

## LEGAL INFORMATICS

# The Evolution of Data and Freedom of Expression and Hate Speech Concerns with Artificial Intelligence

**Abstract:** This opinion article, by Channarong Intahchomphoo and Christian Tschirhart, explains the evolution of data and how it becomes useful information and then insightful knowledge. In the current era we are witnessing a high increase in the development and adaptation of artificial intelligence (AI) in society. AI technologies have the ability to process large volumes of data and information to help in finding insightful knowledge. However, AI is not perfect and there are ethical concerns, particularly when unintended negative consequences result from it; this paper also discusses ethical concerns currently confronting our society related to the freedom of expression and hate speech issues with AI. Importantly, this paper notes that governments are working to find ways to regulate social media and internet companies through legal channels as governments are no longer confident in the ability of social media and internet companies to self-regulate and thereby to guide society on what content is right or wrong. This is a critical new development in internet and AI governance that information and technology professionals and public and private organizations need to monitor closely the situation as it evolves.

**Keywords:** data; freedom of expression; hate speech; artificial intelligence; AI

## EVOLVING FROM DATA TO USEFUL INFORMATION AND INSIGHTFUL KNOWLEDGE

The evolution of science and related technologies has developed from research activities based on observation, lab experiments, animal and human trials, and testing inventions and hypotheses. Those activities are normally well-documented, with many small pieces of data to prove findings and research in-progress, when that research data is then documented and organized into categories, tables, or other formats, to be used later on to interpret the collected data in a way that can be easily understood using graphs, charts, and other visualization techniques. When the research data is organized to help in the analytical process, it then becomes 'useful information'. While presently researchers and organizations can gather a lot of data and organize a lot of information, it is harder now to determine what is actually useful information. For example, when we look at data collection from internet-based activities like the number of posts that people around the world generate and share on Twitter, Instagram, Facebook, YouTube or TikTok in each day or hour or minute, it is impossible for humans to organize

those data using traditional methods. There is simply too much data. To process such large volumes of data, often referred to as 'big data', humans have developed AI technologies to do the information organizational tasks. Social media companies and internet companies use AI technologies in their user interface systems to present potentially useful information users may find interesting through ads, news feeds, and other posts based on the user's previous interactions with their systems, web search activities, and other traceable criteria.

Another example of AI technologies being used to automatically process data and information which we would like to discuss is Covid-19 health data. As the whole world endured several years with great difficulty, health professionals and organizations at local, national, and global levels have been collecting data about the virus, including: the number of newly infected patients who tested positive with the deadly virus, the number of people hospitalized (including overall patients, and the number in intensive care), the number of people who have and have not received Covid-19 vaccinations, etc. There is a lot of related data if we look at the global scale. To understand such a large scale of data and to find

useful information as new data continues to be added every single day, data scientists and computer scientists have used AI technologies including machine learning and data or text mining techniques to find obvious and hidden data patterns and to find the 'insightful knowledge' to make predictions as to what might likely happen or to better understand the current situation. The data of hospitalized patients with Covid and their residential postal codes could be linked together to tell which neighborhoods are the most at risk of Covid-19 infection. Another way is to use the genome sequencing data of a given Covid variant to find insightful knowledge to support public health surveillance networks. The virus' genetic information can now be analysed faster by using computers and software with AI technologies to determine the genetic code of the virus.<sup>1</sup>

The examples above aim to describe the evaluation of data and how it changes to become useful information and insightful knowledge with the help of AI. The following section will describe ethical concerns of data, information and knowledge confronting our current society in the AI era, particularly related to the freedom of expression and hate speech issues.

## FREEDOM OF EXPRESSION AND HATE SPEECH CONCERNS WITH AI

Freedom of expression is the human right to seek the truth and propose new ideas while hate speech is considered as a limit to freedom of expression because of the negative and harmful consequences caused by such expression. To prevent hate speech, governments are legally responsible to step in to restrict its citizens through legal enforcement. Online hate speech mostly relates to harassment, cyberbullying, and defamation. An example of this is the 6 January 2021, Capitol riot in Washington, DC, after sitting President Donald Trump's defeat in the November 2020 presidential election. Facebook was blamed because they allowed posts from Trump supporters that contained hate speech to circulate on its platform. Facebook was also criticized for deciding not to take down the post by former President Trump on 28-29 May 2020, where, in reference to the George Floyd protests happening in across the US, he stated "*when the looting starts, the shooting starts*". This touches on the complexity of freedom of expression and hate speech with AI systems on social media.

If we look in more depth at the principle of freedom of expression – the term that is mentioned in Article 19 of the Universal Declaration of Human Rights,<sup>2</sup> as an example, in Canada, the term freedom of expression is also a core principle of the *Canadian Charter of Rights and Freedoms*.<sup>3</sup> Nonetheless, it is not an absolute right and freedom if one's expression causes harm to people. In a well-known 2001 case at the Supreme Court of Canada, *R Sharpe*,<sup>4</sup> the accused was charged with possession of

child pornography and he argued that his freedom of expression was infringed under the Canadian Charter of Rights and Freedoms. In the end, the court rejected his challenge based on the fact that the harms of child pornography outweighed the harms to his freedom and expression.

Hate speech is another challenging societal issue in the internet and AI era. We are starting to see more news reports on websites or social media channels from reputable news agencies deciding not to allow people to freely post personal comments on certain webpages and social media posts. This is because they have experienced a lot of hate speech comments, especially when the news piece is about people from a different race, gender, age, nationality, and so on. In the internet and AI era of today, judges and courts do have to power to order internet and social media companies to remove hate speech, defamation or harassment posted on their online platforms. But the legal process requires documentation and evidence to be submitted for consideration and it often takes some time and financial resources to draft a legal order or receive an official letter from the judge to send to the relevant tech companies. It really does not work well anymore, considering how fast and easily hate speech can be widely circulated on the internet and social media. That is why online harassment and cyberbullying remain a big concern in our society, with no end to it in sight. An example from November 2020, was when a young Canadian professional hockey player found guilty by a Swedish court after he took a photo of a young woman during a sexual encounter without her consent and uploaded the photo on Snapchat with her name and age attached and shared it with his hockey teammates who were playing with him in the professional league in Sweden.<sup>5</sup> This kind of online harassment and hate speech continues to happen.

In our previous academic research work on hate speech, we conducted and published a systematic review on social media and youth suicide in 2018.<sup>6</sup> That review showed that there are both positive and negative links between social media and youth suicide. The positive link concerns youth suicide prevention including detecting youth at risk of suicide, suicide prevention awareness campaigns, and consultations with youth at risk via social media. The negative link concerns how social media is used as a tool to pressure youth towards suicide through such means as cyberbullying, sexting, and disseminating harmful information.

There is now an expectation that the ethical censorship practice against online harassment and hate speech must be implemented first by the internet and social media companies. In recent years we have noticed that large technology companies have constantly updated their community guidelines on what users can and cannot post and how harmful posts and users will be dealt with, whether through deleting the harmful posts or through the outright banning of users found to have violated the rules from using their platforms. Facebook

calls their rules 'community standards',<sup>7</sup> YouTube names theirs 'community guidelines'<sup>8</sup> as is the case with TikTok,<sup>9</sup> whereas Twitter simply calls the 'rules'.<sup>10</sup> When we looked closely and compared those social media companies' community guidelines, we found that they all apply the same principles to prevent and mitigate the risks of societal and individual harm, against ideologies that promote hate. Interestingly, the YouTube community guidelines and policies clearly state that their policies development work is something that is 'never finished', which means that they are constantly evaluating their community standards. It reveals how difficult it is to effectively and fairly moderate and govern online content, even for big, well-resourced tech companies such as Google and Meta/Facebook.

People generate and share new content on social media every second and there is no perfect AI system able to detect all harmful content in real time. As we saw with the recent incident of Amazon's Alexa voice assistant giving a voice response suggesting that a 10-year-old girl "plug in a phone charger about halfway into a wall outlet, then touch a penny to the exposed prongs",<sup>11</sup> after it had received a voice command from the girl, who was seeking a suggestion for a fun physical activity to do as a challenge. Upon the request, AI systems inside the voice assistant went to search for a popular challenge activity that people posted on TikTok, YouTube and other social media platforms and unintentionally selected the dangerous activity for the innocent girl. This is a case of unintended negative outcome of AI systems. When Amazon teams were making Alexa and while running various tests before launching the system they may never have considered the need to filter out data such as the harmful videos that people share on social media, as the AI systems inside Alexa were designed to recognize social media posts about challenge activities as a 'good' data source that it could use and analyse and then select the potential best option matching with the input command received with the existing algorithms and then report back with the output (the voice suggestion) to the human user (the innocent girl). This is our explanation and argument as to why all social media and internet companies must know that their work to develop community guidelines must never end; it is a living document of policies and practices constantly changing. Therefore, we think that

social media and internet companies need to invest more human, financial and other resources into their community standards teams with the main aim of detecting and deleting the harmful and hateful content as soon as possible.

Unfortunately, most countries do not yet have a national body that is specifically responsible for internet content and AI governance. This includes, for example, Canada where the government has the Canadian Radio-Television and Telecommunication Commission. For the most part they regulate internet services and then there is the Office of the Privacy Commissioner of Canada, which has a mandate to protect business and government data;<sup>12</sup> so we end up relying on social media and internet companies to self-regulate content on their platforms.

However, currently there are a lot of discussions about how we need governments to step in to find ways to regulate social media, the internet, and AI companies through the legal channel because many governments and experts see the self-regulation performed by the tech companies as inadequate for the task of guiding society on what is right or wrong content, given the harmful content users are allowed to generate and share on their platforms. Recently, the head of a US Senate panel stated during a hearing on Instagram's negative effects on young people that "*I believe that the time for self-policing and self-regulation [of content on Instagram] is over*".<sup>13</sup> This could be a critical development in internet and AI governance that information and technology professionals and public and private organizations need to monitor closely.

## CONCLUSION

The meaning and application of data has changed over the years; it has become a powerful resource for providing insight into many societal issues. With the wide usage of internet, social media and AI technologies, this paper notes that we are now seeing increasing indications that governments will step in to regulate internet and AI technologies used by social media and other high-tech products, to prevent the potential risks or harms to society and individuals. Information and technology professionals need to pay attention to this recent development in internet and AI governance.

## Footnotes

<sup>1</sup> Andre Hudson and Crista Wadsworth, (2021) 'Genomic Sequencing: Here's How Researchers Identify Omicron and Other COVID-19 Variants'. <https://theconversation.com/genomic-sequencing-heres-how-researchers-identify-omicron-and-other-covid-19-variants-172935> accessed 2022.

<sup>2</sup> United Nations, 'Universal Declaration of Human Rights'. <https://www.un.org/en/about-us/universal-declaration-of-human-rights> accessed 2022.

<sup>3</sup> Department of Justice, Government of Canada. Constitution Act, 1982: Part I Canadian charter of rights and freedoms. <https://laws-lois.justice.gc.ca/eng/const/page-12.html> accessed 2022.

<sup>4</sup> R v Sharpe [2001] 1 SCR 45. <https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/1837/index.do> accessed 2022.

- <sup>5</sup> Daniel J Rowe, (2021) 'I know I Caused a Lot of Harm': Habs Draft Pick Logan Mailloux Apologizes for Sharing Sexually Explicit Photos Without Consent'. <https://montreal.ctvnews.ca/i-know-i-caused-a-lot-of-harm-habs-draft-pick-logan-mailloux-apologizes-for-sharing-sexually-explicit-photos-without-consent-1.5521894> accessed 2022.
- <sup>6</sup> Channarong Intahchomphoo, (2018) 'Social Media and Youth Suicide: A Systematic Review'. *Proceedings of the 2018 European Conference on Information Systems: Beyond Digitization-Facets of Socio-Technical Change*, Portsmouth, UK. [https://aisel.aisnet.org/ecis2018\\_rp/13/](https://aisel.aisnet.org/ecis2018_rp/13/) accessed 2022.
- <sup>7</sup> Meta Platforms. (2022) Facebook Community Standards. <https://transparency.fb.com/policies/community-standards/> accessed 2022.
- <sup>8</sup> YouTube. (2022) YouTube's Community Guidelines. <https://support.google.com/youtube/answer/9288567?hl=en> accessed 2022.
- <sup>9</sup> TikTok (2022) Community Guidelines. <https://www.tiktok.com/community-guidelines?lang=en> accessed 2022.
- <sup>10</sup> Twitter (2022) The Twitter Rules. <https://help.twitter.com/en/rules-and-policies/twitter-rules> accessed 2022.
- <sup>11</sup> BBC News. (2021) 'Alexa Tells 10-year-old Girl to Touch Live Plug With Penny'. <https://www.bbc.com/news/technology-59810383> accessed 2022.
- <sup>12</sup> Office of the Privacy Commissioner of Canada. (2020) *Investigations*. <https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/> accessed 2022.
- <sup>13</sup> CBC News. (2021) 'U.S. Senators Go After Head of Instagram Over How Platform Can Harm Children'. <https://www.cbc.ca/news/world/instagram-senate-hearing-1.6278644> accessed 2022.

## Biographies

**Channarong Intahchomphoo**, PhD, cinta088@uottawa.ca, is a Research Fellow in Digital Transformation and Innovation at the Faculty of Engineering, University of Ottawa, Canada, and an Affiliated Researcher at the Canadian Robotics and AI Ethical Design Lab (CRAiEDL). Channarong is currently also employed as a Data Research Librarian at the University of Ottawa Library.

**Christian Tschirhart**, MA, is an independent scholar holding two master's degrees in Global Studies from the University of Vienna (Austria) and the University of Leipzig (Germany).