

SomeBuddy

Case study

Name of initiative

CrimeDetector

Implementing organization

SomeBuddy

Initiative type

Artificial intelligence (AI) system

About the initiative

The CrimeDetector system helps support children and adolescents who have potentially experienced online harassment. When children report incidents, such as cyberbullying, the system automatically analyzes the case using natural language processing (NLP) and prepares a first-aid kit for the affected child, which is always reviewed by a 'lawyer-in-the-loop', to provide legal and psychological advice.

Alignment with UNICEF Policy Guidance on AI for Children

The initiative has aimed to:

- Ensure inclusion of and for children
- Protect children's data and privacy
- Provide safety, transparency and accountability for children

Location

Finland and Sweden

Launched

2019

Overview



Researchers and academics have thoroughly explained how social media is often experienced by young users as an ungoverned space – a ‘wild west’ – where they lack a precise and robust understanding of their digital rights. With the support of legal authorities and legislation to handle the perpetration of online harassment, including cyberbullying, victims can seek legal protection. The core mission of SomeBuddy, a Finnish start-up, is to provide children and adolescents with a diligent, tailored legal assessment and psychological guidance when they are faced with online harassment.

SomeBuddy’s design is centred around the online experience of the most vulnerable target demographic – children aged 7 to 18 – who are turning to digital solutions more than ever for learning, socialization and play.

SomeBuddy’s design is centred around the online experience of the most vulnerable target demographic – children aged 7 to 18 – who are turning to digital solutions more than ever for learning, socialization and play. Not only has the digital service been conceived with the insights of social media experts and psychologists, child-rights experts and lawyers, but it was also built through active co-creation with children. From this inclusive learning process, a ‘first-aid’ platform was created to help minors report unwanted online harassment and cyberbullying. SomeBuddy uses a combination of artificial intelligence (AI) and human legal supervision to identify and categorize or diagnose types of online harassment, providing victims with a criminal assessment, as well as recommendations for the next steps.

Context and project origins

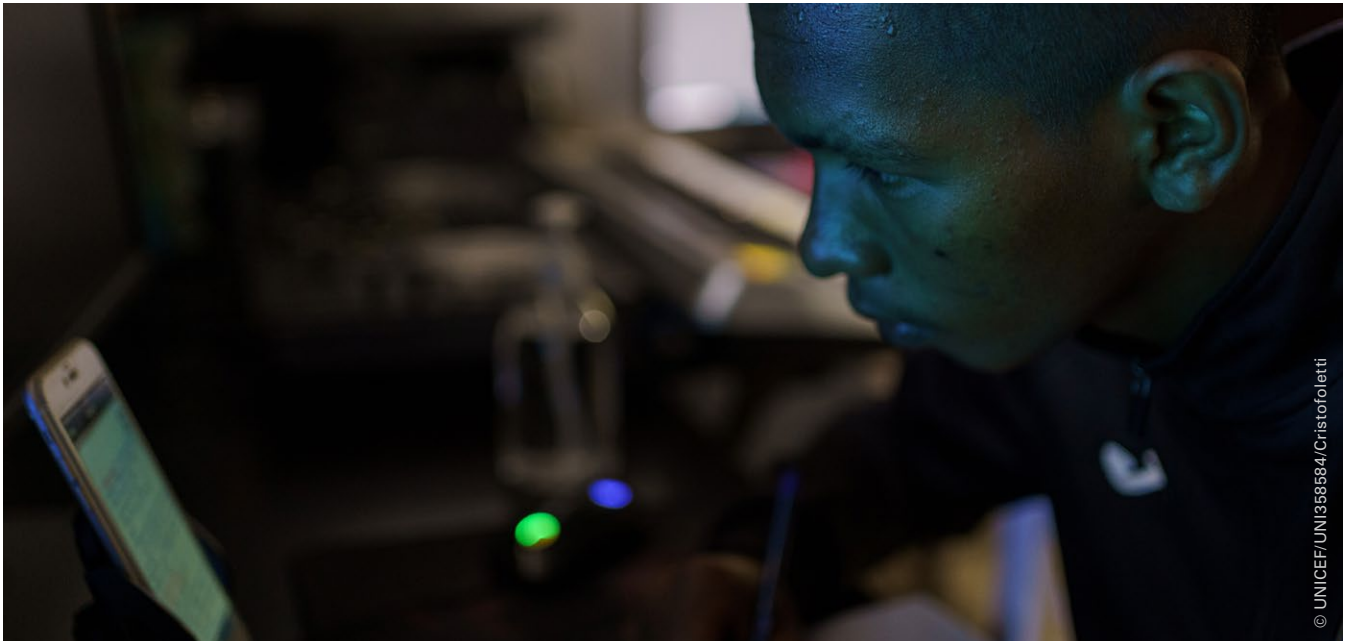
Cyberbullying is becoming an endemic problem with corrosive implications for children's affective and physical health and well-being. As defined by UNICEF, [cyberbullying](#) harnesses the use of digital technologies – social media, mobile phones, messaging and gaming platforms – and consists of repeated behaviour that is aimed at scaring, angering or shaming those who are targeted.

Yet SomeBuddy considers a larger spectrum of online harassment behaviours that may impact young users' well-being or constitute a threat to their safety. SomeBuddy's objective is to provide support in all unpleasant and conflictual situations that children may face on social media platforms and help define when these situations constitute a crime. While the most common types of online bullying or harassment among young people are insults, name-calling or verbal bashing, more serious cases of online harassment can involve death threats, rape threats, image-based sexual assault (also known as revenge porn) and sexual harassment. As social media and cyber harassment are such new phenomena, SomeBuddy's professional team has gone through a comprehensive literature and case law review to define a legal approach and framework in accordance with the criminal law of both Finland and Sweden.

SomeBuddy's objective is to provide support in all unpleasant and conflictual situations that children may face on social media platforms and help define when these situations constitute a crime.

A 2019 UNICEF poll indicates that many forms of online harassment, while not new, are [rising](#) in magnitude and scope for young users across the globe. In Finland, cyberbullying is a matter of concern that has gained increasing awareness. A 2018 poll commissioned by the Finnish branch of Save the Children has found that over [30 per cent](#) of respondents between the ages of 12 and 17 have witnessed some kind of sexual harassment of their peers on digital media in the last year. A 2020 study by Plan International Finland reveals that [42 per cent](#) of women and girls surveyed in the country had personally been harassed or abused online.

Victims of cyberbullying often suffer from social, physical and psychological disorders, ranging from lowered self-esteem, depression, anxiety and suicidal ideation. This harmful phenomenon is growing as internet connectivity increasingly becomes the standard method for young people to learn, bond and share life's experiences.



Recognizing the impact that digital technologies could have on children's emotional resilience, three female founders (a social media researcher, a lawyer and a school counsellor) shared, and translated into practice, a decade of multidisciplinary learning to establish SomeBuddy as a first-aid platform to help minors report unwanted online harassment and cyberbullying. The platform is based on an AI system which discerns children's colloquial writing, identifies potential crimes based on their responses, and then – confirmed by a human-in-the-loop – presents meticulous legal guidance and an empathic action plan that can be understood by young users. By gaining credible and anonymous age-appropriate legal advice through this system, young people are empowered to seek justice when needed and with the support of trusted adults. SomeBuddy understands that children are often hesitant to bring up acts of abuse with parents or teachers. Therefore, by providing an anonymous and online platform to report abuse and present concrete steps that minors can take for mediation or legal action, SomeBuddy seeks to reduce the discrepancy between reported incidences of online harassment and the actual amount of harassment that occurs.

After an initial pilot phase from 2018 to 2019, during which time 1,400 victims of online harassment were assisted, SomeBuddy started partnering with Finnish municipalities and schools, rolling out their service to protect school children. Currently, SomeBuddy operates in Finland and Sweden, but plans to expand its service internationally to better meet global needs.

About the AI system

SomeBuddy's purpose is to provide swift, robust and accurate legal and emotional support to victims of cyberbullying and online harassment. From its inception, SomeBuddy understood that utilizing AI to identify potential crimes and provide pre-prepared guidance would efficiently alleviate the strain on legal experts, thus reducing the time needed to respond to harassment claims and increasing the number of victims served. SomeBuddy has developed both a crime detection software and a 'survival kit', to help young users identify and report acts of online harassment while protecting their online privacy and safety.

To report online harassment, a user logs into SomeBuddy and fills in the interview tool, a fully automated questionnaire where they are asked to describe the incident. Upon completion of the questionnaire, SomeBuddy's CrimeDetector AI software examines the answers and determines if the incident fits into one of the pre-determined criminal categories. These categories correspond to the most common criminal offences reported to SomeBuddy in Finland, including for example, defamation, dissemination of information violating personal privacy, sexual offences against children, stalking and identity theft. As a next step, the AI software prepares a semi-tailored response, known as a survival kit, which determines whether or not a crime has been committed and provides a detailed list of steps that the young user may take.

Once this analytic phase is complete, SomeBuddy's humans-in-the-loop, a lawyer and a social psychologist, review the case. These experts ensure that the algorithmic analysis is correct, polish the remaining components of the legal and emotional support guidelines, and then authorize release of the survival kit. Via their digital platform, SomeBuddy then sends a detailed response to the victim, which includes a summary of findings and suggested supportive steps to take for mediation or legal action. This response is clear, concise and free of confusing legalese. In this way, a child easily understands whether they have been the victim of a crime and what measures they can take to resolve it.

FIGURE 1: HOW SOMEBUDDY WORKS, CASE OF LOTTE, AGE 12





Because SomeBuddy is designed with children in mind, and over 70 per cent of SomeBuddy's users are minors, the next steps are focused on helping the child to effectively address their victimization with a trusted authority figure. For example, a child who has received a pornographic picture from an adult, or who has been pushed to send a nude image to an adult, might be identified as a victim of child sexual abuse. The victim's survival kit would include a pre-prepared message with SomeBuddy's findings and recommended measures that the victim could send to their parents or caretakers, eliminating the need to re-live the incident. Additionally, if a crime has been committed, the survival kit includes instructions to report the incident to law enforcement.

Approximately 53 per cent of the cases reported to SomeBuddy are diagnosed as a crime by the digital service. The cases that are not defined as crimes are being treated with the same diligence by the conflict resolution and social media experts, so that these users will also receive the help and tools they need. In most cases, the user is experiencing a social conflict that is important to resolve whether the case is diagnosed as a crime or not. This need for mediation has been taken into account in assembling the professional team's competencies that represent the human-in-the-loop in SomeBuddy's system.

Application of child-centred requirements

SomeBuddy was designed from the ground up with children in mind, and the target users were included in the design process from day one. Within this inclusive learning journey, SomeBuddy focused intently on the requirements of inclusion, privacy, safety, transparency and accountability, as detailed in the [UNICEF Policy Guidance on AI for Children](#). These three requirements served as the bedrock of SomeBuddy's design philosophy because the team wanted:

- > Children to feel that SomeBuddy was an inclusive, fair and respectful environment where they could report harassment;
- > To limit data collection to only the most necessary information to make their legal assessment, such as age, but no names or other identifiers; and
- > Data from sensitive reports to be securely handled, stored and accessed. Users could trust that real lawyers prioritize victims' safety when reviewing the cases, offer a clear legal determination assessing if a crime had been committed, and provide concise and empathetic recommendations.

Ensure inclusion of and for children

To create this child-centred, AI-powered solution, SomeBuddy partnered with a multidisciplinary group of child psychologists, social media and child rights experts. Through hands-on engagement and a co-creation process, the start-up team also sought inputs from children, high schoolers, college students and YouTubers in order to understand how to best tailor their approach to youth. This meant understanding how a young user would want to engage with the platform, how they would speak with SomeBuddy and how the survival kit could offer optimal solutions that a user would actually rely on. For instance, during the pilot stage and before launching the web application, SomeBuddy partnered with students at Sipoonlahti School to ensure that the language used in the system was simple and child-friendly. Furthermore, prior to publishing a new website in 2020, the start-up's legal team sought inputs from upper secondary school students about the platform's design, language and style. This focus on young user engagement proved essential to bridge the gap between the legal professionals offering guidance and the victims who didn't necessarily communicate in the same manner.



To make the system's access as inclusive and wide-ranging as possible, SomeBuddy decided against creating a native app, opting for a browser-based experience instead. This is due to the fact that Nordic students typically have budget phones with limited memory. The need to install an app would therefore pose a roadblock to students wanting to report harassment, whereas a browser experience would not require downloading additional software. With the same inclusive ethos in mind, the legal start-up decided to offer an online platform as it lowers the threshold for reporting an incident. Since cyberbullying and diverse forms of harassment and abuse occur online, having a digital support service means that victims can remain in the same cyber environment to receive assistance. Lastly, SomeBuddy championed fairness and equality by expanding the languages offered from Finnish to Swedish, allowing Swedish-speaking minority groups in Finland, and most children in both Scandinavian countries, to use the service.

Protect children's data and privacy

Due diligence on privacy and children's data protection underpins SomeBuddy's decision to make its service anonymous. By ensuring that all reporting is divorced from personally identifiable information, the platform's design increases the likelihood that victims will seek help. Global cyberbullying [statistics](#) suggest that up to 50 per cent of students experience cyberbullying, but only 10 per cent tell a parent or other trusted adult. Protecting children's privacy is a core aspect of SomeBuddy's support, and has been a guiding principle throughout its iterative design process:

- > First, SomeBuddy opted to exist as an external (not embedded social media) service, thus allowing messages between SomeBuddy and its users to be encrypted and secure. If a social media website is hacked, SomeBuddy does not have to worry about its communications being breached.
- > Next, SomeBuddy has designed the login system so that the information used to login cannot be connected to private information related to the incident. SomeBuddy's users are only known by user IDs, allowing for maximum anonymity. Names of victims or assailants are not used in the automated interview, furthering their commitment to privacy.

Provide safety, transparency and accountability for children

With regard to safety, SomeBuddy stresses that its legal experts thoroughly review the cases to prevent what is called 'false positives' (a crime is identified when one was not actually committed) or 'false negatives' (no crime was detected when in fact one was committed). The AI software provides a sturdy backbone in the analysis, but its role is primarily to facilitate and support the legal expert's decision-making process. As false positives or false negatives could have considerable ramifications for victims, the legal expert ultimately remains the final arbiter on any decision, maximizing the safety and comfort of young users. This prudent combination of machine and human skills also contributes to transparency. SomeBuddy's users can be confident that a real human examined the facts of their case, explaining and justifying the legal reasoning process. The suggestions offered are therefore valid safeguards in procedural and legal terms.

In the design process, SomeBuddy initially trialed a virtual assistant interface for the interview tool but discovered that the chatbot proved ineffective since it was too conversational in nature. Due to children's curiosity to chat with the bot, the responses were too frenetic and hasty, and the reports didn't quite succeed in conveying the severity of all cases. This meant that information was not captured correctly, slowing down the process of providing legal and emotional guidance to users and presumably increasing the risk that an inappropriate response would be provided. SomeBuddy thus opted for a more simple, automated, but comprehensive interview questionnaire that gave users more time to elaborate on their written inputs, letting them pace their recollection effort. This design option helped improve and therefore streamline the intake process, allowing more users to receive the aid they needed – a strong example of SomeBuddy placing safety at the forefront of the design process.

With regard to safety, SomeBuddy stresses that its legal experts thoroughly review the cases to prevent what is called 'false positives' (a crime is identified when one was not actually committed) or 'false negatives' (no crime was detected when in fact one was committed).

Results, findings and impact

Since its deployment, SomeBuddy has enrolled over 25,000 users and has assisted 2,000 users who reported incidents of online harassment. Of those 2,000 incidents, 53 per cent were identified as criminal acts. Of those criminal acts, a large number were identified as defamation, a criminal act that encompasses spreading false information, embarrassing or insulting language. The next greatest number of offences were dissemination of information violating personal privacy and sexual abuse of children. Relying on SomeBuddy's ability to help them identify and report criminal acts, young users have better access to justice and online safety. This, in turn, reduces cybercrime, mental health spending, and streamlines work within the justice system. Over a three-year period, the legal start-up estimates that cost savings for schools, health-care partners and judicial institutions in Finland amount to 47 million euros.

FIGURE 2: SOMEBUDDY HAS PROTECTED SCHOOLCHILDREN IN THE FINNISH CITY OF TURKU SINCE THE FALL OF 2019.

These results cover an 8-month period

98

cases were handled by SomeBuddy's team of experts

85%

of cases were reported by students

39%

of the handled cases constituted an offence

2,517

students activated the service for their protection

15%

of cases were reported by teachers or other professionals

6%

of the handled cases had a risk of crime. Help has been given before the situation has escalated into a crime

Common platform used by offenders



Snapchat 26%



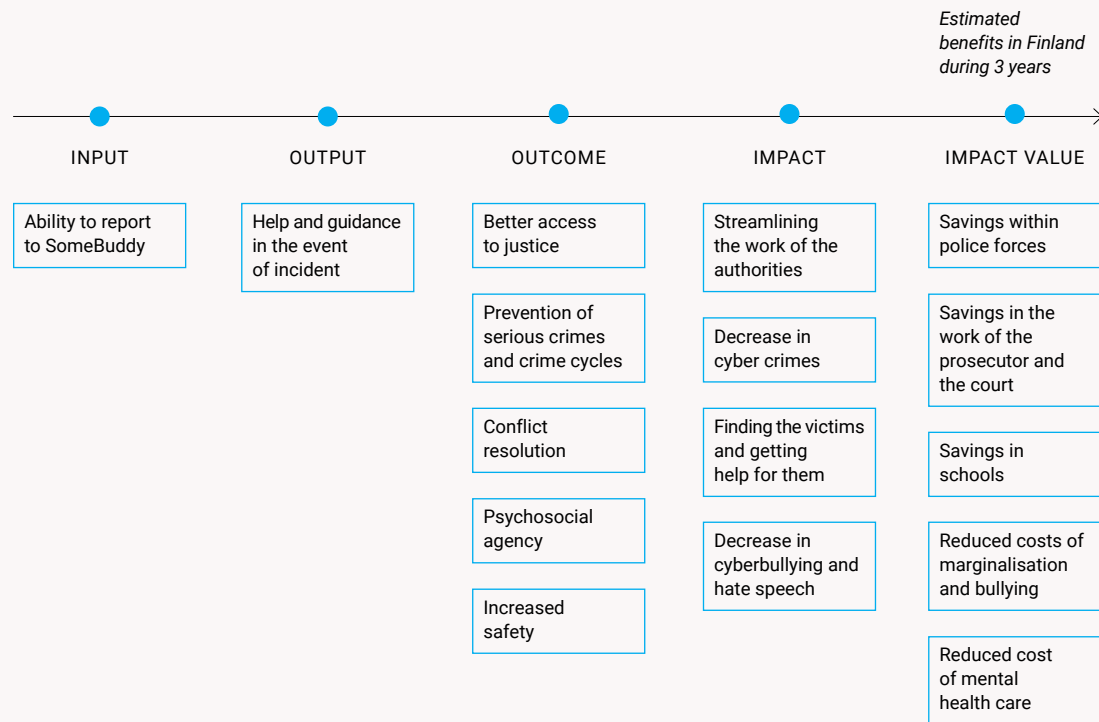
Instagram 19%



Whatsapp 16%

The most common criminal offences on social media include sexual commenting, suggesting or sending sexual images or other pornographic material to a child under 16 years of age and spreading sexual images of a minor under 18 years of age. These acts make up 34% of all sexual offences against children on social media. Offences involving defamation also account for 34% of illegal acts on social media, including the use of insulting language, shaming or embarrassing comments and spreading lies.

FIGURE 3: SOMEBUDDY IMPACT MODEL



Children have distinct rights that must be championed and upheld, including in cyberspace. Unfortunately, young online users are often not provided with a clear, tailored understanding of their digital rights, as well as what constitutes a cybercrime. While people may understand that it is a crime to assault someone on the street, they may not understand that someone sending them a sexually explicit message is also a crime. Ultimately, one of the greatest results of SomeBuddy is in empowering children, providing guidance and education.

About this case study

This case study is part of [UNICEF's Artificial Intelligence for Children project](#), which aims to better understand how AI-powered technologies can protect, provide for and empower children. Through a partnership with the Government of Finland, the project developed a draft [Policy Guidance on AI for Children](#), which contains practical recommendations and principles for governments and industry to create child-centred AI policies and systems. In order to inform and improve future versions of the guidance and inspire others, organizations were selected to pilot the guidance and share their diverse experiences to illustrate how AI-based initiatives could be designed to be more child-centred. The approaches taken, lessons learned and insights gathered will be one contribution to the global effort towards AI policies and systems that support children's development. View all of the case studies on our [website](#).

Contact

info@somebuddy.com

Acknowledgements

This case study was produced by the Office of Global Insight and Policy (OGIP) under the guidance of Melanie Penagos and Steven Vosloo, and was written by Eleonore Pauwels. UNICEF is grateful to Kaisa Rissanen and Eija-Leena Koponen from SomeBuddy for their support in developing this case study. It was copy-edited by Eve Leckey while design and layout was done by Mariana Amaral and Mardiyah Miller. Cover: Photography by M.T ElGassier/Unsplash; artwork by MacVell

UNICEF works in the world's toughest places to reach the most disadvantaged children and adolescents – and to protect the rights of every child, everywhere. Across 190 countries and territories, we do whatever it takes to help children survive, thrive and fulfill their potential, of early childhood through adolescence. And we never give up.

The Office of Global Insight and Policy serves as UNICEF's internal think-tank, investigating issues with implications for children, equipping the organization to more effectively shape the global discourse, and preparing it for the future by scanning the horizon for frontier issues and ways of working. With dedicated expertise in seven policy areas – digital technology, human capital, governance, the environment, society, markets, and finance – the Global Insight team assists the organization in interpreting, and engaging in, a rapidly changing world. Visit us online to learn more: unicef.org/globalinsight

Office of Global Insight and Policy
United Nations Children's Fund
3 United Nations Plaza, New York, NY, 10017, USA

© United Nations Children's Fund (UNICEF), August 2021

This is a working document. It has been prepared to facilitate the exchange of knowledge and to stimulate discussion. The text has not been edited to official publication standards and UNICEF accepts no responsibility for errors. The statements in this publication are the views of the author(s) and do not necessarily reflect the policies or the views of UNICEF. The designations in this publication do not imply an opinion on legal status of any country or territory, or of its authorities, or the delimitation of frontiers.



This document is interactive and designed for digital viewing.



Please consider the environment and refrain from printing.