SOCIAL STRIDE ADVOCACY
JOHN HUMPHREY CENTRE FOR PEACE & HUMAN RIGHTS

BY YOUTH FOR YOUTH

Learning About Online Hate & Advocacy - The Things No One Taught Us About

2022

ABSTRACT

As a result of public health measures, the use of online forms of communication and socializing is no longer optional but required. The reality of identity-based disinformation and discrimination urges to be addressed primarily because reporting and remediation mechanisms are inadequate and inaccessible to address online harm or pursue remedies for people subjected to online hate and discrimination.

This paper seeks to capture the learnings of the Social Stride network, a community-based, youth-led network that aims to address hate and discrimination online, using three different data collection methods: 5 focus groups with advocates, the SenseMaker® app to collect stories of advocates' experiences, and the analysis of the 57 cases reported by advocates between 2021 and 2022.

The data collected is analyzed and organized in this report to answer the following guiding questions:

1. What trends and best practices are young people using as remedies to address online hate, discrimination, and dis/misinformation?
2. What do young people in Canada see as necessary practices for other young people and institutions (e.g., media) to prevent online hate, discrimination, and dis/misinformation, and for remediation when these do happen?

ABSTRACT

Suite aux mesures de santé publique, l'utilisation des réseaux sociaux en ligne n'est plus facultative, mais obligatoire. La réalité de la désinformation et de la discrimination fondées sur l'identité doit être abordée principalement parce que les mécanismes de réparation sont inadéquats et inaccessibles pour traiter les préjudices en ligne ou rechercher des recours pour les personnes victimes de haine et de discrimination en ligne.

Cet article cherche à démontrer les enseignements du réseau Social Stride, un réseau communautaire dirigé par des jeunes qui visent à lutter contre la haine et la discrimination en ligne, en utilisant trois méthodes de collecte de data différente : cinq groupes de discussion avec des défenseurs, l'application SenseMaker® pour collecter récits d'expérience d'avocats et l'analyse des 57 cas signalés par des avocats entre 2021 et 2022.

Le data recueil sont analysées et organisées dans ce rapport pour répondre aux questions suivantes :

1. Quelles tendances et pratiques les jeunes utilisent-ils comme remèdes pour lutter contre la haine, la discrimination et la désinformation en ligne?
2. Quelles sont, selon les jeunes au Canada, les pratiques nécessaires pour les autres jeunes et institutions (par exemple, les médias) pour prévenir la haine, la discrimination et la désinformation en ligne et pour y remédier lorsque cela se produit?

By Youth For Youth

Learning About Online Hate & Advocacy - The Things No One Taught Us

Table of Contents

Background to Social Stride's Research

The idea for the Social Stride program emerged in 2020 when young people who had been actively participating in the John Humphrey Centre for Peace and Human Right's (JHC) Stride Advocacy project highlighted the need to provide young people across the country with the tools to address hate, discrimination and dis/misinformation in the online environment. Since then, Social Stride has trained and mentored a team of 38 young volunteer advocates across Canada, reported 57 incidents of hate and discrimination online, supported the exploration of remediation mechanisms to 15 of those 57 cases, and collectively designed digital literacy material to share their learnings online and offline. Notably, the latter is the direction that the Social Stride project is headed as the team continues to evolve the project to respond to community-identified needs.

The lack of theoretical and experiential literature on effective advocacy against hate and discrimination online motivated a research partnership between JHC, the Social Stride team of advocates and the Illuminate Lab housed at the University of Alberta (UofA). This research sought to capture the learnings of the Social Stride team using three different data collection methods: five focus groups with advocates, the use of the SenseMaker® app[1] to collect stories of advocates' experiences, and the analysis of the 57 cases reported by advocates between 2021 and 2022. The data collected is analysed and organized in this report to answer the following guiding questions: 1) What trends and best practices are young people using as remedies to address online hate, discrimination, and dis/misinformation? and 2) What do young people in Canada see as necessary practices for other young people and institutions (e.g.,

[1] SenseMaker® is a mixed methods data collection tool that allows users to collect narratives on smartphones or computers. It is a tool that allows for exploration aimed at improving a situation, with iterative cycles of data collection, analysis, and reflection. As data collection is ongoing, narratives accumulate over time. The SenseMaker® tool creates visual representations of the narratives that allow users to come together to 'make sense' of common themes and narratives they are seeing emerge. For this part of Social Stride's work, the youth team was invited to record stories on a daily basis of their experiences both monitoring online spaces for hate, discrimination and mis/disinformation and providing remediation support to people who experience online hate or discrimination or are the targets of mis/disinformation.

media) to prevent online hate, discrimination, and dis/misinformation, and for remediation when these do happen?

The five focus groups with advocates aimed to identify best practices and challenges in the work of Incident Reporters, as well as provided an opportunity for the team to collectively analyse the data gathered in the Incident Report forms and through SenseMaker®. While the focus groups fostered group reflection on best practices, the SenseMaker® app allowed for an in-depth and comprehensive understanding of how this work affected the young people involved and, drawing from this understanding, what they see as necessary for creating safe online spaces. The JHC staff and the Illuminate Lab team co-designed the interpretative framework of the SenseMaker® collection tool to make sure it allowed advocates to input daily stories of their experiences providing support and addressing online hate and discrimination in a way that captured the details and nuances of this work. The stories gathered through SenseMaker®, referred to as "micronarratives", were collected between November 2021 and February 2022. Drawing from the 57 cases, the Social Stride team also collected data on the nature of the nature of the incident including information on primary protected ground (e.g., race, gender, etc), and social media platforms in which hate and discrimination seemed to be most prominent. A database was created to house this data with the aim of using it to inform an Artificial Intelligence (AI) tool, but the diversity and extent of the incidents collected limited the capacity to narrow the AI and its effectiveness. The analysis of this data has been informing the digital literacy campaigns, outreach efforts, and the development of an agenda on key points and practices to hold actors accountable for hate, discrimination and dis/misinformation online.

In this report we organize the findings through four primary themes that emerged from the research: the relevance of online advocacy work, the impacts of hate and discrimination online in young people and their advocates, the tools available to look for remediation and the gaps we have found, and the definition and characterization of safe online spaces.

Background to Social Stride

Social Stride is an initiative by-youth-for-youth through which youth advocates support reporting and remediation in cases of hate and discrimination online. Through the work, three key positions emerged among the network: incident reporters, case managers and communications team. Within Social Stride, each youth advocate takes on a role of their choice which collectively contributes to addressing online hate, discrimination, misinformation and disinformation. Incident Reporters and Case Managers navigate reports of online hate, discrimination, dis/misinformation that they come across or that are reported directly to Social Stride. In either situation, cases of online hate, discrimination and dis/misinformation are documented via a reporting form which is in turn documented into the Social Stride database. If appropriate, Incident Reporters will also reach out to the person who has experienced hate or discrimination and ask if they would like to be relayed to support services and/or explore avenues of remediation. Other activities which Incident Reporters engage in include letter writing campaigns to media outlets and mass reporting of social media posts and accounts. The primary task of Case Managers is to ensure cases are moving forward. When Incident Reporters struggle with cases, primarily pursuing remediation avenues, Case Managers are responsible for identifying and pursuing alternative routes (e.g. letter writing campaigns, bring public awareness to an issue, solely providing support services, reporting to appropriate social media outlet, etc). And lastly, the the primary task of the Communications team is to research and curate digital literacy education content on Social Stride's Instagram and Twitter accounts. This content is often created in partnership with other organizations (e.g., Alberta Hate Crimes Committee, Don't Click! Youth Initiative) or in response to current events (e.g., anti-covid public health measures demonstrations). Lastly, this team of youth advocates is mentored, guided and led by staff at JHC, a non-profit based out of Edmonton, Alberta (Canada) in partnership with the Alberta Hate Crimes Committee.

One of Social Stride's major strengths is the team's ability to leverage partnerships. Responding to hate and discrimination online is a need for which there are not many

resources or remediation mechanisms available. Therefore, partnerships and research have been a key component in the development of the Social Stride processes and frameworks of advocacy. Among Social Stride key partners are grassroots, youth-led organizations (Canadian Voices Against Racism, La Connexional, Don't Click! Youth Initiative) and non-profit partners from across Canada (e.g., Jack.org, Alberta Hate Crimes Committee, REACH Edmonton), as well as a partnership with the University of Alberta.

Theme 1: The Relevance of Social Stride's Online Advocacy

The Canadian Race Relations Foundation & Abacus Data (2021)[2], as well as the report of the Special Rapporteur, Irene Khan, "Disinformation and Freedom of Opinion and Expression" (2021)[3], note that identity-based disinformation and misinformation is on the rise and that these incidents can be the precursor to and galvanize in-person violent attacks against marginalized community members. During the pandemic, anti-Asian discrimination and hate speech increased online at alarming rates, a trend that has translated to a 700% rise in anti-Asian violence in cities across Canada. Statistics Canada (2020)[4], has also recently reported that 33% of those who responded to a crowdsourcing survey initiative shared that they had experienced discrimination online during the pandemic. Further, Nanos Research and the Canadian Race Relations Foundation (2021)[5] published findings showing that nearly six in ten Canadians (58%) have stated that racist online content and behaviours is a "major problem" in this country, a finding which was consistent across all age groups and regions in Canada.

Research out of Europe has also noted that young people value and need more informal online spaces to report online hate and that processes need to be established for young people to receive appropriate support when they do make a

[2] Online Hate & Racism
[3] Disinformation and Freedom of Opinion and Expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan
[4] The Daily — Canadians' Perceptions of Personal Safety Since COVID-19
[5] Online Hate and Racism: Canadian's Experiences and What To Do About It

report (2019)[6]. Similarly, through Social Stride, we have noted that people often take to social media to increase public awareness of the harms they have experienced because of specific disinformation and misinformation they have seen online, however, they have no one and nowhere to access support and remedy[7]. Social Stride advocates clearly connect the relevance of their volunteer work within Social Stride to their personal experience of being at the receiving end of online hate for which they did not have any remediation mechanism available and were not prepared to address. One of the advocates shared:

> [A] lot of people who get into this work have experienced hate in some way, and that has been true for me as well. As a queer person and having that perspective, I didn't want other people who belonged to marginalized communities to have that or even worse experiences [...] it is important for Canada to confront the realities of our society.

In all focus groups the idea that young people who belong to Social Stride have not had the opportunity to learn digital literacy at home or at school was a prominent theme and, individually, as well as as a group, they were grateful for having learned the "etiquette and risks" of social media.

The reality of misinformation and disinformation on social media has become more urgent to address in the last two years. The COVID-19 pandemic has compelled society to blur the line between the online and offline world. For many, and perhaps for a large part of the Canadian population who have attended school, work, and parties through a screen, the dichotomy between the online world and the 'real world' does not exist anymore. In fact, the online world is part of the 'real world'. This is particularly the case for young people who have grown up with the internet as a part of their daily lives and for whom online spaces are important places of social interaction, places where they express and develop their identities, and places where they access various kinds of information on a daily basis (e.g., news, health information, professional opportunities, social activities and events). During the pandemic, interacting online rather than in-person has become a necessary protection against the risk of infection and the only way to keep connected to daily

[6] Hacking Online Hate: Building an Evidence Base for Researchers
[7] Online Hate and Advocacy: The Things No One Taught Us About

obligations and activities. Online environments are now our workplaces, our classrooms, and where we interact with friends and family. As public health measures ease, online spaces continue to be present and necessary nonetheless. Many workplaces have chosen to continue with a hybrid model of work (Harvard Business Review, 2021)[8], while some provincial governments are increasing investments into developing virtual schooling options (e.g., Aversa, 2022)[9] and many universities have maintained a number of courses and programs online (Piper, 2022)[10]. That said, these online spaces and interactions should not simply be understood as required or imposed upon people by external sources (e.g., employers, universities and other institutions). They are also sources of information, connection and social organizing, as demonstrated by anti-racism and social justice movements such as Black Lives Matter and the Indian Farmers' Protest.

Social Stride advocates describe their engagement online as "not optional." In fact, some of our volunteer advocates having a social media presence began as early as grade 6 or 7, when they were around 11 or 12 years old. As one of them attests:

> Some of us here are the first generation that grew up with a computer in a household. Today, it is much more relevant being online, and our online activity has been 'required' for people to be online and so much more true for youth who have to have a social media presence their entire life. The digital literacy piece [...] of our work is perhaps the most important because it has not been filled by anyone, at least in my experience. Youth are more aware of what is going on in the world, racism, hate, discrimination in general. Online presence gives us better access to information, and we are really curious to learn better about things like misinformation. The rest of the world has to catch up to where we are and we need support to face the hard parts of what we learn and are exposed to online.

On the one hand, research has highlighted the ways that engaging online (e.g., through social media, online communities, etc.) has a number of benefits, particularly for young people. For example, online communities can provide psychological and social support that is sometimes not available at home or school;

[8] Designing the Hybrid Office
[9] Ontario Schools will Offer Remote Learning for at Least One More Year
[10] Flexibility of Virtual Learning Prompts Some Post-secondary Students to Pursue More Online Studies

young people engage in a number of supportive activities online such as sharing media content, promoting the visibility of social issues, offering guidance to one another and building a sense of community and solidarity among people who identify as members of marginalized groups (Kaveladze, n.d.[11]; Ito et. al., 2020[12]). On the other hand, we know that engagement online can sometimes lead to harmful consequences. For example, Tao & Fisher (2022)[13] note that racialized young people are often the target of racism both "individually (directed to them personally) and vicarious (observed to be directed to others of one's race)". Additionally, they reported that social media is one of the most used online spaces by racialized young people to connect with peers and engage online. Thus, it is important to recognize that the harms of online hate, discrimination and dis/misinformation are most felt by racialized and other minoritized community members.

While early research on the topic of race and racism online argued that virtual spaces could reduce instances of racism that is often experienced in the offline world, more recent research has noted that, in fact, the online world makes racialized people more susceptible to hate and discrimination because of the need to reveal aspects of one's identities. Further, in addition to the lack of anonymity that was initially assumed would be part of the online world, technology often creates a type of permanent record of one's experiences, ideas and opinions, which may also make them more vulnerable to being targets of hate and discrimination (Tynes, 2015)[14]. In the same vein, online environments can create conditions for "bullying, fatalistic worldviews and mob mentalities," which can have serious mental health impacts that have been associated with increased depression and anxiety, as well as increased substance use in young people (Kaveladze, nd.). Similarly, extensive research has demonstrated that online hate and discrimination is associated with

[11] Yes, Online Communities Pose Risks for Young People, But They are Also Important Sources of Support
[12] Social Media and Youth Wellbeing: What We Know and Where We Could Go
[13] Exposure to Social Media Racial Discrimination and Mental Health Among Adolescents of Color
[14] Online Racial Discrimination: A Growing Problem for Adolescents

increased depression, anxiety, and lowered motivation in schools (Castaño-Pulgarín, Suárez-Betancur, TilanoVega, Herrera López, 2021)[15].

The line between discrimination online and offline is really thin in the experiences of members of Social Stride. For them, social media is a place where the ideologies that ground the offline world exist, collide, and converse. Hateful ideas online become hateful actions offline, hence, education and advocacy become fundamental to foster healthy social media spaces and to prevent hateful offline actions. An advocate shared that:

> A lot of the discourse of human rights issues and local issues happens online now - even the core ideas are moved online first, like the Farmers' Protest. I found in social media a space for activism as well as a space for posting pics and having fun. However, racism and discrimination became more evident when doing any activism. Now, I can use my skills to address misinformation. There needs to be education and training on the dangers of social media; hard and soft skills.

Most young people who participated in focus groups noted that the use of their personal social media to express their opinion on matters related to discrimination was interpreted as a direct invitation to people who did not share their point of view to post their disagreement and demand explanations from them. In short, public ideological or political confrontation was demanded at different levels of aggressiveness.

Given the increasing time that young people are required as well as desire to spend online, and the lack of avenues for redress (both legal and otherwise), we need to be concerned about their exposure (direct and indirect) to online hate, discrimination, dis/misinformation that targets them and their communities; the harms of which will be felt by them, their families and their communities. In her report, Special Rapporteur On the Promotion of the Right to Freedom of Opinion and Expression, Irene Khan (2021)[16] notes that disinformation and misinformation is not the only

[15] Internet, Social Media and Online Hate Speech: Systematic Review
[16] Disinformation and Freedom of Opinion and Expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan

problem; so too are the “reactive, inadequate, and opaque” responses from institutions that, at best, focus their responses on improving content moderation and, at worst, ignore human rights concerns related to the indirect, direct, collective and individual harms that online disinformation and misinformation causes. In line with what the Special Rapporteur describes, the lack of intentional and informed anti-discriminatory responses (including the simple act of moderation) from institutions that use social media to promote themselves is, perhaps, one of the most clear findings of this research. Volunteer advocates shared 57 stories in which 56% of them would have improved the outcome for the complainant with a change in the policy or online behaviour of an institution or a social media platform. The connection of this lack of intentional action against discrimination and dis/misinformation with the mental health of advocates and complainants will be further developed below. In the meantime, it is important to understand this point as an essential one when taking action and applying the learning of this report. In what follows, we highlight our learnings from the work done in Social Stride to support those who have been harmed by online hate, discrimination and dis/misinformation and to address the need to center young people in improving reporting and support mechanisms.

Theme 2: The How, the When, and the Impacts of Experiencing Online Hate and Discrimination

Online hate and discrimination starts at a young age and the frequency and severity of it increases with age as people spend more time online (Media Smarts, 2019)[17]. In line with the research literature we have noted above, Social Stride youth advocates noted in their work that when young people speak out online about discrimination and advocate for themselves and members of their communities, they can become targets of hate. For example, drawing from the experiences of the Social Stride team, people put a lot of pressure on social justice advocates and make assumptions that because one is vocal, they are obligated to respond (e.g. “Oh, you’re anti-racist. Tell me why racism is…”). Advocates are repeatedly required to back up and prove one's

[17] Young Canadians Pushing Back Against Online Hate

ideas, feelings and opinions. This is not only inherently problematic because of the ways that it pressures those who have to live with the harms of daily discrimination to also constantly educate others about discrimination in ways that make them increasingly vulnerable, but also, this kind of individual moderation comes with a high cost to people's mental well-being.

The Link Between Mental Health and Remediation

Though online hate and discrimination is pervasive, our experience supporting and reaching out to people who are targeted and to complainants who reach out to us for support, is that it is difficult for them to even identify discrimination or hate incidents because hate and discrimination has been normalized. Also, on some social media platforms it is really difficult to present evidence for reporting and in some cases there is no formal reporting mechanism. For example, on Instagram there is the option to report accounts that are "detected as in defiance of their Community Guidelines." However comments and accounts often appear and then disappear rapidly. This creates a situation in which complainants are harmed by the experience of discriminatory and hateful comments and dis/misinformation, and harmed by the experience of having to carry the burden to 'prove' their complaint in a world where evidence disappears in the blink of an eye. This combination exacerbates the harms of discrimination, hate and dis/misinformation, and makes reporting difficult and, in some cases, impossible and painful.

Additionally, the research with the Social Stride advocates highlighted that impacts on mental health are severe and increase with feelings of hopelessness and helplessness when remediation cannot be pursued because there is a sense of being unprotected and vulnerable. Importantly, the mental health of volunteer advocates were also impacted in cases where there was no reporting or remediation mechanisms available. In 89% of the stories told by Social Stride team members, volunteers identified cases they worked on as threatening to the advocate's mental health or a combination of their mental health and their digital safety. When asked to explore that perception of threat, advocates connected their mental health wellness with the intention they have to support the complainant creating a report or a

complaint. Thus, when all evidence has disappeared, or when media outlets or institutions are not open to discuss or explore remediation, advocates feel they cannot do their work and people are left hurt by the incident. The following triad shows all the responses given by volunteer advocates when asked to place a dot on the side of the triangle that better represents the threat they were facing when working on the specific case they were reporting on. As the triangle clearly shows, advocates overwhelmingly identify mental health disruption as the major risk of the work they do.

[18]

In line with research on this topic, it is imperative that we understand how deeply the impacts of online hate, discrimination and dis/misinformation are felt in every part of one's life, including those around them. Since the initial goal of the Social Stride project was to connect with and support youth who had experienced online hate

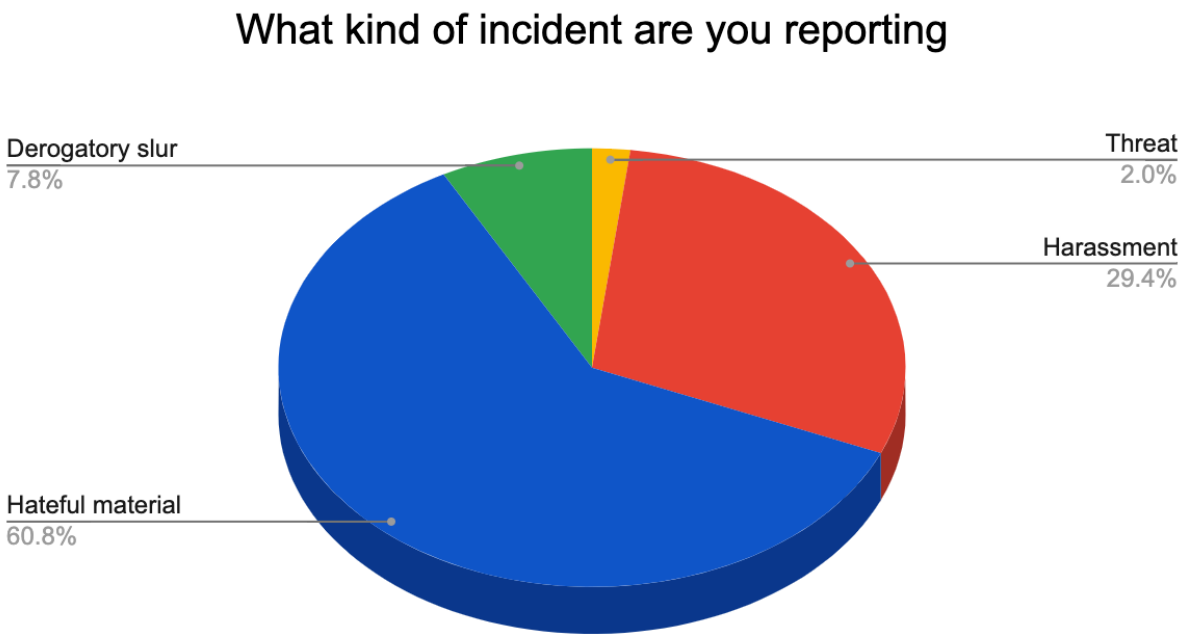
[18] Each dot in this triad represents a micronarrative submitted by a Social Stride volunteer through SenseMaker®. When responding to a question, such as 'What was threatened for you was?' volunteers completed a multiple choice with the options 'digital safety', 'physical safety' or 'mental health'. The closer the response is to a corner of the triad, the stronger that this description fits in relation to the experience.

and discrimination, we assumed that the greatest impacts of online hate, discrimination, and dis/misinformation would be felt mostly among young people. However, as the project continued we found that the work we were doing started to resonate with people in older age groups who wanted to engage in our public educational events and who approached us with an interest to volunteer in Social Stride. As a result, we see that there is a significant need for intergenerational education that deals with identifying and safely responding to online hate, discrimination and dis/misinformation as well as creating safe profiles online. As intermediaries, whether that be a parent, other relative, guardian, teacher, or youth support worker, they often worry about what young people in their lives are looking at online and how that affects their offline social relationships and skills. As it became clear that older age groups were wanting and needing education on these topics as well, Social Stride began extending training and workshop sessions beyond the intended youth audience: for example, we conducted workshops for Big Brothers Big Sisters Calgary youth support workers and mentors (see here for example). We were able to share our learnings with these individuals and help them develop an in-depth understanding of: digital literacy; various social media platforms/applications; the effects of online hate, discrimination and dis/misinformation; the importance of maintaining a social media presence; and, available reporting mechanisms. There is an important responsibility that adults play in this work. Young people are asking for support and role models who they can talk to and learn from, and who can support their online engagement, and thus, adults need to fulfill this role.

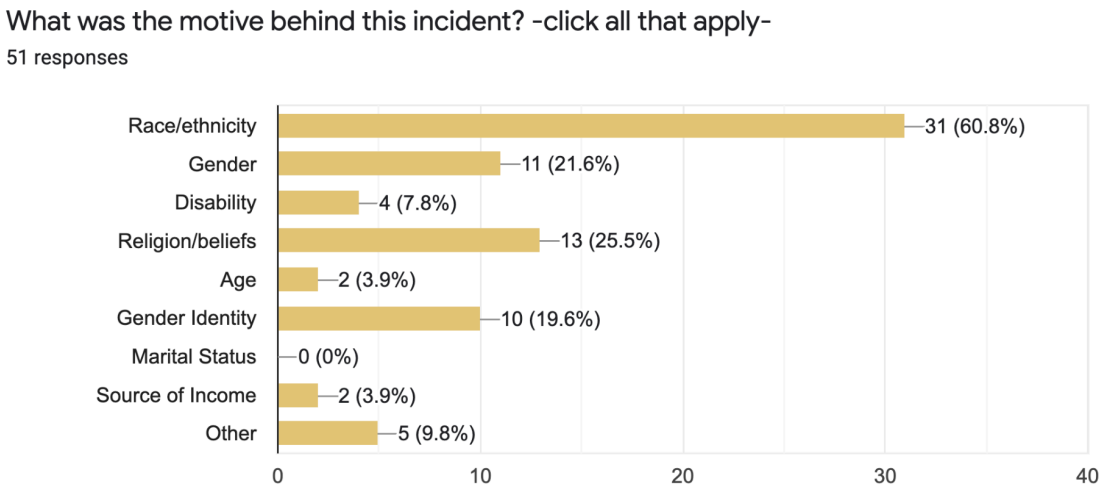
Common Forms of Hate and Discrimination

Looking at the numbers, Social Stride has collected data on what hate looks like online based on cases the youth advocates addressed. Importantly, the advocacy group and Incident Reporters were active primarily in Alberta until February 2021. After February, we expanded our advocacy scope nationally. Therefore, the data we are using to describe the tendencies of hate and discrimination are coming fundamentally from Central Alberta, where the group started and is more widely known.

Going deeper into the data collected by advocates in the Incident Report form, we noticed that they described the kind of incident they are reporting using four categories: 63.4% of the incidents are related to “hateful material”, 24.4% are categorized as “harassment”, 9.8% of the cases are described by advocates as “derogatory slurs”, and 2.4% are uttering a threat”.



Advocates received training to identify and document the motive of the cases they report (i.e. bias indicator). In the Incident Report forms collected by advocates, the most common instances of online hate are motivated by discrimination and prejudice about race/ethnicity, religion, gender and gender identity, and disability. There were also a few cases where the age and income of the person were the motive of hate or discrimination. Discrimination due to race and ethnicity represented 60.4% of the reported incidents, being the most visible and recurrent cases the team has worked with. Further, cases that sit at the intersection of religion and race represented 19.6% of the cases. Importantly, we noted that racialized Canadians that practice or identify with religions other than Christian, are the most targeted by hate and discrimination.



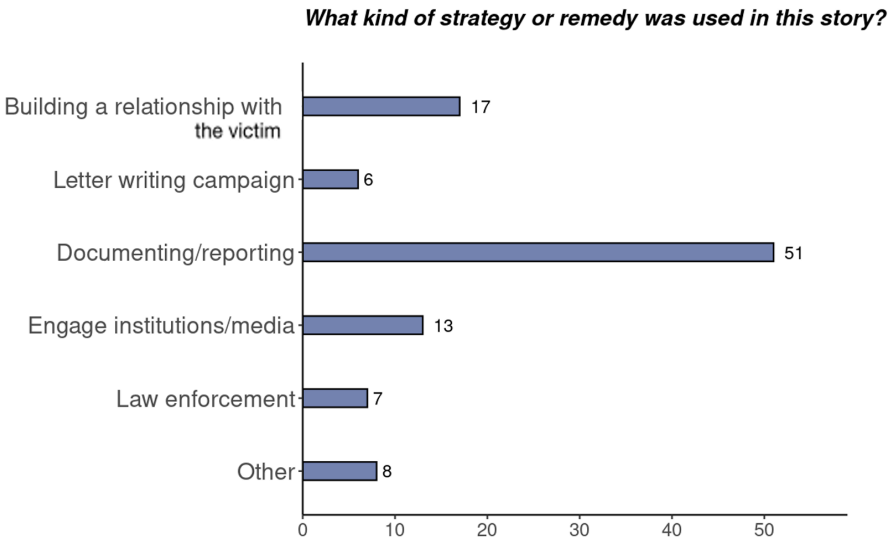
In one of the National Round Tables the Social Stride team hosted, presenters reflected on the common responses to online hate and discrimination in order to think through best practices to support. Some of them argue that there were three basic responses: (1) Fight; (2) Freeze; and (3) Flight. In the context of responding to a hateful message, fighting is often seen through replying to the post with hate. This response is counterproductive because it continues to cultivate an environment of hate online and uses discriminatory remarks to "offend" the person or organization who created pain in the first place. Freezing is usually seen through reposting, liking, discriminatory remarks without critically reflecting on them. This response often lacks tone and context and gives room to the continuation of stereotypes. Finally flight is seen when people ignores a discriminatory post or a comment because they understand their discriminatory remarks, but do not engage with it. Flying eventually results in the disappearance of that comment (e.g., it no longer shows up on people's feeds because no one is liking it or sharing it). This classification and available responses are not only important to understand in order to be able to provide appropriate support to people who have experienced online hate and discrimination, but also categorizing responses in this way reveals the complexities and tensions that young people must navigate when they come face-to-face with online hate, discrimination and dis/misinformation under a public eye and with all

the social pressure on them. This reflection also highlighted the importance of the work of Social Stride as our intention is to educate and support young people to seek remedy to hate and discrimination online in productive and effective ways. To this end, using a peer-support approach, we explored the best ways to respond to online hate, discrimination and dis/misinformation to meet the outcome desired by the complainants as well as to provide them with a wider spectrum of reactions than what this triad offer them.

Effective Strategies for Addressing Hate and Discrimination Online

Through our research we have been able to identify trends of how online hate, discrimination and dis/misinformation manifest on social media platforms. Most significantly, we have found that social media companies indirectly validate discriminatory posts/comments which has normalized hateful language online and, subsequently, made some of us more vulnerable to it. The normalization of hate, discrimination, and dis/misinformation online can also be understood as a larger social acceptance of discriminatory ideas if we see it through the lens of the online/offline relation between ideas and actions. When users come across hate, discrimination and/or dis/misinformation online they have the option of using reporting mechanisms provided by the social media platform which usually begins with a request for the post(s) or comment(s) to be removed. For example, if a racial slur is used in a comment to a post on an Instagram account that is not one's own, a person can choose to report it to Instagram for violating Community Guidelines. Drawing from Social Stride's work, on average it takes approximately 24 hours before platforms review a report and we have found that it can take up to 2-4 business days before users receive a response. Therefore, covert hateful content (that is not initialy detectable by artificial intelligence in the same way that overt discrimination is), can remain visible and widely shared by other users on that platform for a number of days before it is deleted by Instagram. The only instance when content is immediately dealt with is when reporting is done en masse (e.g., multiple users report the same post and/or content) because social media platforms will prioritize reviewing, moderating, and responding to content that has been reported multiple

times. Since mass reporting seems to be the most effective advocacy measure to remove hateful and discriminatory comments, it is not surprising that when talking about elements that affected the outcome of the case, Social Stride advocates identified 'Digital Community Support' as one of the most effective strategies. The following triad highlights how important strong digital community support can be to achieve a positive outcome when one reports online hate, discrimination and dis/misinformation. Here we see that 42% of stories told by advocates explained that what mattered in being able to achieve a successful outcome through reporting was having sufficient digital community support that could help the complainant amplify their concerns. This is an important finding because it tells us that building a trusted community online is important for addressing online hate, discrimination and dis/misinformation, and also because it highlights for us that those who do not have access to this kind of community may be even more vulnerable to experiencing this kind of online violence.



Notably, we also found that in 51 of the 57 stories, Social Stride advocates shared that they themselves had done the reporting for the victim of their experience of online hate, discrimination and dis/misinformation through different social media companies' reporting mechanisms. Furthermore, we found that when we did the reporting as a Social Stride advocate we saw faster responses from social media companies as well as from law enforcement in cases where they were also involved.

Through our work in Social Stride exploring various forms of remediation and reporting with the people who have experienced online hate and discrimination, we hope to create pathways for young people to safely and effectively address hate and discrimination online. Neither the Canadian legal framework nor the regulatory procedures of institutions or the government have provided effective ways to remediate hate and discrimination online. Young people are not provided support or education to immerse themselves in the vast, fast, and sometimes harmful world of social media. Social Stride is hoping to shine some light on these dimensions of the problem and the possible ways to address them. All forms of online hate and discrimination are harmful and impact every aspect of an individual's life. Unfortunately, online spaces are not something one can leave behind and choose to opt out of, and it should never come to the point where someone feels that this is the only choice they have left. As we note throughout this report, in this day and age,

social media facilitates social connections, professional opportunities, access to education, and more; thus, going completely offline is rarely (if ever) an option.

Theme 3: Tools and Strategies to Address the Reality of Online Hate and Discrimination

Research with Social Stride advocates has taught us that, in online contexts, it can be quite difficult to provide people who report experiences with hate and discrimination with viable avenues to pursue remedies and support. We have faced a number of legal and institutional barriers when supporting people with finding routes to remedy, healing and justice, including:

- the lack of immediate action by online and social media platforms to remove the hateful or discriminatory content;
- the threshold online hate incidents may need to reach in order to be considered a hate crime;
- the financial burden it costs and time it takes to pursue remedy through legal representation and the court system; and,
- the overall slowness and inefficiency of courts, law enforcement and human rights mechanisms to find adequate justice and remedy for people who experience online hate and discrimination and who are the targets of dis/misinformation.

Social Stride's research has found that because of these barriers, the most important support we can offer is often the connection with peers who can understand the difficulties of an issue (e.g., someone hacking a profile to spread hate from there); someone who could listen and who could help them translate their concerns in writing; and, someone who could connect them to the required social support (e.g., mental health supports). While these are all important, there is a critical difference between a sense of remedy, healing, and justice that is rooted in accountability to make change by those who have perpetrated the harm, versus individual-based support that addresses individual needs with no recourse for systemic redress. Thus, we maintain that an important tool for addressing online hate and discrimination is

to continue to put pressure on institutions that are responsible for ensuring safety in the online spaces that they have created and that they maintain (e.g., media institutions, social media companies). Additionally, we note that there are very few places where people can report online hate due to a number of factors, including the very high threshold for what counts as hate established in anti-hate legislation. In other words, once something is deemed as not fitting within the legal definition of a hate crime or hate speech, a person is left with little to no option for reporting the harm they have experienced and, subsequently, little to no option for remedy.

It is important to highlight that regardless of the response employed, an important lesson in this work has been the realization that the person providing support needs to address their own safety and well-being as well. We have found a number of measures needed to be in place in order to so, including choosing times throughout the week to 'unplug', knowing one's own capacities in finding remediation that meets people's needs (particularly within the context of the limits of legislation and current lack of accountability mechanisms for institutions to better monitor their own online spaces), and doing this work with a group of like-minded people who can support you as well as having the organizational infrastructure behind you to provide appropriate access to mental health support. More specifically, Incident Reporters and Case Managers have reported that they benefit from being able to collectively identify next steps and to work through difficult issues together.

One of the most important sources of information and learning of effective mechanisms to address hate and discrimination online were the National Round Tables hosted between January and March 2022. Particularly, the first two Round Tables addressed the ways in which youth-led organizations and advocates were addressing hate and discrimination in their own social media pages, the gaps they see through this work and the needs they have moving forward. In order for us to share these learnings, we have divided them into three categories, individual responses, community-based responses and legal responses.

Individual Responses

Individual responses are ones that are led by an individual (both by the person who experienced the hate and discrimination and also by individuals within organizations such as JHC/Social Stride). Individual responses that were employed and that Incident Reporters supported others in leading include: reporting cases of online hate, discrimination and dis/misinformation to existing mechanisms, despite their limits (e.g., law enforcement, StopHateAB.ca, reportinghate.ca, Social Stride); blocking accounts; disabling comments on one's personal account; deleting harmful comments; making decisions to exclude certain users from one's social media bubble (e.g., denying new user requests or blocking users); and, choosing to ignore and not respond to others' comments online. Additionally, we found that supporting individuals in setting personal boundaries and making choices to go online only when they feel that they are emotionally safe and healthy was an important response.

Social Stride advocates shared that they recognize that, unlike community-based responses, these individual responses are limited in terms of their capacity to address hate, discrimination and dis/misinformation from proliferating online. In addition to what is described above, as well as below, advocates have seen first-hand the inconvenience of available reporting mechanisms in terms of mobilizing systemic and institutional change. For example, in addition to the ineffectiveness of social media reporting mechanisms noted above, online trolls can reverse the effect of these mechanisms to target the advocate or person who experienced the online hate or discrimination instead. This was the case for Edmonton blogger Linda Hoang, who after speaking out about a discriminatory experience with Ottawa convoy supporters was mass reported by those same perpetrators which resulted in Instagram disabling her account. That said, while strategies such as shutting off comments may not address hate and discrimination directly, and while there are concerning barriers to accessing and using available mechanisms for reporting, when people do report they can make online spaces safer for those being targeted and can assist in moderation efforts by protecting both content creators and readers.

Further, reporting online hate, discrimination and dis/misinformation is vital to collecting data and exposing this issue.

Community-Based Responses

Community-based responses are deemed the most effective ones by young people who participated in the first Round Table, demonstrated through the research presented by Farshad Labbaf. The effectiveness is related to the fact that those responses are built from the experiences of people who are facing the effects of discrimination and the challenges and barriers to report and achieve the desired remediation. Therefore, it is important for organizations such as Social Stride to maintain their relevant work and continue to draw learnings and do research from their experience. Overall, community-based responses were described by Roundtable participants as effective, accessible and trustworthy. Community organizations and initiatives such as JHC and Social Stride have created tools and work in collaboration with other organizations to make their support accessible and center on the complainants' goal and safety. For example, given the lack of efficient and effective options for reporting online hate and discrimination, as well as the aim of Social Stride to act as a reporting and support mechanism, the team had to create tools to identify online hate, discrimination, misinformation and disinformation. To this end the Social Stride team monitored online spaces such as Facebook, Instagram, Twitter, blogs, Reddit, video streaming sites (e.g., Twitch, YouTube) and the comments sections of news articles. Specifically, trending hashtags were an important monitoring tool (e.g., #dropthet, #freedomconvoy) as well as colloquially popular buzzwords, keywords, and dog whistles (e.g., "make Alberta great again").

Importantly, most community-based responses have included research and education. Digital literacy and media literacy are important. Educating people on how to respond when they see online hate, discrimination and dis/misinformation and how to support those who have experienced it is critical for ensuring safe online spaces. Importantly, understanding digital media is necessary in addressing the structural roots that allow hate and discrimination to proliferate online and offline. Additionally, education for young people is more meaningful if the teaching content

and tools are created and led by young people. Thus, Social Stride developed tools for community-based responses that ensure that community and young people can work together to respond to hate motivated issues online. These tools are based on lived experiences, making them relevant and relatable to those who are the targets of hate and discrimination. These tools are intended to account for the gaps left by institutional responses (e.g., legal responses, reporting to police), in which people often face a lot of barriers in attempting to access (e.g., mistrust in institutions, lack of response from institutions, limited legislation that allows for any meaningful follow through) and for achieving any kind of remedy (e.g., these processes can be very lengthy, retraumatizing, and cost prohibitive). Examples of tools that have been created by Social Stride include tips on how to engage in by-stander intervention safely and effectively. Stop Race-Based Hate and Stop Hate AB are also two examples of community-based responses that we often share and explain how to use for reporting and responding.

Legal Responses

As discussed earlier, Canada has limited laws that define what online hate is and how to prosecute or how to pursue remedies for people who have experienced online hate and discrimination. Within the context of the current legal framework, freedom of speech has often been invoked to stop any attempt to limit online hate speech, as some jurists have argued that Canadian courts have infringed upon freedom of expression, a protected right under section 2(b) of the *Canadian Charter of Rights and Freedoms*, through adopting legislation or taking any action to restrict hate speech to protect marginalized groups of people (Gill, 2020).[19] Under section 2(b), the *Charter* protects "freedom of thought, belief, opinion and expression" which includes "freedom of the press and other media of communication"[20], recognizing that this right along with freedom of conscience, religion and peaceful assembly belongs to every person. However, it is important that we respond to attempts to use the right of freedom of speech to harm others; when people actively propagate hate, while

[19] The Legal Aspects of Hate Speech in Canada
[20] *Canadian Charter of Rights and Freedoms*, s 2(b), Part 1 of *The Constitution Act*, 1982, Schedule B to the *Canada Act 1982 (UK)*, 1982, c 11 [Charter].

they are technically exercising their freedom of speech, they are infringing on the rights of others through threatening their safety and right to life, liberty, and security of person.[21] Additionally, the right to freedom of speech has been used to oppress other marginalized groups historically. The most targeted group, as presented above according to the data collected by Social Stride, are young people who are not white and identify with minority faith groups or religions. This is a group who has been the target of online and offline discrimination and propagating hate against them excused under 'freedom of speech' further marginalizes them and continues to make them targets of identity-based verbal and physical attacks.

Hate speech directly contradicts the values underlying freedom of expression and *Charter* rights. It threatens the safety and well-being of its targets, usually comprised of vulnerable and marginalized communities. Hate speech silences and intimidates its victims and further limits their freedom to participate in civil society online. The *Charter* also posits that the rights and freedoms it enshrines can be limited 'reasonably' by the state where those limits "...can be demonstrably justified in a free and democratic society".[22] Although courts apply a legal analysis to determine what limits are demonstrably justified, through Social Stride we have learned that when people actively propagate hate, those who are marginalized are further threatened and intimidated from engaging in the online social sphere and in expressing their views. Additionally, while different platforms have created more ways for users to report and limit what other users can comment on their posts, social media companies cannot be relied on to solve the issue of addressing hate, discrimination and dis/misinformation online. Not only do they often profit from controversial posts, they also do not have the capacity to respond outside of automated responses. Instead, those major online platforms propose to address online hate with self-regulatory tools that have ended up targeting media literacy campaigns while still maintaining the conditions for rampant online hate and discrimination. What needs to change within the governing legal framework addressing online hate is:

[21] *Universal Declaration of Human Rights*, Art. 3
[22] Charter (1982) s 1.

- a robust definition of what online hate speech consists of, including veiled racism and discrimination;
- social media platforms should be required to remove hate speech expeditiously and in a more efficient manner than what is currently available;
- an impartial external Commission or Tribunal should be created and designated to oversee and monitor online hate and discrimination to make determinations of whether content contitutes hate speech with the powers to compel the removal of content from online social media platforms and websites; and,
- avenues for real remedies should be made available at the level of the federal human rights mechanisms for victims of online hate and discrimination to address online hate speech, including financial awards for victims, punitive measures against perpetrators, and the immediate removal of online hate speech posts.

On platforms such as Instagram, personal, professional and business accounts whose content includes or refers to any social justice issues or has advocacy-related content are at risk of being shadow banned[23]. Discussing online hate and discrimination —the subject Social Stride focuses on— is especially susceptible to shadow banning; Instagram claims this content violates their Community Guidelines because information within these campaigns have the ability to shift another users' opinion and perspective about how social media platforms and applications function. For example, an event poster describes a session where youth will discuss their experiences with online hate and discrimination, and the poster is then shadow banned due to the language on the poster describing online hate. Subsequently, the user is not able to boost (i.e. promote) the event on that same platform. Applications like Meta (formerly Facebook) and Instagram would state that the request was rejected because something within the content is prohibited according to their Community Guidelines. Most social media companies deny the existence of shadow banning and continue to dismiss users' personal experiences (e.g. a user's content is

[23] Shadow banning refers to making posts invisible by most accounts. Only close friends might be able to see post, even if they are public. This happens because the platform identifies in the post language that they consider dangerous.

hidden from a hashtag they have used, the prevention of their content from appearing on popular pages such as “for you”, “recent” and/or “explore”) (Nicholas, 2022)[24]. Documenting, reporting and creating records is key to demonstrating how hate and discrimination manifest online and to ensuring authorities will address perpetrators of online hate speech.

Theme 4: Youth Voices on Creating Safer Online Spaces

Notably, the research revealed that when people feel unsafe to interact in online spaces because they are fearful of being targeted, it has a strong impact on their overall wellbeing. We found that in 74% of the stories being shared through SenseMaker® youth advocates explained that the people they supported felt the impact on their self-esteem, their sense of identity, or both. In a similar vein, the stories shared revealed that the impact of the work was felt mostly on youth advocate’s mental health as explained above, and that the less resolution a case had, the more negative an impact suffered by the advocate. Importantly, not only did Social Stride members share that their mental health was generally affected by the work, but they also highlighted that it was often the inability to access an appropriate remedy mechanism or opportunity for healing for the person who experienced the online hate or discrimination that was the most difficult. This is unsurprising given how difficult it can be to access remedy, healing and justice for incidents of online hate and discrimination. However, what this reveals is that it is not only the experiences of online hate and discrimination in and of themselves that have an effect on people doing the work to combat this kind of online violence. The barriers to mobilizing remedy, healing and justice for people who have experienced online hate and discrimination also have a significant negative effect on their wellbeing and mental health.

[24] Shadowbanning Is Big Tech's Big Problem - The Atlantic

The current laws under the *Criminal Code of Canada* addressing hate and discrimination are woefully inadequate to address online harms.[25] The Government of Canada attempted to enact Bill C-36, legislation intended to modernize and revise laws on online hate and discrimination. Bill C-36 would have amended and restored Section 13 of the *Canadian Human Rights Act*[26] concerning online hate speech and expanded the powers of the Canadian Human Rights Tribunal. The amendment would have defined a new discriminatory practice of communicating hate speech online and to improve the complaints process available of online hate speech. These changes would have given anyone in Canada the ability to take action if they encountered hate speech online by filing a formal complaint to the Canadian Human Rights Commission. This would have been a change long called for by advocates and restored an important civil remedy for victims of hate speech. The second element of the bill amends the Criminal Code to improve the prosecution and prevention of hate crimes by redefining hatred in line with the Supreme Court of

[25] *Criminal Code of Canada*, ss. 318, 319, and 320.
[26] Canadian Human Rights Act, s. 13.

Canada definitions. Additionally, Bill C-36 would have also established a tool for those who fear, on reasonable grounds, that another person will commit a hate crime or hate propaganda offense. These individuals would be able to seek a court ordered peace bond (a court order that seeks to prevent criminal conduct by allowing a court to impose tailored conditions designed to prevent a crime from occurring) with the consent of the Attorney General to prevent those crimes from occurring. However, despite the Government of Canada's continued attempts to enact Bill C-36, legislation which would have provided stronger protections against hate speech, the opposition against the omnibus bill proved to be stronger and the legislation failed to make it past its First Reading in Parliament.[27] This continued lack of concrete and real remedies available to people who experience online hate has contributed to further frustration and victimization of those most marginalized in Canadian society.

The Social Stride team shared that coming together with strategies to do this work in a healthy way (e.g., identify days and times for the work so that you are not always online and exposed to the negativity of online hate incidents and discrimination, block times to be offline, check-in with yourself to see if you are in a good place to engage in this work, conduct team member check-ins regularly) was critical and created a sense of hope. The narrative experiences shared by the Social Stride team through SenseMaker® made it clear that while many of the stories shared were difficult, coming together to talk about the experience of bearing witness to people's harm allowed the group to reframe some of these experiences in positive ways which engendered a sense of hope and possibility in the work they were doing. For example, while the young people were impacted by the barriers to accessing avenues for remedy, healing and justice, they also found hope in what those they supported shared with them about how significant it was for them to be able to have someone to share the experience with and process the harm done. Additionally, the team identified that creating more possibilities for collaborative and collective work, such as those that Social Stride offers, is critical so that this work is not done in isolation by individuals. In fact, 78% of the stories had a combination of outreach and collaboration consolidation. Similarly, the Social Stride team highlighted the

[27] Bill C-36

significance of relationships with each other, community members and leaders, and other stakeholders in this work and noted that opportunities to come together in ways that employ restorative practices might allow for the co-creation of solutions to creating safe online spaces. Finally, there was a very strong articulation that simply not having an online presence is not an option. The reality is that people's identities and socialization are strongly connected to their online profiles and lives and they put a lot of time and intention into creating these spaces. When asked about a success story that the core team wanted to share as part of this report of learning, they all concluded that the core team, the growth in terms of learnings, and the research were a great summary and measure of Social Stride's success. We have faced many barriers and growing pains yet continue to come together to address those issues.

> [It is] really amazing to see how everyone has grown and how people have worked together throughout this year. I see myself grow more confident. Before my experience with Social Stride, I hated talking to people, but now I can count facilitation and outreach as part of my skillset. I have participated in Provincial consultations where I saw people that I admire. I have made connections with organizations such as Big Brother Big Sisters - Calgary with whom I talked about Social Stride and say we are doing things right and we can share our experience with you. I am so grateful to everyone here. There would've been nothing without everyone here. Social Stride exists now and has supported people, and has become a tangible work that I am proud of.

Conclusion

The findings from the research between JHC, UofA, AHCC and Social Stride advocates are in line with the notion of ontological security which has been studied in relation to the use of social media. Ontological security is the expression and confidence that one's self-identity exists over time. Ontological security is based on the notion that "...people have a fundamental desire for their existence to have inherent value" (Areni, Momeni, & Reynolds, 2022)[28]. Thus, it has been argued that "continuity of the self involves creating and maintaining meaningful connections between the past and the present such that the self is bolstered or affirmed in some

[28] Ontological Insecurity, Nostalgia, & Social Media: Viewing YouTube Videos of Old TV commercials Reestablishes Continuity of the Self Over Time

way" (ibid). Collectively, our need for ontological security has grown in response to our ever increasing and expansive (across time and space) social networks, the sometimes alarmingly quick pace at which change happens in our present-day world, an increasing sense of uncertainty and unpredictability, and increasing human longevity (Areni, 2019)[29]. Some studies have found that ontological security is a strong motive for engaging with social media that highlights and helps people reconnect with their memories of their past selves, as well as establishing connections with others through sharing autobiographical memories (Areni et. al., 2022). Similarly, studies have shown that the increasingly strong need for ontological security in our rapidly changing world drives a desire to continuously document one's daily events in order to create a linear and affirming life history, which social media provides us access to (Areni, 2019). That said, this desire is not only about writing one's life story, but about re-writing one's life story as well, which, again, social media allows us access to in a way that not many other forums do.

Social media, and technology more broadly (e.g., smartphones), has changed the ways we structure and understand our realities which historically relied on face-to-face interactions and is now largely built around our spatial and temporal distance from each other. Mass communication technologies have not only changed how we structure our lives more generally, but also at the level of our everyday lives, as we build in time to "post", "scroll", "like", "comment/reply", "search" and "play" that are both linked to current routines and create new ones (Amigo, Osorio & Bravo, 2017)[30]. Notably, this new way of being has come with benefits, and negative impacts as well. For example, on the one hand, the growing and diverse technologies that people have access to "facilitates problem solving and daily routines, and also [provides] an effective and efficient interface for interpersonal relationships" (ibid). On the other hand, people often express concern about "information overload, loss of privacy, the end of face-to-face interactions, among others" (ibid). Thus, it is incumbent on us as a society to ensure that these online spaces are safe.

[29] Ontological Security as an Unconscious Motive of Social Media Users
[30] Mobile Communication Technologies and Ontological Security

THANK YOU

Social Stride has been in the works for two years, and would like to extend an immense thanks to those that have supported and funded our efforts:

Alberta Hate Crimes Committee (AHCC)

John Humphrey Center for Peace and Human Rights (JHC)

Government of Canada Department of Canadian Heritage

REACH Edmonton

University of Alberta Illuminate Lab

And a final thank you to the youth and partner organizations that engaged in conversation and were willing to openly share their experiences with online hate and discrimination with the community:

Canadian Voices Against Racism (CVAR)

Don't Click! Youth Initiative

Jack.Org

La Connexional

Funded by the Government of Canada | Financé par le gouvernement du Canada | Canada

JOHN HUMPHREY CENTRE for PEACE and HUMAN RIGHTS