Artificial intelligence (AI) has progressed rapidly recently, substantially impacting various aspects of our lives, transforming many aspects of our society, including communication, education, and entertainment. While AI systems offer immense potential to enhance our lives, they also bear that same potential to create negativity. Bullying, once confined to physical spaces like schools and workplaces, has now extended its reach into the virtual world due to the propagation of social media platforms, AI chatbots, and deepfake technology.

Social Media Platforms:

Social media platforms have become breeding grounds for bullying due to the expansion of AI-enabled algorithms developed by social media companies. These AI generated fake profiles aim to maximize user engagement, often by prioritizing controversial, sensational, and emotionally charged content such as political clickbait, bigoted memes and celebrity gossip. As a result, harmful content can garner more visibility, which perpetuates toxic interactions online.

AI algorithms can also analyze sensitive information shared online, such as personal details, psychological profiles, or private conversations and tailor content to specific individuals. Just the other day I was talking to a friend about wanting new shoes and an advert for the Atmosphere store chain came up on my phone while I was researching this article. I'm sure you've had this happen to you at some point. This specifically tailored content paves the way for personalized bullying, where AI systems can identify weaknesses or vulnerabilities and target victims with malicious content.

Chatbots:

AI driven chatbots are often designed to interact with users in a conversational manner. These seemingly innocent programs have also played a role in facilitating bullying through their ability to mimic human conversation. Cyberbullies exploit this technology by creating automated chatbots that harass, deceive and victimize individuals. Such AI-powered chatbots can be programmed to mimic human behavior, enabling them to harass, threaten, or provoke victims anonymously and to bombard victims with offensive messages, threats, or hate speech. It is near indistinguishable from a real person's online profile. Because it all happens automatically once cyberbullies

produced a clip that definitely looks like it could have been the real Mr. Cruise. There's also famous deepfakes of President Obama and many, many Arnold Schwarzenegger ones on Youtube.

Some of the aforementioned pieces are actually pretty funny, and it is easy to see why people are drawn to them. This technology, however, has opened the floodgates for cyberbullies to generate explicit, defamatory, and/or embarrassing content of just about anyone, and attribute it falsely to the victim. I certainly won't list them here, but there does exist a plethora of websites which can be used to create deepfake pornography without the victims even knowing. Using various programs, cyberbullies can seamlessly create and distribute content that deceives many individuals and destroys reputations. Just recently there was controversy over Bruce Willis and a Russian company who thought they actually owned the rights to his face. They were mistaken.

The emergence of deepfakes poses a dangerous and significant risk; the sophisticated manipulation of media means the friends, family or colleagues of victims may find it challenging to differentiate between genuine and altered content, leading to a myriad of outcomes. The victim may experience intensified social isolation, self-doubt, professional damage and sometimes, long-term psychological damage if deepfake content is believed over the truth. Moreover, deepfake technology undermines societal trust in digital media, which can bring forward a doubt in the usefulness of AI to humanity as a whole. While its detriments are significant, AI does stand to benefit us in many essential ways in almost all critical industries. These above examples, however, highlight how AI can be weaponized to perpetuate bullying

set them up, these chatbots can operate relentlessly, targeting vulnerable individuals at all times of the day so that there is never a second of peace. This inevitably will induce psychological distress, anxiety, panic attacks and feelings of helplessness in victims.

Deepfake Technology:

Deepfake technology is an application of AI that allows the manipulation of images, videos, and audio to create highly realistic videos, often indistinguishable from real footage. Have you seen any Tom Cruise deepfakes? They involved the work of a skilled actor along with algorithms known as Generative Adversarial Networks (GANs) and

and personal discrediting. Interestingly, if AI is used responsibly, it actually could play a pivotal role in combating these types of bullying by helping with early detection and prevention.

- AI-powered algorithms can analyze vast amounts of online content and identify signs of harassment, hate speech, or cyberbullying. By flagging such content, social media platforms and administrators become aware and can take necessary actions to tackle the issue promptly.

- AI can aid in monitoring the behavior of actual users, detecting patterns that indicate bullying tendencies, and issue warnings or intervene when necessary. This not only protects victims but also contributes to creating a safer online environment.

- AI can enhance the support networks available for those impacted by bullying and other traumas. Chatbots and virtual assistants powered by AI can provide instant support, guidance, and resources to victims. These tools can offer a non-judgmental and confidential space for victims to express their thoughts and feelings. Furthermore, AI can help connect victims with appropriate helplines, support groups, or counselors, providing them with essential assistance in dealing with emotional trauma. In order for this technology to be used to its most effective, social media platforms as well as AI developers, and policymakers must prioritize ethics when in the design and deployment stages of their AI systems. Guidelines can be established to ensure that AI algorithms prioritize user safety and well-being. For example, interdisciplinary collaborations involving psychologists, educators, and AI experts can help design AI systems that are more sensitive to linguistic nuances and contextual factors, thereby improving their accuracy in the detection of bullying behavior. Education also plays a crucial role in addressing the complex
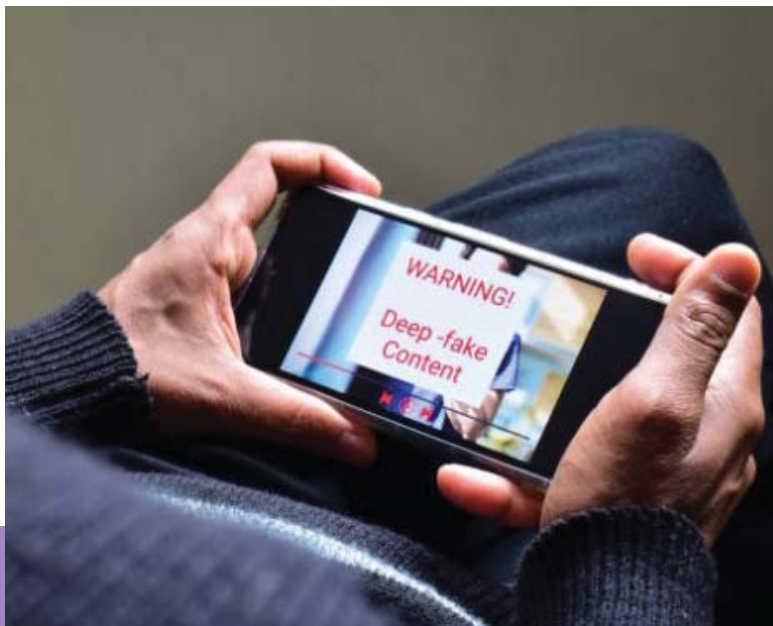
relationship between AI and bullying. Teaching early digital literacy, as well as general ethics to young people is essential to empower them as responsible digital citizens of the future. Equipping students with the knowledge and skills to navigate the online world, can help them understand the potential risks of AI-enabled bullying and encourage them to use technology responsibly.

As we have seen, AI possesses both positive and negative implications for bullying. On one hand, it offers solutions for early detection, prevention, and support for victims. On the other hand, it can be exploited to amplify bullying behavior through deepfake videos or malicious chatbots. To harness the potential benefits of AI while mitigating its adverse effects, responsible AI implementation, interdisciplinary collaborations, and education are integral. Ultimately, by utilizing AI ethically and responsibly, we can create a more inclusive and supportive digital environment, free from the menace of cyberbullying.

Thanks for reading.   Written by EDC Slobodian