

**School of Computer Science and Statistics**

**Individual Assessment Submission Form**

|  |  |
| --- | --- |
| **Course Title** | TEP |
| **Module Title** | TEP 00062 |
| **Lecturer(s)** | Prof Dave Lewis, Prof Vincent Wade |
| **Assessment Title** | Individual Assignment on Ethical Digital Engagement |
| **Date Submitted** | 27/03/2024 |
| **Word Count** | 3930 |
| **Student Number** | 22314780 |
| **Student Name** | Alina Lavrova |

**Assignment Submission Declaration:**

I have read and understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at: <http://www.tcd.ie/calendar>

I have also completed the Online Tutorial on avoiding plagiarism ‘Ready, Steady, Write’, located at <http://tcd-ie.libguides.com/plagiarism/ready-steady-write>

I declare that this assignment, together with any supporting artefact is offered for assessment as our original and unaided work, except in so far as any advice and/or assistance from any other named person in preparing it and any reference material used are duly and appropriately acknowledged.

Any use of Chatbots/Generative AI tools in researching the materials research for this report is fully described at the end of the report. I confirm that no text produced by such tools has been directly used in the report.

Signed ………………Alina Lavrova……………… Date …………27/03/2024…………

Description of use of Chatbots/Generative AI tools:

………………………………………………………………

Investigating and Analysing Ethical Risks of a Digital Engagement Application

Description of Application

Snapchat is an American messaging app that allows users to exchange photos and videos. This application originally was developed by Snap Inc., which is a technological company founded in 2011 by Evan Spiegel, Bobby Murphy, and Reggie Brown. At a time when the app was launched, it provided a new way of how people communicate online because whereas most messaging apps focused on text messages that users share with each other, Snapchat shifted its attention to photo and video communication. Users of the app can send and receive photos and videos for a particular amount of time which usually does not exceed 10 to 15 seconds *(Velten et al., 2017)*. Whenever the timer is up, the content is permanently deleted from the chat. What is prominent is that each user is able to set the timer for a message right before it is sent and even receive a notification if the screenshot is made. Because of this feature, Snapchat was able to establish itself in the digital space during a time when privacy concerns were growing. It helped to reduce the risk involved in gathering and storing personal data online that can be then exploited by third parties against users.

Snapchat is mostly popular among children and teenagers who tend to quickly share photos and videos in large amounts and forget about what was sent right after. Moreover, the timer became a valuable privacy feature for those young people who want to keep their chats private from parents, friends or teachers. Older users too, in fact, respect the application's safety features and it is one of the most popular messengers in the world today. Furthermore, those who wish to preserve specific chat messages are able to do so as well so that not all photos, videos or text messages have to disappear for ever.

In addition to chats with friends, Snapchat offers a lot more features such as an opportunity to post ‘My Story’, play mini-games, watch entertaining videos and news, and use different photo and video editing tools and effects, and many more *(Elgersma, 2018)*. Moreover, for many years an extremely popular element among users has been the ability to apply ‘Face Masks’ while taking photos or videos. These masks transform users’ faces and people can choose the one they like among hundreds of available. However, recently Snapchat also added an AI Chatbot called ‘My AI’ to its application opening a new set of opportunities for its users. As Snap Inc. describes it: *“In a chat conversation, My AI can answer a burning trivia question, offer advice on the perfect gift for your BFF’s birthday, help plan a hiking trip for a long weekend, or suggest what to make for dinner.” (Snapchat.com)*. In general, this chatbot is a simplified version of ChatGPT integrated into the app and listed as one of the user’s friends in the chat list. The chatbot may be given a name, be added to groups, and have its avatar's look changed by users. Moreover, a person can share images in the chat as well. The bot replies to messages rather than contacting users first. Because My AI captures users’ data such as location and basic information like the device used, it offers personalized solutions and enhances engagement. This chatbot attracts people not by only answering questions and offering features that other AI chatbots do but also by recommending places to visit near the user specifically chosen according to the enquire, providing directions to a place and identifying animals, plants, objects through photos that can be sent into the chat. As a result, Snapchat becomes a multipurpose application and increases its capabilities leading users to spend more time within it and retain more frequently.

Identification of Stakeholder Roles Involved in the Application and its Governance

AI Provider

The chatbot integrated into Snapchat is powered by OpenAI and is based on the latest version of ChatGPT. However, Snap Inc. modified the initial version and trained AI to make it more suitable for the Snapchat application. For example, it uses AR and visual processing in addition to being a chatbot and playing games with users. Although, the main difference between ChatGPT and My AI is that it has more restrictions and safety measures integrated. It can be seen that My AI is trained specifically in a way that aligns with the company’s safety and privacy guidelines. Any swearing, violence, sexually explicit content, or opinions about dicey topics like politics are restricted in My AI, it also politely declines to help with academic essays *(Heath, 2023)*.

Direct Data Providers

The primary categories of people using Snapchat are people from the age groups of 18-24 and 13-17 *(Statista, 2023)*. Those are children, teenagers and young adults who together constitute the target audience of the application. People from these groups most value the ability to preserve messages, photos and videos only for a limited time. At the same time, as the main users, these groups are the primary providers of data utilised by the application and AI integrated within it. Gathered information helps to increase performance, whereas gained statistics allow to learn more about clients’ needs and enhance the quality of time spent by users in the application by offering a personalised experience.

Indirect Data Providers

However, direct users of Snapchat are not the only category of people whose data can be captured by the provider. It can be seen that whenever children or teenagers use the application to record live moments, take photos or videos with funny face masks, post them in ‘My Story’ or just save in the app memory, people who are also captured on them on purpose or accidentally, are automatically affected. For example, the personal information of friends and family members of users can appear in the application with people even not being acknowledged. All information can be saved in the chat with My AI which, in turn, saves users’ data even after the content stops being displayed in the chat. Another possible group of stakeholders who can be influenced by the use of Snapchat is teachers at schools. It is common for teenagers to make jokes about classmates and school staff. For this reason, teachers’ photos can be shared and stored in the app, as well as in the AI chatbot.

It is apparent that even when Snap Inc. cannot fully control data being sent to the application, it is responsible for ensuring its protection. Snapchat has a policy of being transparent in the information it captures and how it is used but it still shares some data with third-party applications and its partners. The main strategy of Snapchat to increase data protection is to store the information for as little time as possible. Initially, it was a competitive advantage of the application that differentiated it from other messengers. Moreover, Snap Inc. accepts its social responsibility and understands that its competitive advantage and market share are strongly connected to the extent to which users feel that they can trust an application. Whenever the data is not protected enough, it becomes a potential target for hackers and other people with illegitimate motives, information leaks put users in danger and break the trust.

Oversight Stakeholders

Whereas users play an important role in Snapchat’s decisions, the application’s AI is also a subject of scrutiny by regulatory bodies around the world. In addition to the detailed privacy policy and Snapchat’s guidelines on maintaining security within AI and the application itself, the application follows data protection regulations which are specific for different regions where it operates. For example, in Europe, Snapchat is accountable to the European Data Protection Board and has to remain in compliance with the Digital Services Act (DSA), in Canada this role is played by AI and Data Act (AIDA) and supervised by the Ministry of Innovation, Science and Economic Development and Office of the Privacy Commissioner of Canada (*Global AI Regulations Tracker: Europe, Americas & Asia-Pacific Overview*, n.d.). Nowadays, AI regulation is still a new area where different ways to protect people are still developing. Whereas some countries are already working on the adoption of security measures, others only starting to plan possible legislation.

Identification of Ethical Risks

According to statistics from Statista.com, Snapchat was able to achieve more than 414 million daily users worldwide in the third quarter of 2023 *(Statista, 2024)*. These incredible results point out that an application created by Snap Inc. is one of the most popular in the world. As a result, once the programme is used so widely, further concerns about the morality of the chatbot and messaging AI arise, which may undermine the AI Social Responsibility Principles. When some of the risks associated with the Social Responsibility of AI are only potential and could affect stakeholders only in theory, others are more likely to occur and people need to keep them in mind while using My AI.

Human Rights

First, and most dangerous of the consequences of AI is the violation of human rights. Inaccurate and biased information from My AI can have an impact on the dignity, social, cultural and even economic rights of a person. Even with additional training from Snap Inc. the chatbot, as it was trained on a large amount of data from various sources, can be biased, provide discriminative information, or affect the well-being of stakeholders. Specifically, in My AI the used LLMS is reduced in order to diminish the likelihood of such responses. However, it still can occur as AI is far from perfect and the technology is only in the evolving stage. Such offensive data can influence any users of My AI but the younger audience is more sensitive and likely to believe in the content provided. Older users, at the same time, can keep in mind those risks and critically evaluate gained information. Moreover, the danger of AI in this context is also based on the fact that My AI appears as a friend in the user’s chat list and is created in a manner that gullible teenagers can forget that it is not their friend but only a technology when all conversations are designed in a very human-like manner. This may have an impact on the well-being and mental integrity of a user when parents of the children affected could not even know that their kid treats AI as a real friend. Moreover, the implications of this risk are high, imperfect information can affect people’s perception of the world and can lead to unpredictable outcomes. It can be seen that the situations with parents worrying about their children after Snapchat launched My AI have been real from the first day *(Kelly, 2023)*. However, in my opinion, the likelihood of violation of human rights on a such high level is under question because usually the conversations with chatbots are short and simple, and the scale of influencing information is more likely to be minimised.

Consumer Issues

Second, My AI can play a role in consumer issues when personal data of users gained during their time in the app, is misused by developers or exploited by third-party. My AI tracks users’ locations, saves data provided with it and utilises it in the future to enhance performance and user experience. However, location and other personal data are private information that in case of exploitation, can lead to issues with stakeholder security. Whereas the risk of leaking the user's location at one point in time could look harmless, information gained by Snapchat together with proper use can tell a lot about the person, and where they live. For example, My AI tracks location in a background mode, it stores users’ responses about preferable places and hobbies, and it can even have a photo of the user if this person sent his face to the chat. Similarly, this risk increases when the users are a younger generation as they have a lower level of privacy concerns. Despite Snapchat employing various techniques to secure confidential data and prevent leaks, any user or stakeholder who uses the app is at certain risk of having information about them compromised if the leakage happens. This also includes friends, family, coworkers, teachers, and other contacts of those whose information is misused. Not to add that the danger is really serious when it occurs. Although there have not been any occurrences in the past few years lowering the likelihood of the risk, Snapchat has a history of several data breach incidents. For instance, in 2017 more than 200,000 photos of users leaked because the partner photo storage was hacked. Another prominent example of unauthorised usage of user data is when in May 2019, it occurred that Snapchat employees were spying on users and had access to their personal information and contact details *(Reed, 2021)*. As a result, this is a real issue with dangerous consequences and it has to be carefully guided.

Labour practices

It can be seen that nowadays AI technology influences how work is done and its role becomes more important every year. Different AI-based improvements help workers to simplify some tasks, save time and become more productive. These technologies can be used across all parts of the business in different departments but there is also a risk associated with companies’ transition to use more of artificial intelligence. One of the potential threats of AI integrated into Snapchat could be a decrease in employee demand because of labour displacement. In spite of the fact that My AI does not directly affect employees at first glance as it is an additional feature rather than a substitute for existing operations, workers can be affected by the application of AI in internal business structure and workflow. For example, automatisation of some functions within Snap Inc. can diminish the need for low-skilled workers or the number of employees needed in general can be reduced. In this case, if a company does not spend some resources on retraining or restructuring the workforce, a part of the workers can appear to be fired which negatively affects their lives and the lives of their families. Another case can be when AI technologies used for recruiting new employees, discriminate against specific groups of people *(Discrimination and Bias in AI Recruitment: A Case Study, n.d.)*. Nevertheless, the impact of AI on employees in Snap Inc. is hard to determine and the danger of My AI for employees is more likely to be low as well as the likelihood of its occurrence. The reason for this belief is that there were no incidents in Snap Inc. regarding job losses or discrimination caused by AI but we still need to carefully research and control this issue because it has severe consequences in case of occurrence.

The Environment

One of the major reasons why artificial intelligence can affect the environment is its water usage. In general, significant water consumption is needed both directly and indirectly for AI systems to operate. The water footprint, the amount of water used in operation, has a significant impact on the environment. *“Direct water use includes the water used in cooling systems of data centres and the production of microchips used in AI models. Indirect water use includes the water used in the production of electricity used to power the data centres that run AI models”* *(View of the Environmental Impact of AI: A Case Study of Water Consumption by Chat GPT, n.d.)*. Despite that water usage in the AI industry is relatively smaller than in other industries, it still has an impact on the world, especially in places where water is hard to find. For Snapchat specifically, it can be estimated that the water footprint is not inferior in volume to other AI systems because, as mentioned above, Snap Inc. provides its chatbot based on ChatGPT from OpenAI. As a result, we can approximate the environmental impact of My AI as an impact that ChatGPT has on the planet complemented by resources needed for additional training and operation. Therefore, even when the impact on the environment is lower than, for instance, in agriculture, it is still severe and realistic and the actions against it have to be made.

Fair Operating Procedures

In general, algorithms drive My AI's operation, dictating its reactions and behaviours. Therefore, ethical risks may occur if the procedures used are not transparent to stakeholders. Lack of transparency can lead to biased responses and practices, as well as unfair treatment of Snapchat users. When AI algorithms are biased, My AI responses can be used to marginalise certain social groups or, for example, manipulate the public political process in a hidden way, using a friendly conversation manner. This can directly affect users’ point of view, which can be later translated to their friends, family and people whom they interact with. However, I believe that the risk of this happening in the case of My AI particularly is insignificant and the likelihood is low. The reason for such a claim is that Snapchat’s AI is designed in a way that limits any discussions concerning political matters or other sensitive topics. When the user writes questions including these themes, My AI refuses to answer and to continue the conversation about it.

Community Involvement and Development

Whereas the chatbot created by Snapchat, as mentioned, can create ethical risks for its users and people close to them by undermining personal data privacy, causing threats to well-being, providing misleading information and so on, its practices can also affect whole different communities in some cases. Some of the unique features of My AI that are designed to be its main competitive advantage, can also bring damage to society more broadly. For instance, Snapchat’s chatbot is popular for its ability to suggest users places to visit and recommend different activities. The user just needs to put their enquiry in the form of a message to the chat and My AI will provide recommendations and even show them on the map with the designed route. The application itself collaborates with external platforms that provide data on different places to the chatbot. However, there is a potential problem deriving from such recommendations. It is possible that some small local businesses, cafes, tiny restaurants and other not-very-well-known places can be affected by this technology. Suggesting users only places that have a lot of feedback and information online, will prevent those people from visiting local places, the chance of accidentally finding a great location diminishes when the route is automatically provided to you and places are chosen without your active participation. Though the chances of such events are not excessive because local places can still be offered, and the effects could not be seen to be catastrophic for society as a whole because not everyone uses this feature, some individual business owners can still experience damage.

Discussion of Mitigation Measures for Risk

Personally, I believe that the influence of My AI chatbot on children constitutes the highest ethical risk of all listed above because includes the most vulnerable part of society. In situations when the younger generation using technology is not able to separate it from real friends, it puts human rights in danger. My AI is created in a way that users can give this chatbot a name, change its avatar and customise the chat. Moreover, it is displayed as just one of the dialogues in a chat list next to the conversations with the user’s friends. For all these reasons that make My AI unique in the first place, together with simple friend-like immediate responses, the risk that some people would treat it more like another human being rises exponentially. As a result, the well-being of those who are attached to artificial intelligence can suffer. For instance, children can ask for a piece of personal advice or mental help but the chatbot does not have any empathy as well as specifically trained knowledge. Not only this but children’s mental integrity and health can be at risk if other types of information provided by My AI contain inappropriate or ethically unacceptable responses. Therefore it is important to define possible measures that the developer and the broader community can use to decrease this risk.

The Developer

First, Snapchat can undertake several actions itself. It is crucial to ensure that Snap Inc. acknowledges this issue and acts towards its mitigation. Snapchat can re-design My AI with the purpose of making it look less like real users. The customisation features available to children can be reduced or more signs pointing out that it is just a technology can be added. Moreover, guidelines with important information regarding the safe use of the feature can be shown before every conversation or reminders in the format of push-up notifications can be introduced. It should be clearly communicated to children that My AI is an artificial intelligence rather than a real person. Not only this, but Snapchat can establish an additional age boundary for the My AI feature that not all users (who can be any person 13 and more years old) can interact with it but the ones who are older that certain safer age.

External Stakeholders

Second, the broader society and people close to potential victims can raise awareness. In the era of AI children’s education on the safe use of the technology and ways to protect themselves should become a priority. Schools should create lessons concerning it, parents should communicate the risks to their children and control the use of the technology if the child is too young, and entertainment platforms should provide more material about it. All of those initiatives must have as their primary goal informing young people about the dangers associated with artificial intelligence and raising their level of attention to their safety, personal information, and general well-being. Not only this but together with the actions taken directly by technology providers, it would decrease the risk of young people’s human rights becoming affected.

Legislation Bodies

Lastly, regulators should not stay away from this issue. Some of the additional legislation measures that can be implemented to reduce the ethical risk of children being affected can include specific certifications and licences that companies, whose target audience include young people, should receive before the technology goes public. Those certifications can include additional regulations on how AI has to be designed and the scope of answers it can provide. Moreover, legislation bodies can collaborate with technological organisations, educational institutions and organisations protecting child rights to ensure better ethical control and address challenges concerning children’s privacy. Together, those parties can develop safer methods and design innovative ways to deal with the issues. Moreover, legislators can create educational forums that would help schools, parents and other organisations to better educate children and design awareness programs.

Conclusion

To conclude, My AI powered by Snapchat is a powerful tool enabling users to enhance their experience and gain information, tips and advice on various topics. It can help people to find interesting places, get ideas on recipes, what to buy as a present and many more. However, AI technology is currently in a developing stage and there are still a lot of risks and challenges that society has to address. My AI when increasing the time spent by users on Snapchat, can cause harm to users and other stakeholders. One of the main threats is concerns with young people’s human rights who are the main target audience of the application. Therefore, Snap Inc., external stakeholders and regulators have to carefully manage the technology and continuously develop ways to reduce any potential dangers arising from its use.

References:

Snapchat Support Centre

<https://help.snapchat.com/hc/en-us/articles/13266788358932-What-is-My-AI-on-Snapchat-and-how-do-I-use-it>

Velten, J. C., Arif, R., & Moehring, D. (2017). Managing Disclosure through Social Media: How Snapchat is Shaking Boundaries of Privacy Perceptions. *The Journal of Social Media in Society*, *6*(1), 220–250. <https://thejsms.org/index.php/JSMS/article/view/214/116>

Elgersma, C. (2018, June 18). *Everything you need to know about Snapchat*. Phys.org; Phys.org. <https://phys.org/news/2018-06-snapchat.html>

*Artificial intelligence in social media / What do parents need to know about Snapchat’s new chatbot My AI?* (2023, September 1). Klicksafe.de. <https://www.klicksafe.de/en/news/was-muessen-eltern-zum-neuen-chatbot-my-ai-bei-snapchat-wissen>

Statista. (2024). *Snapchat daily active users 2024 | Statista*. Statista; Statista. <https://www.statista.com/statistics/545967/snapchat-app-dau/>

Statista. (2023). *Average age of Snapchat users 2023 | Statista*. Statista; Statista. <https://www.statista.com/statistics/933948/snapchat-global-user-age-distribution/>

*Global AI Regulations Tracker: Europe, Americas & Asia-Pacific Overview*. (n.d.). Legalnodes.com. <https://legalnodes.com/article/global-ai-regulations-tracker>

‌Reed, C. (2021, December 8). *Snapchat Data Breaches: Full Timeline Through 2021*. Firewall Times. <https://firewalltimes.com/snapchat-data-breach-timeline/>

*View of The Environmental Impact of AI: A Case Study of Water Consumption by Chat GPT*. (n.d.). Puiij.com. <https://puiij.com/index.php/research/article/view/39/23>

Kelly, S. M. (2023, April 27). *Snapchat’s new AI chatbot is already raising alarms among teens and parents | CNN Business*. CNN. <https://edition.cnn.com/2023/04/27/tech/snapchat-my-ai-concerns-wellness/index.html>

Heath, A. (2023, February 27). *Snapchat is releasing its own AI chatbot powered by ChatGPT*. The Verge. <https://www.theverge.com/2023/2/27/23614959/snapchat-my-ai-chatbot-chatgpt-openai-plus-subscription>

*Discrimination and bias in AI recruitment: a case study*. (n.d.). Lewis Silkin. <https://www.lewissilkin.com/en/insights/discrimination-and-bias-in-ai-recruitment-a-case-study#:~:text=An%20algorithm%20can%20be%20tainted>