance." It affects not just cisgender heterosexual women but also transgender, intersex, non-binary, genderqueer people, and men. Misogynistic acts are often motivated by underlying sexist ideologies and can neither be defined as purely structural nor as purely individual actions.

Online gender-based violence (OGBV)
OGBV is defined here as a subset of technology-facilitated gender-based violence (TFGBV): this refers to any "act that is committed, assisted, aggravated, or amplified by the use of information communication technologies or other digital tools, that results in or is likely to result in physical, sexual, psychological, social, political, or economic harm, or other infringements of rights and freedoms." For a more detailed review and discussions of terms and definitions please refer to ISD's report "Misogynistic Pathways to Radicalisation."

Very large online platforms (VLOPs) and very large online search engines (VLOSEs)
Online platforms and search engines used by an average of 45 million monthly users or higher (equal to 10 percent of the population in the EU), as defined by the European Union's Digital Services Act (DSA). As VLOPs and VLOSEs pose particular risks regarding the dissemination of illegal content and societal harms, they are subject to the most stringent requirements under the DSA. Examples of VLOPs and VLOSEs include Meta (Facebook and Instagram) and Google (Google Search and YouTube), respectively. A regularly updated list of designated VLOPs and VLOSEs can be found on the website of the European Commission.

Introduction

In the 2024 European Parliamentary election, the number of women Members of the European Parliament (MEPs) decreased for the first time in 45 years. The influence of women MEPs in parliamentary committees also fell, even where representation within committees was improved. This coincides with the rise of broader right-wing populist and anti-feminist movements who have mobilised against gender equality and gained traction across the EU.

Online Gender-Based Violence (OGBV) has become a pervasive threat in the digital age and space, undermining democratic processes, silencing marginalised voices and perpetuating systemic inequality. This phenomenon is particularly acute during election cycles, where female candidates and gender-diverse individuals face disproportionate levels of harassment and disinformation. This was evident in cases including the 2019 European Parliamentary elections, the 2022 US midterms and the 2024 South Africa elections.

Harassment, threats and abuse (both in person and online) are often simply considered "the cost of doing politics" for women and gender minoritised people. As a result, a third of women parliamentarians in Europe felt that the violence they had been subjected to during their term had affected their freedom of expression negatively; 21 percent did not want to pursue another term in office. Public attacks against women also drastically inhibit other women's political ambitions and civic engagement, decreasing their political participation and threatening representative democracy. Misogynistic online discourse has been found to play a significant role in this 'chilling effect.' Although it is considered an issue on all major social media platforms, little research has examined hate directed toward political candidates specifically on TikTok.

This project was designed to examine the different ways in which the harmful stereotypes and beliefs underpinning OGBV manifest on TikTok, an increasingly vital space for political expression especially among younger audiences. This project explored multiple dimensions of OGBV, building on ISD's extensive research into online disinformation, hate speech and gender-based violence (GBV).

The research included an audit of TikTok's search algorithm and the platform's content moderation practices on sexual violence, as well as analyses of the prevalence of misogynistic content in Hungary and the spread of gendered disinformation in Germany. ISD also examined harmful content targeting candidates during the 2024 EP campaign period on TikTok and the French legislative election campaign.

Every piece of research conducted within this project highlighted the same broader systemic challenge in combating OGBV: normalised misogynistic and discriminatory language. ISD found that such language, including remarks often dismissed as minor or in "bad taste", remains a significant systemic barrier in political spaces, including on social media platforms like TikTok.

This report offers a comprehensive understanding of how digital platforms like TikTok shape the landscape of gendered violence and provides evidence-based recommendations for creating safer and more inclusive online spaces that uphold these principles. It further highlights the urgent need for coordinated action from policymakers, platforms and civil society organisations to effectively combat OGBV.

Online gender-based violence (OGBV)

OGBV can be committed, assisted, aggravated or amplified using information and communication technologies or other digital tools. It can result in physical, sexual, psychological, social, political or economic harm, or "other infringements of rights and freedoms." Like any form of GBV, women and the LGBTQ+ community, especially queer and transgender individuals, experience disproportionate rates of OGBV.

Their experiences are further shaped by the interrelationships of social categories such as gender and nationality, as well as gender and religion, which can result in specific forms of discrimination. Ahead of the US 2022 midterm elections, US representatives Ilhan Omar and Rashida Tlaib (who are both Muslim) were subject to specific forms of abuse including accusations that they support terrorist organisations and are foreign agents.

A range of multilateral fora and multistakeholder initiatives including UN Women, the World Health Organisation (WHO), the UN Population Fund (UNFPA) and Global Partnership recognise the need to address the role of platforms in enabling and exacerbating OGBV. UN Women, for instance, has underscored the importance of tackling technology-facilitated violence against women and girls through comprehensive strategies for prevention and response. Similarly, the WHO has recognised online violence as a critical public health and human rights issue, calling for integrated approaches to prevent and address it. Furthermore, the Global Partnership for Action on Gender-Based Online Harassment and Abuse highlights the necessity for international cooperation to tackle this issue across digital platforms. The Christchurch Call has recognised the need to deepen and explain the evidence base on the links between terrorist and violent extremist content (TVEC) and misogyny, advocating for a more robust, evidence-based analysis to strengthen efforts to combat both violent extremism and gendered disinformation.

OGBV has tangible offline impacts. Offline harms including mass violence, intimate partner violence and stalking, can also be extended and amplified online. This can take the form of livestreamed terrorist attacks, deepfake pornography and doxxing. It can also include verbal attacks and stereotypes against women and gender minoritised individuals, further fuelling radicalisation and offline violence. For example, British police have reported that misogynist influencers such as Andrew Tate have contributed to the radicalisation of men and boys, leading to harmful actions both online and offline.

These manifestations of GBV – from "common-sense" online misogyny to violent attacks – form a 'continuum of violence'. Rather than treating acts of gender-based violence as isolated incidents, this concept highlights its interconnected and systemic nature, rooted in deeply ingrained misogynistic beliefs and biases. Instead of creating a "hierarchy of abuse", it recognises the cumulative impact and serious consequences of all instances of physical and non-physical gender-based violence. These include actions often perceived as 'minor' such as sexist jokes.

While OGBV affects women and gender minoritised individuals disproportionately, cisgender men and boys also experience OGBV, especially if they fall outside patriarchal norms of masculinity. At the same time, it is important to recognise that OGBV occurs within an ecosystem characterised by a gender digital divide that is rooted in structural gender inequalities, in which those who design – and, in some countries or regions, access and use – communication technologies are disproportionately male. Consequently, the underlying concepts of patriarchy, male supremacy and the gender binary play a key role in legitimising OGBV.

This is particularly evident in the rise of the global Manosphere – a network of anti-feminist actors who operate both virtually and in real life. These actors promote varying degrees of misogynistic, sexist and male supremacist ideologies. Their persistence and influence are largely sustained by the interconnected nature of the digital landscape and have triggered strong concerns among human rights advocates and users.

As a direct consequence of this online toxicity, women in politics tend to be more guarded than male politicians and expend more resources at a greater risk overall. This can result in self-censorship, withdrawal from public social media channels or even an exit from politics. These dynamics dissuade women and gender minoritised individuals from pursuing public leadership roles or entering politics, demonstrating how OGBV can result in the deterioration of democratic principles. A growing body of research on OGBV, anti-feminism, and online misogyny highlights the increasing prevalence of these issues across all major social media platforms.

OGBV on TikTok

Since its global launch in 2018, TikTok has become an integral part of daily life for many, amassing over 955 million users worldwide. In France (24.7 million users), Germany (22.8 million users) and Hungary (3.1 million users), the platform plays a significant role in shaping online discourse and influencing public opinion. The platform is especially popular among young audiences, and constitutes a crucial space for political expression. However, prior research by ISD has shown downsides to this trend: misogynistic and anti-LGBTQ+ hate amplified by TikTok's algorithm-driven recommendation system and surges in hateful hashtags and gendered abuse and disinformation during elections in the US, South Africa and other countries.

In 2021, TikTok announced a partnership with UN Women and the Web Foundation to promote online safety for women and raise awareness of gender-based violence. The stated goal was to "start the conversation about gender-based violence and educate the TikTok community." This included new functionalities for users to control their comment sections, a prompt reminding its users to be kind, updated community guidelines on deadnaming, and an awareness-raising initiative against GBV. TikTok's Newsroom has not published any further updates on OGBV or GBV more widely.

As a Very Large Online Platform (VLOP) under the EU's Digital Services Act (DSA), it is TikTok's responsibility to create a safe and fair digital space, curb gender-based violence, and mitigate negative impacts on peoples' fundamental rights (as outlined in DSA articles 1, 34 and 35). The DSA came into effect for Very Large Online Platforms (VLOPs) like TikTok in August 2023 and applies to all services from February 2024. It establishes a crucial regulatory framework aimed at creating a "safe, predictable, and trusted online environment that fosters innovation" while protecting fundamental rights enshrined under the Charter of Fundamental Rights of the European Union. Many of its provisions are designed to mitigate online risks related to democratic processes, extremism, disinformation and hate.

TikTok's Community Guidelines (last updated in May 2024) currently include a section on "Safety and Civility" which outlines prohibited behaviours such as "dehumanizing someone on the basis of their protected attributes" (including gender and gender identity, sex, and sexual orientation), degrading individuals based on their "personal appearance," sexual harassment, and the promotion of misogyny and anti-LGBTQ+ ideologies. Under "Sensitive and Mature Themes," the platform states that nudity, sexually suggestive content involving minors, and graphic violence are restricted. However, discussions about sexuality, reproductive health and sex education are permissible.

To enforce its guidelines, TikTok employs a combination of automated technology and language-specific human moderators to detect and manage harmful content. Machine learning (ML) algorithms analyse various signals from the videos including keywords, images, titles, descriptions, audio and metadata to identify potential violations. According to the platform, potential violations are detected by automated moderation technology which either immediately removes content or passes it on to a moderation team that conducts further reviews. TikTok also works with experts, nonprofit organisations, its Youth Council, and Safety and Content Advisory Councils to ensure their "policies and processes are informed by a diversity of perspectives, expertise and lived experiences."

Despite these stated efforts, ISD has found crucial inconsistencies in TikTok's moderation strategy on OGBV-related content, which have been explored in six separate briefings. These include bias in content moderation and search functionality, gaps in the platform's response to gendered disinformation, instances of over-blocking, unequal safeguarding of users across locations and languages, and the open spread of harmful but legal content and normalised misogyny. Each of these areas has been investigated in more detail in our published analyses, which form the basis of the following observations.

Gendered bias and gaps in content moderation

An analysis of 9000 French, German and Hungarian TikTok comments around the 2024 European Parliament election found that women candidates received 80 percent more harmful comments than male candidates. These included personalised attacks, calls for violence, sexual harassment, misogynistic slurs and gendered disinformation. These tactics were also observed in a case study focusing on women candidates for the 2024 French legislative election.

This work corroborates previous findings that women politicians are at higher risk of receiving abusive content online than their male counterparts. They were also more likely to be subject to abuse and harassment based on their professional experience and qualifications, (dis)ability, gender identity, gender expression and/or appearance, in accordance with classic misogynistic tropes. These attacks sometimes stemmed from male candidates themselves. Furthermore, male candidates' channels have served as platforms for hate speech, defamatory speech, and derogatory or discriminatory content. A significant proportion of harmful comments on these channels were directed not only at the channel owners but also at others, frequently targeting women or gender-minoritised groups. This highlights the broader online abuse ecosystem, where political discourse is increasingly dominated by harmful rhetoric that extends beyond the immediate political figures. These findings indicate a significant gendered bias in TikTok's detection capabilities for harmful content.

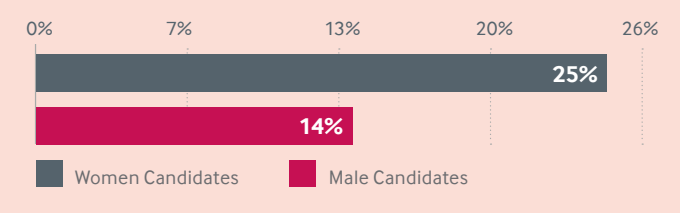

Figure 1: Share of hateful, defamatory, discriminatory and/or derogatory speech comments of all comments received by women candidates and male candidates in France, Germany and Hungary.

ISD also detected gaps in TikTok's approach to misgendering. Commentators degraded women candidates by repeatedly and deliberately misgendering them, especially if they perceived that their gender expression did not conform with concepts of 'traditional femininity.' This behaviour has been linked to physical violence and economic harms targeting transgender, non-binary and genderqueer people, as well as nonconforming heterosexual and cisgender individuals.

The findings of this briefing show how normalised misogynistic language continues to constitute a significant systemic barrier in political spaces including on TikTok.

To mitigate this, TikTok should:

- **Ensure consistent and comprehensive implementation and enforcement of policies related to misogynistic content, gendered abuse and harassment.** This includes the timely removal of illegal hate speech in all languages when the platform has been notified. The code of conduct on countering illegal hate speech online requires signatories, including TikTok, to review valid notices in a timely, diligent, non-arbitrary and objective manner, and to expeditiously remove or to disable access to reported content which violates TikTok's policies or applicable law.
- **TikTok should be consistent in its approach to misgendering.** TikTok prohibits the use of a person's "former name or gender." To be consistent with its own stance, TikTok should recognise that misgendering also applies to gender non-conforming individuals who identify with their gender assigned at birth. As such, a logically consistent approach to their own policy would prohibit misgendering by using a name or gender that does not correspond with a person's gender identity, regardless of whether this person identifies as transgender or not.
- **Ensure a Safety and Privacy by Design informed approach,** taking a victim-survivor-centred perspective. A gender and trauma-informed lens should be applied throughout all stages of development of user interfaces and tools. For example, platforms should proactively provide users with tools that protect their privacy and reduce exposure to hateful attacks.
- **Provide more specific annual transparency reports on content moderation policies and actions.** This will enable external researchers to track and scrutinise the scope and scale of OGBV, as well as assess the enforcement of community guidelines and policies over time. This requires platforms to establish gender-disaggregated, intersectional and standardised transparency reporting.
- **Fully comply with regulatory data access obligations by providing access to public data via researcher APIs.** This is currently limited and would significantly improve research capabilities. While protecting applicable user rights to privacy, TikTok should ensure that the data provided is correct and that access can also be adapted to changes in research projects without unnecessary bureaucracy.

- **Complement the use of AI-based systems to detect and moderate harmful content with greater human oversight and expertise**. This requires teams with specific expertise on (illegal) hate speech against women and gender minoritised people. The nuances of language and the political and social landscape of the relevant region should also be considered. This allows for careful approaches that recognise the role of subtle/veiled misogyny and anti-LGBTQ+ hate. It can also help mitigate algorithmic bias. TikTok should be transparent about the group-specific qualifications of human moderators – for example, expertise or intersectional training in the field of gender-based violence and discrimination.
- **Make risk assessments more transparent and comprehensive.** In TikTok's DSA Risk Assessment Report 2023, gender-based violence content is categorised as a significant risk. However, there is a lack of information and detailed metrics on behavioural and actor-related risks, as well as the effectiveness of mitigation measures.
- **Act in a diligent, objective and proportionate manner in applying and enforcing the restrictions they outline in their terms and conditions in accordance with Article 14 of the DSA.** This also includes the enforcement of their policies for Government, Politician, and Political Party Accounts (GPPPA), which refers to accounts operated by political representatives.

Search functionality bias

TikTok's importance as a search tool is growing worldwide, especially among young users. ISD analysed how TikTok search results are moderated through algorithmic probing, using targeted prompts in English, French, German and Hungarian. In this analysis, all identified slurs were gendered, aimed at women from the respective communities. The search results featured unassuming users who appeared to be targeted by these slurs, even though the prompt's keywords were not present in all the videos displayed.

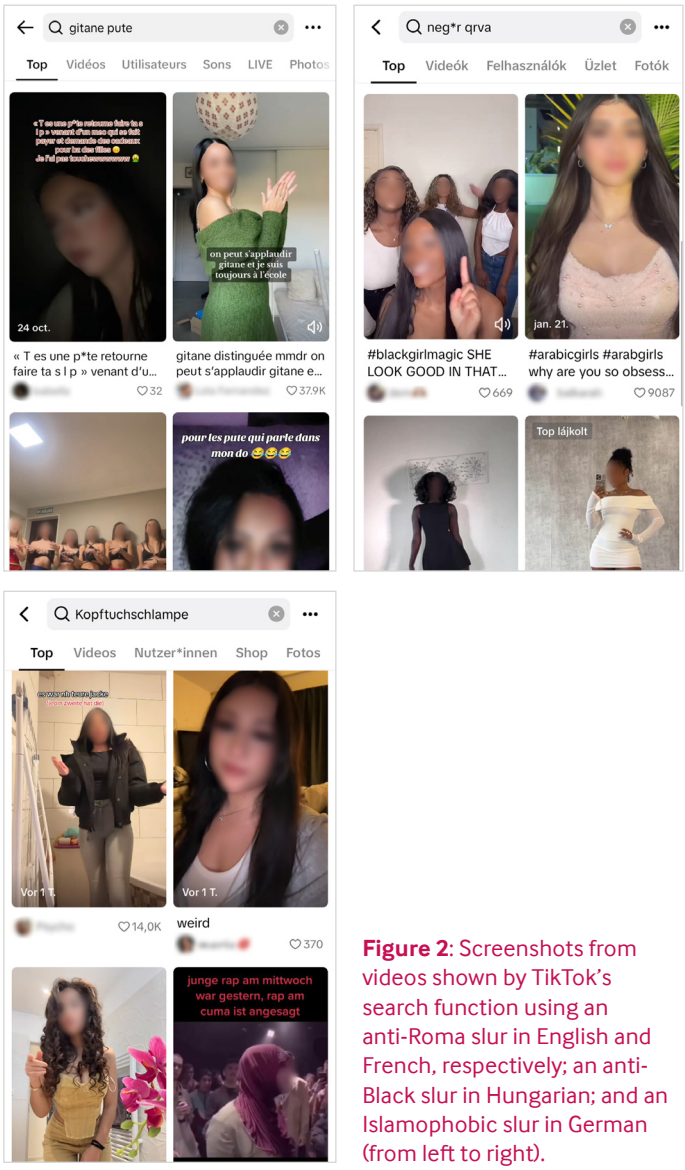

Figure 2: Screenshots from videos shown by TikTok's search function using an anti-Roma slur in English and French, respectively; an anti-Black slur in Hungarian; and an Islamophobic slur in German (from left to right).

This suggests that TikTok's search and recommendation algorithms are not only reflecting but also amplifying harmful societal biases. The fact that slurs were aimed specifically at women indicates a gendered dimension to the content being prioritised. These findings reveal significant evidence of algorithmic bias: across all four languages, search results consistently demonstrated harmful associations that objectify and degrade presumed members of marginalised groups. This pattern suggests that TikTok's search and recommender algorithms reproduce and potentially amplify societal biases to drive user engagement and increase revenue. To prevent this, TikTok should:

- **Incorporate gender analysis and intersectional feminist methodology** when assessing the risks of algorithms and machine-learning (ML) models embedded in their services and ensure that relevant teams (such as those designing, testing, and evaluating algorithms) are diverse and trained on how to conduct gender analysis to detect and mitigate biases and discriminatory patterns.
- **Strengthen its commitment to inclusivity and ethical practices** by identifying, assessing and mitigating the influence of its search and recommender algorithms on systemic risks. This is part of their obligation under articles 1, 34, and 35 of the DSA to create a safe and fair digital space, and to mitigate negative effects on fundamental rights and of gender-based violence.
- **Involve representatives of groups potentially impacted by hateful content in TikTok's DSA risk assessment methodology.** This includes consultations on harms resulting from the design of its search engine and recommender algorithms as well as on the development of related risk mitigation measures. TikTok should publish up-to-date and detailed reports on the results of its risk assessment, including information on how the feedback from representatives of the different groups was considered.
- **Improve transparency to better enable the prevention, detection and ultimately the addressing of discrimination and bias embedded into algorithms.** Although TikTok already provides information regarding the basic parameters used in its search and feed personalisation, it is unclear how exactly content is 'matched' to search queries and user interests. TikTok should provide public information about its search and recommender algorithms' rationale. It should also provide the assumptions regarding potentially affected groups, the main classification choices and what the algorithms are designed to optimise for, the specific relevance of the different parameters, and the decisions about any possible trade-offs.

Gaps in the platform's response to gendered disinformation

The resurgence of the 'National Rape Day' hoax exposed significant gaps in TikTok's content moderation strategy and its ability to stop the spread of gendered mis- and disinformation. The hoax originated from a baseless rumour claiming that sexual violence would be legal on 24 April and that some men were planning to commit a series of sexual assaults. Searches for this trend failed to trigger any labels, warning messages or promotion of content mitigating the spread of this hoax. ISD also found that TikTok's moderation was inconsistent when dealing with different date formats and spelling alternatives to the blocked hashtag "#April24"; this loophole was crucial to the spread of misleading information.

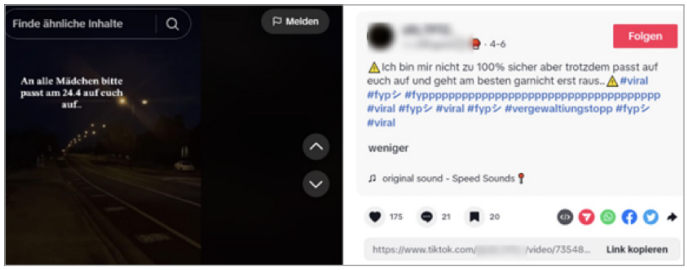

Figure 3. Content in German warning women and girls to not go out on 24 April.

TikTok's inability to tackle this hoax reflects a wider trend: social media platforms consistently overlook and inadequately moderate content that disproportionally cause harm to women and girls as well as transgender, non-binary and genderqueer people. To combat this, TikTok should:

- **Work with external, specialised fact-checkers who can help mitigate the spread of gendered disinformation.** At a minimum, TikTok should strengthen internal content moderation teams to recognise gendered disinformation and mitigate it in a proportionate, transparent, and human-rights compliant manner. Where appropriate, fact-checked content should consistently be labelled and excluded from TikTok's recommender system.
- **Ensure platform design features do not amplify misogynistic content.** This may include algorithms that prioritise sensational, polarising and often harmful content (both legal and illegal). A risk-based or duty of care type approach could help counter the amplification of such harmful but legal content while also preserving rights to speech and expression.

- **Invest in the empowerment of women and gender minoritised people** by offering accessible and bulk reporting mechanisms that enable affected users to quickly flag and report digital attacks against these groups.
- **Develop and operate exchange channels between relevant teams**, including trust and safety as well as content moderation teams, to proactively share information about OGBV incidents such as actors and tactics. This can help map the scale and scope of OGBV, as well as coordinate and inform cross-platform responses.
- **Expand resources to identify influential networks involved in gendered disinformation** campaigns, including means of technical manipulation and other types of OGBV that go against TikTok's Community Guidelines.

Overblocking of non-violent content including reproductive health and sex education

ISD has examined TikTok's content moderation practices on sexual violence in English, French, German and Hungarian, focusing on how these practices align with the platform's policies and user experiences. Although TikTok's Community Guidelines on "Sensitive and Mature Themes" explicitly allow educational content on sexual health, many creators report that their material has been removed or shadowbanned. The lack of transparency provided to users around supposed community guideline violations further hinders their ability to appeal sanctions. Due to TikTok's inconsistent enforcement, some users feel that their topics are "secretly unwanted" although they do not explicitly violate TikTok's Community Guidelines. To mitigate this, they have turned to "netspeak" or "algospeak" to evade algorithmic content moderation. This disconnect between platform policies and user experience underscores broader concerns about content moderation on TikTok.

TikTok's decision to completely block terms associated with sexual violence, such as "rape", leads to overblocking which impacts access to non-violent content including reproductive health and sex education. This content is explicitly allowed under TikTok's own Community Guidelines. By restricting it, TikTok prevents experts, educators, activists and survivors of violence from sharing and receiving support. The banning of such language can further exacerbate and reinforce societal stigma associated with sexual education, gender diversity and surviving sexual violence. This creates barriers to effectively accessing and sharing sexual health content and information on gender diversity, which can inhibit community support and relegate crucial discussions on sexual health, gender and gender-based violence.

Keyword	Language	(VPN)-Location	Search Result
Rape	English	United Kingdom (UK)	Search blocked with support message; link to TikTok's sexual abuse resource support page (language of the prompt appears to be based on device location regardless of VPN); The Survivors Trust phone number displayed
Rape	English	Germany	Search blocked with support message; link to TikTok's German-language sexual abuse resource support page
Rape	English	France, Hungary	No block, produces search results
Vergewaltigung	German	UK, France, Germany, Hungary	Search blocked with German-language community guidelines note
Viol, le viol	French	UK, France, Germany, Hungary	No block, produces search results
Nemi erőszak, megerőszakolni	Hungarian	UK, France, Germany, Hungary	No block, produces search results

Table 1*:* Table indicating search results for the term "rape" in different languages across different VPN locations using an incognito browser window (UK, France, Germany, Hungary).

Efforts which are highly focused on keyword enforcement are destined to fall short due to the vast and creative variations users employ to continue promoting this content, whether with malicious intent or in good faith. At the same time, they are likely to inadvertently limit important discussions on sexual and reproductive health and safety. ISD has found that TikTok is failing to manage negative effects in relation to OGBV alongside negative effects on civic discourse and freedom of expression, an act of balance it is required to perform under Article 34 of the DSA.

To mitigate its shortcomings, TikTok should:

- **Transparently outline the reasoning for blocking specific keywords** to make the information clearer to both the users and creators affected by these decisions. This should include an explanation of how they achieve a proportionate approach with regards to freedom of expression when blocking certain keywords. In addition, searches using keywords that produce content on sexual assault could prompt notes of support guiding users to helpful resources and at the same time produce search results for educational content and other non-violent content. This should be consistent across languages and geographies.
- **Apply more nuance and context in moderation decisions**. This should include implementing comprehensive and context-sensitive keyword detection mechanisms to ensure a more balanced and supportive strategy for monitoring and removing harmful content whilst preserving crucial conversations on sexual education and violence. At the same time, necessary support to survivors, activists and experts should be provided. This could include a human layer of moderation for specific contexts such as the discourse on rape and sexual assault in general.
- **Work closely with gender experts, survivors' organisations**, women's and LGBTQ+ organisations who focus on OGBV and content creators to better safeguard educational and meaningful conversations around gendered harms including rape and sexual assault.

Unequal safeguarding of users across locations and languages

ISD found that TikTok's moderation practices and safeguarding measures vary across different languages, locations and devices. ISD analysis of content moderation found significant inconsistencies for the term "rape" in English, French, Hungarian and German, particularly in the application of search filters and in the support offered to users encountering sexually violent content. For example, searches for "rape" on the desktop version triggered supportive messages and links to resources for survivors in the UK and Germany but produced no formal notices in France and Hungary. Additionally, mobile prompts in Hungarian, French and German, independent from the language settings of the phone, redirected users to English-language resources and UK-based hotlines. This raises concerns about TikTok's support based on language, location and device. These findings further call into question the platform's overall moderation practices.

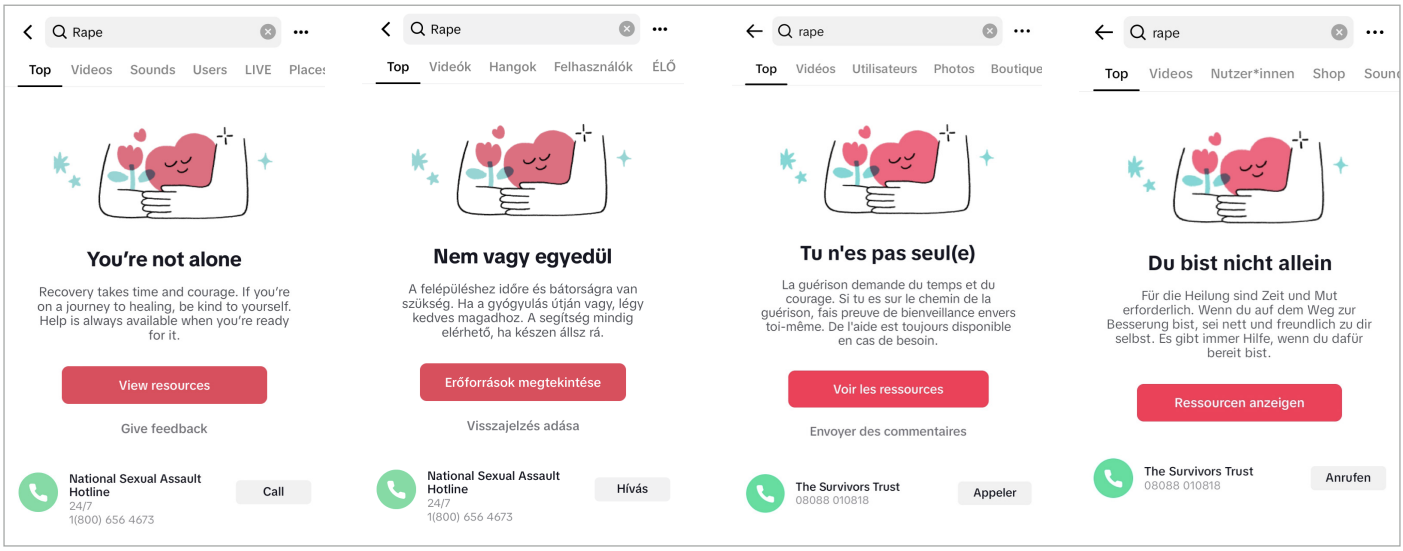

Figure 4: Screenshots showing a supportive message for survivors/victims of rape when searching for the English keyword "rape." (from top right: English, Hungarian, French and German)

Moreover, TikTok's allocation of content moderation resources neglects less widely spoken languages such as Hungarian. This leads to harmful content remaining unchecked within these communities. The platform currently has a disproportionate number of English-language moderators (2334) compared to other languages (e.g., French: 650; German: 837; Hungarian: 47). A crucial first step to addressing these gaps would be to allocate resources more proportionally across languages to ensure equitable access to support and information.

These differences in moderation practices and resources create an uneven playing field for users, as inconsistent moderation practices imply that users are not equally safeguarded across all locations and languages. To address these gaps, TikTok should:

- **Allocate proportional resources across languages to ensure equitable access to support and information**. According to the DSA Transparency Report from October 2024, TikTok currently employs 62 Hungarian-speaking moderators for 3.1m Hungarian users. This is proportionately lower than the numbers allocated for other Central and Eastern European countries: there are 63 Czech-speaking content moderators for 2m Czech users, 41 Slovakian content moderators for 1m Slovakian users and 41 Slovenian content moderators for 500k users. More generally, TikTok should make sure users of all languages are fairly and equally safeguarded and demonstrate that the resources applied to each language are proportionate and sufficient to ensure user safety.
- **Policies, mechanisms and content moderation teams should be inclusive and culturally sensitive.** This includes upskilling content moderation teams on issue-specific topics such as OGBV, local contexts and nuances of language. This will give them the ability to better address OGBV and other harmful content.
- **TikTok should further enhance transparency in its moderation decisions to ensure equitable access to support and information for all users.** Transparency reports should provide information on the different measures applied across locations and languages, especially regarding sexual and reproductive health content and other related thematic areas.

Normalised misogyny in Hungary and other widespread harmful but legal content

ISD also identified a range of harmful content. While often perceived as less overtly damaging than more extreme forms of online violence, this discourse fosters an environment conducive to misogynist radicalisation. This was especially evident in the analysis of male Hungarian TikTok users posting relationship advice, comments beneath videos of female candidates for the 2024 European Elections in France, Germany and Hungary, as well as the French Legislative Election. Such content fosters a culture that undermines gender equality and encourages the marginalisation of women in public and political spheres.

Hungarian misogynistic actors on TikTok were found to employ tactics such as slut-shaming, demonising women, stalking and emotional manipulation in videos sharing relationship advice, reinforcing toxic relationship dynamics and misogynistic narratives. They appear to reach a significant portion of the estimated 3.1m Hungarian TikTok users, with some accounts receiving more than 10m likes across their videos. In line with findings on global Manosphere actors, they strategically blend self-improvement content with misogynistic narratives and monetise through subscriptions, business ventures and cross-platform content strategies.

Theme	Prevalence among the misogynistic and anti-feminist videos and images identified
Reinforcing gender stereotypes, demonising women and slut-shaming	51.4 percent
Control, manipulation (vengeance, jealousy, possessiveness)	30.0 percent
Trivialising domestic violence, implying the use of physical or psychological violence	27.1 percent
Stalking	11.4 percent

Table 3: Share of themes identified in 70 relationship-advice-related videos on TikTok. 14 videos contained multiple themes. Topics overlap hence the total score is more than 100 percent.

This reflects a wider rise in online misogyny on social media platforms and fosters an online environment where women's autonomy is limited, abusive male dominance is normalised and women are objectified.

To improve this, TikTok should:

- **Address the spread of legal but harmful misogynistic content by moving from a strictly "content-based" approach focused on moderation and removal to a broader "systems-based" approach to online safety.** Rather than deplatforming harmful content, TikTok would instead rebuild its systems to demonetise harmful speech and reduce its technology-facilitated reach. This should prioritise user safety and transparency while upholding freedom of expression.
- **Refine content detection mechanisms to better address satirical, humorous and implicit content.** TikTok's success relies on playfulness, humour, memes, satire and trend-following. This means it is crucial for the platform to thoroughly study, investigate and understand when humorous and implicit narratives cross into potentially harmful territory.
- **Consider limiting access to monetisation tools (like ad-revenue sharing) for actors that spread harmful but legal content.** This is particularly important for accounts whose content may implicitly incite violence but fail to reach strict criteria for removal.

Beyond platforms: The role and responsibility of other stakeholders

Social media platforms like TikTok do not operate in isolation but are deeply intertwined with real-world events, societal dynamics, and political processes. The responsibility for coordinated action to combat OGBV and build safer, more inclusive online spaces is not limited to these platforms alone: it also extends to policymakers, media and civil society organisations.

Gaps in holding political representatives accountable

Political representatives publish and engage with TikTok content. This means they also play an active part in emerging trends that shape public discourse. For example, ISD found that some male candidates in the 2024 EU Parliamentary Elections, particularly those running for far-right parties, platformed and incited harmful language. Given TikTok's growing role as a news and information source, it is essential that stakeholders interact with content on the platform in a responsible manner.

TikTok defines accounts operated by political representatives – including policymakers, party officials, and candidates—as 'Government, Politician, and Political Party Accounts' (GPPPA). To ensure these accounts contribute responsibly to the online discourse, individuals operating them should:

- **recognise the broader dynamics of online hate and harassment which affect women and gender minoritised politicians disproportionately, especially during elections, and act accordingly.** Political representatives and their communication teams should prepare for possible online abuse by drafting appropriate response strategies; these should outline when to engage a platform's reporting mechanisms and/or law enforcement and consider other resources that might be necessary (e.g. mental health professionals).
- **foster a healthier discourse and reduce the risk of harm by developing and enforcing safeguarding strategies.** Political representatives should assign clear roles and responsibilities, especially for large accounts managed by multiple people. They should also make use of the content moderation tools provided to creators by TikTok. Male political representatives in particular should reflect whether their content might amplify gendered hate.

Flawed response mechanisms for gendered disinformation

The resurgence of the 'National Rape Day' hoax in 2024 serves as a recent example for the gaps in official responses to gendered disinformation and forms of OGBV. Over the span of four years, the 'National Rape Day' hoax has appeared in at least four countries and across three languages. After first going viral on English and French-language social media in 2021, this hoax resurfaced in Germany in April 2024 when a government official of the State of Berlin ("senator") publicly warned about it. German media outlets, blogs, and social media platforms jumped on the statement and contributed to the hoax' spread – despite previous experiences in the US and the UK.

Although no credible threats were identified, misinformation linked to the hoax had real-world impacts. These include fears of sexual assault, reduced freedom of movement for women, transgender, non-binary and genderqueer people, and the perpetuation of harmful myths around 'stranger danger' in discourse on sexual violence.

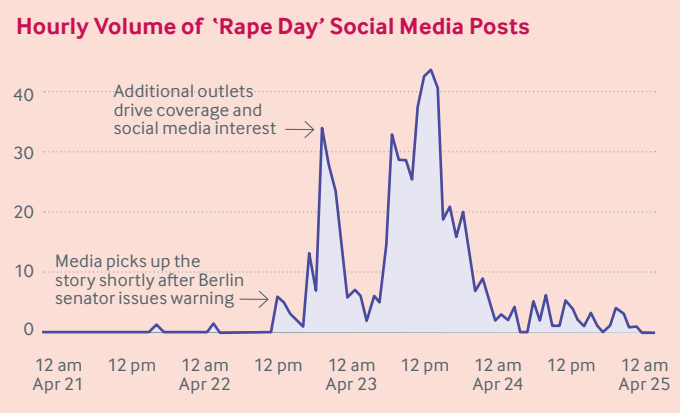

Figure 5: Hourly volume of social media posts in German referencing 'Rape Day' on the days around the warning. As signalled on the graph, the peaks of social media interest align with media coverage of the issue.

To avoid similar scenarios in future, governments and law enforcement should:

- **Provide training for politicians and their communication teams, particularly at the local and municipal levels** on effective communication strategies to address gendered disinformation campaigns and OGBV. This can help prevent officials from unintentionally reinforcing harmful narratives.
- **Collaborate with traditional media regulators, such as the German state media authorities (Landesmedienanstalten),** to develop guidelines for reporting on GBV and gendered disinformation. Ensure accountability mechanisms are in place for media outlets that fail to adhere to these standards.
- **Expand digital and media literacy training to equip schools, educators, public authorities and journalists** with the skills needed to critically assess and report on disinformation, with particular emphasis on gendered disinformation.
- **Strengthen support mechanisms for women and gender minoritised people in politics**, including legal resources, digital security training and emergency response frameworks.

Irresponsible reporting

Media attention was crucial in the spread of the 'National Rape Day' hoax in Germany in 2024. Yet, reporting on the hoax contravened established guidelines for reporting on gender-based violence and failed to effectively debunk misinformation. German far-right outlets successfully co-opted the narrative for their own benefit, while other articles failed to clarify from the outset that the story was a hoax.

This is not an isolated incident. From COVID-19 to elections around the world, experts have shown that media can play an important role in amplifying disinformation. This sometimes means bringing fringe beliefs into the mainstream – particularly with "alternative" media and blogs. Reporting on gendered disinformation and sexual crimes adds another layer of complexity, as seen when media outlets prioritised inflammatory coverage of the 'Rape Day' hoax over nuanced reporting on sexual violence which avoids sensationalism and aims for accurate and neutral language.

To ensure factual reporting and combat the spread of disinformation, media organisations and news outlets should:

- **Provide specialised trainings for journalists on recognising gendered disinformation and OGBV.** These should focus on reporting on such issues in a victim-centred, trauma-informed, gender-sensitive, accurate and factual way that avoids amplifying disinformation.
- Establish industry-wide partnerships to standardise best reporting practices on disinformation and OGBV. This should include stronger, more responsible guidelines for reporting on gender-based violence: ethical, fact-based coverage and clear labelling of misinformation are critical to preventing harmful impacts.
- Offer comprehensive and solution-oriented coverage that adequately captures gender-based violence:
- Collaborate with experts and other journalists to properly contextualise isolated abuse cases to address the systemic issues that foster gender-based violence.
- Encourage early intervention and prevention measures by focusing on stories that go beyond criminal justice proceedings, community support and governments' systemic response to GBV to hold perpetrators accountable and highlight the detrimental effects of victimisation.
- Provide resources for survivors, such as local, national and international services, and direct them towards specialised centres and support networks.
- Form specialised fact-checking teams within media outlets that focus on assessing gendered disinformation and OGBV reports. These teams can benefit from collaborating with public health and gender specialists.

Gaps in ensuring platform accountability

While the full impact of the DSA has yet to be assessed, ISD's research indicates potential shortcomings in ensuring platform accountability and the mitigation of OGBV. Key concerns include limited data access and a lack of algorithmic transparency – both areas the DSA aims to address. Despite the DSA's intent to address issues such as algorithmic transparency and data access, these provisions have yet to be effectively enforced.

Previous regulations, such as Germany's NetzDG (Network Enforcement Act), also had shortcomings, particularly concerning fringe platforms. This demonstrates the potential limitations of both national and EU-level frameworks in holding platforms accountable. The lack of consistent enforcement across jurisdictions continues to allow harmful content (including gendered disinformation and hate speech) to proliferate, undermining the broader goals of these regulations.

To address this, relevant stakeholders should:

- **Implement mandatory access to non-public data of providers of VLOPs and VLOSEs as mentioned in article 40(4) of the DSA. This should be done promptly so that vetted researchers can detect, identify and understand the negative impacts on users' fundamental rights and the continued exertion of gender-based violence. This access** should entail data on the reach of content and information on internal classification labels, where appropriate.
- **Consider ways to support national EU regulators to independently monitor and respond to the outcomes of algorithmic decision-making on platforms** when enforcing or supporting the enforcement of obligations for providers to manage risks stemming from the platform systems and their usage. This could be addressed through cooperation between the European Centre for Algorithmic Transparency (ECAT), the European Board for Digital Services, and national Digital Services Coordinators (DSCs) to ensure adequate personnel and financial resources.
- **Expand the existing requirements for recommender system transparency.** Article 27 of the DSA mandates that providers of online platforms make public the criteria used for determining recommendations and the reasons for the relevant importance of these parameters. However, this information is insufficient to understand how bias continues to be embedded and reproduced by the underlying systems. Therefore, platforms should be mandated to publish detailed information on how these criteria are factored in to produce results which 'match' user interests and search queries.

Conclusion

ISD identified crucial inconsistencies in TikTok's moderation strategy on OGBV-related content, alongside gaps in the scope and enforcement of the DSA. These include gendered bias in content moderation and the platform's search engine, inadequate responses to gendered disinformation, instances of over-blocking, unequal user protections across locations and languages, and the unchecked spread of harmful but legal content and normalised misogyny.

The findings illustrate how digital platforms like TikTok shape the landscape of gendered violence and how online manifestations of discrimination threaten individual rights while undermining democratic principles and institutions. The research primarily focused on TikTok, a platform whose short-form video content poses distinct challenges in terms of amplification and moderation. However, these insights carry broader implications for understanding OGBV across the digital ecosystem.

Each distinct analysis conducted within this project highlighted the same broader systemic challenge in combating OGBV: normalised misogynistic and discriminatory language. ISD found that such language, including remarks often dismissed as insignificant or in "bad taste", remains a significant systemic barrier. One in three people finds the act of verbally abusing women politicians on social media acceptable, according to a representative German study. The effectiveness of such degradation relies on its connection to deeply ingrained "common-sense misogyny" in mainstream society, which obscures and downplays the detrimental effects of online violence and harassment. This reinforces the harmful perception that for gender minorities, experiences of objectification and disproportional scrutiny are simply "the cost of doing politics". Consequently, addressing OGBV requires a holistic approach that recognises and tackles the deeply rooted belief systems that reproduce and reinforce these dynamics.

Outlook

ISD's findings further corroborate decades of research demonstrating the prevalence and adaptability of OGBV across social media platforms. Despite significant work to identify avenues for effective action, online violence targeted at women and girls continues to increase. At the same time, the rights of LGBTQ+ individuals, especially those identifying as transgender, are regressing. New strategies and insights are needed to mobilise users, platforms, regulators, media organisations, policymakers and other stakeholders to create fairer, safer and more just social media environments for all.

To address these challenges, future research must fill critical knowledge gaps, including:

- The specific impact of OGBV on minoritised gender identity groups like intersex and nonbinary individuals and transgender men and women,
- The influence of gender identity and expression in shaping patterns of OGBV,
- Exploring the effectiveness of moderation strategies in less widely spoken languages, where efforts may be inconsistent or insufficient,
- The role of implicit hate, coded language and other highly context-dependent content that evade automated detection,
- Examining interventions that can proactively direct users away from harmful content, encouraging healthier online behaviours, particularly for vulnerable groups,
- Investigating how to reduce the monetisation of misogynistic content and the financial incentives that enable its spread across platforms.

Moving forward, collaboration among different stakeholders is essential to ensuring that digital spaces uphold democratic values and safeguard users against gender-based harm. Addressing OGBV requires not only stronger enforcement measures but also a broader cultural shift—one that dismantles the systemic biases and normalised misogyny that continue to shape online discourse.

ISD | Institute for Strategic Dialogue

Amman | Berlin | London | Paris | Washington DC

Copyright ©ISD (2025).Das Institute for Strategic Dialogue (gGmbH) ist beim Amtsgericht Berlin-Charlottenburg registriert (HRB 207 328B). Die eingetragene Anschrift lautet c/o Schomerus & Partner mbB Berlin, Bülowstraße 66, 10783 Berlin. Geschäftsführerin ist Sarah Kennedy. Jegliches Kopieren, Vervielfältigen oder Verwerten des gesamten Dokuments oder eines Teils davon oder von Anhängen ist ohne vorherige schriftliche Genehmigung von ISD verboten. Alle Rechte vorbehalten.

www.isdglobal.org