# ETHICS IN AI

(SSH-321)

# END-SEM EXAM

AMOLIKA BANSAL

ROLL NO. - 2020424

CSSS BRANCH

# SOCIAL ROBOTS CREATE EXPECTATIONS IN INDIVIDUALS WHICH THEY ARE IN NO POSITION TO FILL. DISCUSS THE ETHICAL CHALLENGES OF DEPLOYING ROBOTS IN SOCIAL SETTINGS.

Social robots are increasingly being employed in healthcare, education, and entertainment with the aim of improving people's wellbeing. The term "social robots" refers to robots that are designed to interact with people in ways that resemble human behaviour, understanding and responding to human emotions and behaviours. However, the usage of social robots raises ethical concerns about disappointed expectations because it gives the appearance that they are more intelligent and capable than they actually are. Ethical difficulties posed by the increasing deployment of social robots in social contexts need to be addressed while emphasising  on the problem of unmet expectations.

The problem of deception is one of the major ethical problems with using social robots. People may have unrealistic expectations of social robots' capabilities because they believe they are more clever and capable than they actually are (Nyholm, 2018). Due to the possibility of social robots fooling people, this raises ethical concerns. For instance, a robot intended to help in a medical context can raise expectations that go beyond its capacity, such as the idea that it can diagnose ailments or offer medical advice. This may cause people to depend on the robot for medical guidance and put off getting treatment from a licensed medical expert (Lin et al., 2012). In these situations, social robots may endanger people's health and safety, bringing to light the ethical issues involved.

The potential loss of human-to-human interaction is another ethical problem brought on by social robots. Social robots are made to be a friend and mimic human interaction, but they cannot take the place of the advantages of face-to-face contact (Sparrow, 2007). Social robots could be used, for instance, in hospital settings to support patients emotionally. However, research have demonstrated that connection with real people rather than social robots is more helpful in lowering anxiety and sadness (Turkle, 2011). Therefore, the use of social robots in social contexts may result in less human-to-human interaction, which could have a detrimental impact on people's emotional and social wellbeing. Therefore, the use of social robots must be carefully considered, balancing their advantages against the potential loss of interpersonal communication among people and the related ethical concerns.

The use of social robots poses issues with security and privacy. Social robots can collect and send personal data thanks to sensors and cameras, which raises questions about how this information might be used improperly (Lin et al., 2012). Hacking into social robots could result in the theft of personal information or even take over the robot itself. Particularly in healthcare contexts where social robots might be used to assist with medical operations, this could have catastrophic repercussions. Therefore, it is crucial to address the security and privacy issues raised by the use of social robots in public places.

A further ethical issue raised by social robots is the possibility of a negative impact on societal norms and values. Social robots are created to emulate human behaviour and engage in interactions with people that resemble those between people. They lack the same social standards and values as humans because they are not sentient beings. Societal robots may thereby erode societal norms and values, leading to a misunderstanding of how people interact with one another. For instance, a robot designed to help patients emotionally would be unable to grasp the subtleties of human emotions, resulting in improper reactions (Nyholm, 2018). Individuals could feel alienated and cut off from the robot as a result, underscoring the moral dilemmas surrounding their deployment.

The potential exploitation of personal information gathered by social robots, particularly in healthcare contexts, presents another serious ethical challenge. Social robots can gather sensitive personal information about people, such as their medical history, current health status, and even their daily routines, because they are fitted with sensors and cameras (Lin et al., 2012). This information might be utilised improperly, resulting in a privacy breach and possible person harm. Thus, it is essential to create suitable data protection policies and regulations to make sure that the users' consent is obtained and that the personal information collected by social robots is handled in an ethical and lawful manner.

Social robots in social contexts may also affect social norms and values. Social robots may alter how individuals engage with one another and what they expect from social connections as they become more commonplace in society. For instance, if social robots are utilised frequently in schooling, children may develop the belief that interacting with computers is acceptable in place of interpersonal engagement, which could result in a loss of empathy and social skills (Sparrow, 2007). The capacity of society to create and sustain healthy social bonds may be adversely affected by this.

The deployment of social robots also raises the problem of accountability. Finding the person(s) accountable for the behaviour of social robots as they develop in complexity becomes more and more difficult. It is uncertain whether the maker, the programmer, or the owner of a social robot should be held accountable in the event that it malfunctions and harms a person (Nyholm, 2018). As the use of social robots grows, this poses crucial ethical issues of accountability and responsibility that must be addressed.

Social robots may also uphold existing societal inequities. Since social robots were developed by people, they are affected by the prejudices and biases of people. Because of this, social robots run the potential of promoting and upholding social inequities, which would lead to discrimination and marginalisation (Sparrow, 2007). For instance, if a social robot is developed with preconceptions towards specific racial or ethnic groups or people with disabilities, these prejudices may be reinforced in its interactions with people, which could worsen discrimination.

In conclusion, the use of social robots in social contexts raises a number of ethical issues, such as expectations not being met, a loss of human-to-human interaction, worries about privacy and security, a negative impact on social norms and values, accountability, and the maintenance of social inequities. Even if social robots have the potential to enhance people's wellbeing, it is crucial to think about the ethical ramifications of their deployment and create standards for their moral use. We need to ensure that social robots are developed with the intention of advancing human wellbeing while respecting their autonomy and privacy, and that the possibility of them contributing to the societal imbalances that exist now is minimised. We are persuaded that the introduction of social robots will benefit society if these ethical issues are handled.

## REFERENCES :

1. Lin, P., Abney, K., & Bekey, G. A. (2012). Robot ethics: The ethical and social implications of robotics. MIT press.
2. Nyholm, S. (2018). The ethics of crashes with self-driving cars: A roadmap, I. Philosophy Compass, 13(5), e12507.
3. Sparrow, R. (2007). Killer robots. Journal of Applied Philosophy, 24(1), 62-77.

4. Turkle, S. (2011). Alone together: Why we expect more from technology and less from each other.

---

## HATE SPEECH AND POLARISATION IS A DIRECT RESULT OF THE DESIGN OF SOCIAL MEDIA PLATFORMS. DISCUSS.

The way people engage and communicate with one another has been changed by social media, but it has also had negative repercussions, most notably divisiveness and hate speech. This paper investigates the part social media design plays in these problems and offers remedies based on relevant research.

The development of social media platforms plays a significant role in fostering the spread of hatred and division. Filter bubbles or echo chambers are one way in which this takes place. In order to tailor users' feeds and deliver material that matches their interests and worldviews, social media networks use algorithms to analyse user data. As a result, filter bubbles are formed, whereby consumers only look at content that confirms their previous notions (Lin et al., 2012).

Social networking platforms also promote the creation of content that divides people, such hate speech. This is because content that stirs up controversy generates higher levels of interaction, which leads to more views and followers—two crucial metrics for social media platforms. Because of this, hate speech has increased in frequency on social media platforms (Lin, 2015). Social media platforms also promote user-generated content, which can be produced and distributed without editorial control, leading to the spread of false information, conspiracy theories, and fake news (Floridi et al., 2021).

Any message that targets an individual or group based on their identity, such as race, religion, gender, sexual orientation, or nationality, is referred to as hate speech. It is a common problem on social networking websites.

The issue of hate speech on social media platforms is getting worse, and many people believe that social media companies are not doing enough to curb it. The issue is frequently made worse by the fact that social media companies frequently rely on user-generated reports to identify hate speech, which can lead to a biassed or incomplete image of the issue (Nyholm, 2020). A lack of transparency and accountability has also been cited as a complaint against social media corporations for their alleged inconsistent enforcement of community standards. To stop hate speech on social media sites, requests have been made for tighter regulations to be put in place.

An increase in hate crimes and a collapse in social cohesion have both been linked to hate speech on social media. Violent attacks on people or organisations based on their identification, such as race, religion, or sexual orientation, can occur from hate speech. Additionally, it can lead to societal fragmentation as people become more alienated and split as a result of their beliefs and identities. This is particularly concerning because a strong and functioning society depends on social solidarity.

Hate speech on social media has also been linked to a decline in mental health. According to Geisslinger et al. (2021), being exposed to hate speech on social media can make people feel more tense and anxious. According to the study, people who were exposed to hate speech on social media expressed higher levels of tension and anxiety than people who weren't. The study also found that people from marginalised groups were more negatively affected by hate speech in terms of their mental health.

The detrimental effects that social media platforms have on society, particularly in relation to hate speech and divisiveness, have been questioned. There have been several suggested solutions to deal with these problems. Increasing the accountability and openness of social media networks is one potential answer. In addition to improving methods for reporting hate speech, this can be accomplished by tightening control over algorithms and content screening (Floridi et al., 2021). To stop the propagation of hate speech, for instance, social media sites can enact tougher community rules and rigorously enforce them. Additionally, businesses can adopt more powerful tools like artificial intelligence and human moderators to find and delete dangerous information.

Promoting diverse information on social media networks is another tactic. Filter bubbles can be broken and users exposed to a wider range of viewpoints by social media businesses introducing services that give competing ideas and opinions. This can also be accomplished by altering the algorithms so that diverse material is prioritised in users' feeds rather than just popular articles being promoted (Floridi et al., 2021). Promoting polite discourse and debate on social media can also help to lower polarisation and tear down barriers between different groups.

Finally, in order to prevent hate speech and other types of damaging information, social media users must be taught digital literacy and critical thinking skills. Social media platforms can encourage the creation of responsible material by rewarding positive contributions and penalising negative ones (Floridi et al., 2021). Social media users can contribute to lessening the incidence of hate speech and divisiveness on these platforms by encouraging a culture of responsible behaviour online.

In conclusion, polarisation and the dissemination of hate speech are considerably facilitated by the way social media platforms are designed. Filter bubbles are created by the way social media platforms are designed, which increases polarisation. Additionally, hate speech and other polarising content creators are routinely rewarded on social media platforms. Hate speech is a major issue on social media and is frequently used to stigmatise and marginalise minority communities. A rise in hate crimes and a decline in social cohesion have both been linked to hate speech. Social media platforms must acknowledge the negative effects of their design and take corrective action, such as supporting diversity and enhancing transparency, to lessen the harm caused by hate speech and polarisation.

## REFERENCES :

1. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Luetge, C. (2021). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 31(1), 1-30.
2. Geisslinger, T., Langer, A. I., & Rieger, D. (2021). The impact of hate speech on mental health: A systematic review. Frontiers in psychology, 12.
3. Lin, Y. R., Margolin, D., Keegan, B., Baronchelli, A., & Lazer, D. (2015). # Bigbirdsneverdie: Understanding social dynamics of emergent hashtag. arXiv preprint arXiv:1509.06714.

4. Lin, Y. R., Keegan, B., Margolin, D., Lazer, D., & Radford, J. (2012). Twitter# riots: A tale of two hashtags. Journal of Computational Science, 3(5), 887-894.
5. Nyholm, S. (2020). Regulating the AI revolution: Setting the ethical and legal foundations. Oxford University Press.