**Check against delivery**



**Statement by David Kaye**

**SPECIAL RAPPORTEUR ON THE**

**PROMOTION AND PROTECTION OF THE RIGHT TO FREEDOM OF OPINION AND EXPRESSION**

General Assembly

Third Committee

Item 69 (b & c)

22 October 2018

New York



**Mr. President, Excellencies, Ladies and Gentlemen,**

This year has been a difficult one for the freedom of expression around the world, and it remains dire. I speak to you while the international community, including the United Nations, has failed to consider how to address the disappearance and apparent murder of Saudi journalist, Jamal Khashoggi. Just one year ago a brave Maltese journalist, Daphne Caruana Galizia, was murdered by a car bomb in front of her home. The Committee to Protect Journalists conservatively estimates that forty-four reporters have been killed this year alone.

The President of the United States attacks the media as enemies of the people. Hundreds of journalists and activists remain imprisoned by Turkey. Myanmar sentenced two Reuters journalists to prison for reporting on massacres committed against Rohinghya Muslims and other minorities in Rakhine State. Belarus adopted a harsh new media law, cracking down on journalists and publishers. Nicaragua has pursued charges and repression of dissenting voices. Tajikistan convicted anti-corruption journalist Khayrullo Mirsaidov on the pretext of embezzlement, sentencing him to a dozen year in prison. Iran continues to crack down on journalists and media – including on the BBC Persian service – not just the reporters but their families. The Phillippines Government of Rodrigo Duterte has taken to intimidating independent media outlets.

We have seen legal attacks on civil society in Hungary. Cambodia harshly cracked down on media and civil society ahead of elections in the summer. Russia sought to shut down the popular messaging site Telegram, as did Iran. The United Arab Emirates continues to detain human rights defender Ahmed Mansoor, Bahrain continues to detain several activists including Nabeel Rajab, and Egypt detained the well-known poet Galal El Behairy, after writing a song critical of government policies. China’s detention facilities in Xinjian seem designed to repress not merely freedom of expression but the freedom to maintain an opinion without interference.

In digital space it is much the same. Governments around the world continue to shut down network access, from Cameroon to Kashmir, while others filter sites that offer their citizens opportunities for dissent and accurate news. Several governments have taken advantage of the space allowed by major social media companies for propaganda and interference in foreign elections. State and non-state actors have tried to amplify their messages of hate and incitement to violence and discrimination, egging on violence against refugees in Germany and Italy, and ethnic cleansing in Myanmar. I have observed continuing pressure on the tools of digital security, especially encryption and anonymity, in a range of laws, including proposals considered presently in Australia. To many around the world, their one outlet for expression – social media – has implemented rules restricting speech in a black box of opacity.

I apologize for having to begin on a grim note, but while I will spend a few moment describing my formal report, I could not begin by ignoring the vast suffering that mainly governments are causing individuals around the world today. The repression of expression is repression of democracy and rule of law. It is repression of innovation, self-exploration, and connection. I cannot urge you strongly enough to take steps to reverse and resist this trend. I urge your leaders to speak the language of respect for reporting as the crucial public watchdog. I urge you to implement in deed the important normative measures the Human Rights Council adopted earlier this month in its resolution on the safety of journalists. The UN cannot continue on with high-level commitments and limited implementation. That is a recipe for cynicism about the work that you do here, and I dearly hope you can change it.

To be sure, there have been some surprising bright spots over the past year. In Malaysia, a new government has promised to take action to reconsider and repeal some of its most speech repressive laws. In Ecuador, which I visited just two weeks ago, the government has made commitments and taken steps to return its laws and policies to a place that is respectful of freedom of expression. The Gambia has been taking steps toward reversing the repressive environment that persisted for decades. Liberia has taken steps to decriminalize defamation and promote protections for a free press, though it has far to go.

**The Current Thematic Report: Artificial Intelligence and the Freedom of Expression**

This introduction may seem unusual given the subject of my report this year, the role of artificial intelligence (or AI) on enjoyment of freedom of opinion and expression. But they are connected. It is critical that even in this moment of global repression of freedom of opinion and expression we lay the normative groundwork for technologies designed to maintain respect for fundamental freedoms.

In my last report to the Human Rights Council (A/HRC/38/35), I sought to answer basic questions about the role of major social media and search companies: What responsibilities do companies have to ensure that their platforms do not interfere with rights guaranteed under international law? What standards should they apply to content moderation? Should States regulate commercial content moderation and, if so, how? The law expects transparency and accountability from States to mitigate threats to freedom of expression. Should we expect the same of private actors? What do the processes of protection and remedy look like in the digital age?

I reviewed the ways in which governments are seeking to regulate internet companies and urged a number of steps to ensure that they do not undermine freedom of expression in doing so. I also detailed the ways in which companies govern expression and the concerns that governance raises for freedom of opinion and expression. I emphasized throughout the importance of the UN Guiding Principles on Business and Human Rights, adopted by the Human Rights Council, as providing baseline approaches that all Internet companies should adopt. I especially encouraged transparency in all aspects of their operations.

The report before you (A/73/348) continues these themes in the context of artificial intelligence and human rights. It does not pretend to be the last word in AI and human rights. Rather, it tries to do three things: define key terms essential to a human rights discussion about AI; identify the human rights legal framework relevant to AI; and present some preliminary recommendations to ensure that, as the technologies comprising AI evolve, human rights considerations are baked into that process.

I fully understand that many might look at the report, see AI, and think – math, it’s hard. And it is true – the technologies behind AI can be complicated and opaque to people who are not technologists or mathematicians. But it is in part for that reason that AI technologies – whether adopted by public or private actors – be as transparent as possible so that, as they take greater roles in our information environment, we do not lose control of fundamental rights to the machines and the humans who program them.

Allow me to summarize briefly some of the main points of the report.

To begin with, I would urge all those with an interest to begin by reading Section II, which explains the technologies that comprise AI, the key features that set them apart from other ways of organizing information, and the way in which they are used in the information environment. In this connection, the report notes that algorithms and AI applications are found in every corner of the internet, on digital devices and in technical systems, in search engines, social media platforms, messaging applications, and public information mechanisms. I review in particular how they implicate content display and personalization, content moderation and removal, and profiling, advertising and targeting.

The report then turns to the human rights at issue. I note a range of rights but want to emphasize the following here. One, the freedom of opinion. The intersection of technology and content curation raises novel questions about the types of coercion or inducement that may be considered an interference with the right to form an opinion. Content curation has long informed the individual’s capacity to form opinions: for example, media outlets elevate particular stories to the front page with the intention to shape and influence individual knowledge about significant news of the day. Commercial advertising has also sought to induce favourable opinions of and cultivate desire for particular products and services.

Two, the freedom of expression. Among other things, users lack access to the rules of the game when it comes to AI-driven platforms and websites. A lack of clarity about the extent and scope of AI and algorithmic applications online prevent individuals from understanding when and according to what metric information is disseminated, restricted or targeted. Small concessions to addressing this problem such as selective identification of sponsored search results, or social media platforms highlighting when advertising is paid for by political actors, may slightly contribute to helping users understand the rules of the information environment, but these neither capture nor resolve the concerns around the scale at which algorithmic processes are shaping that environment.

The report also addresses AI and the right to privacy, the pervasive obligations of non-discrimination, and the right to effective remedy.

I urge a human rights-based approach to AI. Among other things, I urge the following when it comes to substantive standards:

* Companies should orient their standards, rules and system-design around universal human rights principles. Public-facing terms and guidelines should be complemented by internal policy commitments to mainstreaming human rights considerations throughout a company’s operations, particularly in relation to the development and deployment of AI and algorithmic systems. Companies should consider how to elaborate professional standards for AI engineers, translating human rights responsibilities into guidance for technical design and operation choices. The development of codes of ethics and accompanying institutional structures may be an important complement to, but not a substitute for, commitments to human rights. Codes and guidelines issued by both public and private sector bodies should emphasise that human rights law provides the fundamental rules for the protection of individuals in the context of AI, while ethics frameworks may assist in further developing the content and application of human rights in specific circumstances.
* Radical transparency about an AI technology’s impact in the information environment requires disclosure of, for example, data on how much content is removed by AI systems, how often AI-suggested content removals are approved by a human moderator, how often content removals are contested, and how often challenges to content removals are upheld. Aggregate data illustrating trends in content display should be available for users to inspect, alongside case studies which illustrate why certain content will be prioritised over other content. Disclosure about the sources and beneficiaries of political and commercial advertising is a critical element of radical transparency. Public and private sector actors implementing AI-driven systems should also be transparent about the limits of the AI system, including for example, confidence measures, known failure scenarios, and appropriate limitations on use.
* Tackling the prevalence of discrimination in AI systems is an existential challenge for companies and governments; a failure to address and resolve the discriminatory elements and impacts will render the technology not only ineffective but dangerous. There is ample thought leadership and resources for companies and governments to draw on in thinking about how to address bias and discrimination in AI systems; broadly speaking, it necessitates isolating and accounting for discrimination at both the input and output levels. This involves, at a minimum, addressing sampling errors (where datasets are non-representative of society), scrubbing datasets to remove discriminatory data, and putting place measures to compensate for data which “contains the imprint of historical and structural patterns of discrimination” and from which AI systems are likely to develop discriminatory proxies. Active monitoring of discriminatory outcomes of AI systems is also integral to avoiding and mitigating adverse effects on individuals’ human rights.

I also urge the following when it comes to process:

* Embracing radical transparency throughout the AI lifecycle requires companies and governments to take steps to permit systems to be scrutinised and challenged *from conception to implementation*. Human Rights Impact Assessments are one tool for demonstrating a commitment to addressing the human rights implications of AI systems, and should be performed pre-procurement, development or use, and involve both self-assessment and external review.
* AI must not invisibly supplant, manipulate or interfere with an individual’s ability to form and hold their opinions or access and express ideas in the information environment. Respecting individual autonomy means, at the very least, ensuring users have knowledge, choice and control. Pervasive and hidden AI applications which obscure the processes of content display, personalisation, moderation and profiling and targeting undermine the individual’s ability to exercise the rights to freedom of opinion, expression and privacy. Companies should be mindful of the adverse human rights impacts that flow from AI applications that prioritise commercial or political interests over transparency and individual choice.
* Companies must ensure that users are fully informed about how algorithmic decision-making shapes their use of a platform, site or service. This can be achieved through education campaigns, pop-up boxes, interstitials and other means of signalling when an AI system is determining a user’s experience of a search engine, news site or social media platform. State-imposed disclosure requirements may be an appropriate means to protect notice and consent. Individuals also have a right to know when their data is being collected by an AI application and whether it will become part of a dataset that will subsequently inform an AI application, as well as the conditions on which that data will be used, stored and deleted.
* Adverse impacts of AI systems on human rights must be remediable and remedied by the companies responsible. The pre-condition to the establishment of effective remedy processes is ensuring individuals know they have been subject to an algorithmic decision (including one that is suggested by an AI system and approved by a human interlocutor) and are equipped with information about the logic behind that decision. Beyond that, companies should ensure human review of requests for remedy, in order to provide an appropriate check on the systems and guarantee accountability. Data should be published on the frequency at which remedial mechanisms are triggered for decisions made by AI technologies.

**Mr. President,**

Over the coming year I expect to further explore the issues involved in technology’s dominance of the tools of communication. I will be reporting on the issue of commercial spyware, their export by developed countries and their use in abusive ways by governments around the world. I will also be exploring the way in which social medial companies involve local civil society and other actors in those markets where they dominate public space.

But even beyond these thematic issues, I will continue to monitor closely those issues that I am bound to consider under the Human Rights Council resolutions that established my mandate. In particular I will continue to:

* *gather all relevant information, wherever it may occur, of discrimination against, threats or use of violence and harassment, including persecution and intimidation, directed at persons seeking to exercise or to promote the exercise of the right to freedom of opinion and expression…*

and

* *gather all relevant information, wherever it may occur, of discrimination against, threats or use of violence and harassment, including persecution and intimidation, against professionals in the field of information seeking to exercise or to promote the exercise of the right to freedom of opinion and expression…*

I look forward to working with your governments to implement the norms of the right to freedom of expression. Thank you very much.