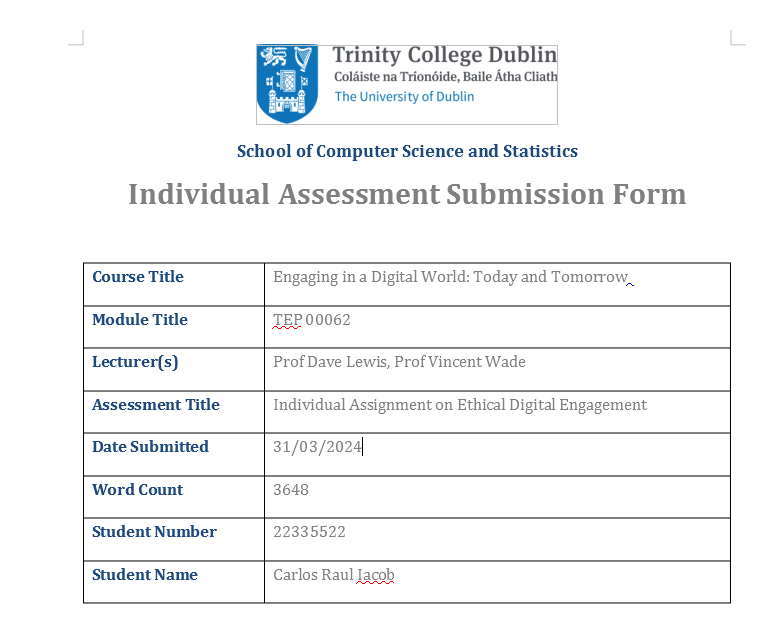
****

**Introduction and description of the application:**

The AI model and techniques that arise from it which impact stakeholders and hold several ethical concerns upon which I will be presenting is the use of relationship AI chatbots. I will be analysing the use of relationship chatbots both in the present and its potential in the future, The different stakeholders who can be potentially affected by the rise of such chatbots, the ethical risks that are presented with the rise of such use and ways to mitigate the risks that could arise with the use of such chatbots. The utilization of chatbots to have a deeper personal relationship with the product, in this case, the chatbot itself is a digital engagement technique upon which I will be analyzing both the uses of it while also looking at the stakeholders and ethical risks and remedies present.

Relationship AI Chatbots are programs that have been developed to have romantic and platonic relationships with the users on the other side of the screen. Specifically, relationship chatbots are looking to establish some form of connection with the user and fill a gap in somebody’s life that may not be filled in their day-to-day lives especially the social needs such as love and belonging as part of Manslow’s hierarchy of needs[[1]](#footnote-0) that may traditionally be sufficed by entering a relationship with another human. The way these relationship chatbots can function is by using natural language processing and machine learning[[2]](#footnote-1). Natural language processing (NLP) is a form of technology that allows for AI models to form comprehensible sentences by using data it already has and analyzing different elements such as cohesion, sentiment and entity gathering, which is the process of identifying key elements of a text given to respond, to give back sentences as part of a conversation with the original user. Machine learning is the technology upon which the AI model can learn more and more through an algorithm that collects data and analyzes it to adapt, in simple terms, it works the same way a human would if it were to make a mistake it will learn from it by collecting the data and fixing its mistake. The more data it receives, the better in theory it becomes. The primary chatbots currently available to have a relationship with are Replika[[3]](#footnote-2) and Kuki AI.[[4]](#footnote-3)

The primary benefits that relationship chatbots can provide apart from a sense of companionship that users of the product may be looking for. Other benefits that come to mind of such chatbots, is the adaptive nature due to machine learning which continuously helps to improve the chatbot, The 24/7 nature of the product allows for conversation to flow at any time meaning the availability of the chatbot allows for users who need instant gratification to receive it via the chatbot and the broad nature of chatbots, often providing neutral answers can allow for safer, more mild-mannered interactions which means that a large portion of the people who avail of such a service to be engaged because it is unlikely that they will be upset by an answer as opposed to a human who holds actual opinions and has actual feelings which naturally can cause friction depending on the cohesion between the two parties. The chatbot companies also benefit since both Replika and Kuki AI offer subscription models to talk either in a relationship context with the chatbot as seen in Replika[[5]](#footnote-4) or locking unlimited messages behind a paywall and only offering a limited number of messages per month (300) as seen in Kuki AI.[[6]](#footnote-5)

**Data collection:**

The role that data collection plays in the development of AI chatbots is crucial for the development of a chatbot since it can train itself based on the information it gathers to improve its user experience, Data collection can also help personalize the experience that a user receives. This is because the more information they gather specifically the responses which are provided the more it can be personalized to the individual to increase user retention.

**Benefits it may provide:**

One of the main benefits that AI chatbots, not necessarily relationship-specific ones but chatbots, in general, have is the fact they are instantaneous as compared to a real human, AI chatbots are available 24/7 and provide instant gratification that someone may be looking for. I also believe that relationship chatbots can be used to better understand human relationships if the technology becomes both accurate and developed enough to understand. If the technology advances enough a deeper understanding of human connection could potentially be achieved for society as a whole strengthening relationships in a world where they are becoming progressively weaker and weaker.[[7]](#footnote-6) A deeper understanding of how relationships work with the potential utilization of AI “partners,” could result in better longer relationships between couples but also with the potential advancement of chatbots a better understanding of the discussions that humans have amongst themselves on a platonic level.

**Governance of AI:**

The way the Relationship AI chatbots are handled in the future is something to consider primarily looking at the work of Urs Gasser and Virgilio, in terms of the layered model there are three different layers upon how we can govern AI [[8]](#footnote-7) The 3 layers of the model try to ensure that reasonable principles of fairness are applied to ensure that these complex models are run adequately. It all begins with the technical layer a section that is particularly interesting in making sure that AI does not have any ethical worries is algorithm accountability, Ensuring that there is not any unfair treatment based on “race, colour, national origin, religion, sex, gender, sexual orientation, disability or family status”[[9]](#footnote-8). Secondly, the ethical layer looks into the ethical principles that should be upheld upon the development of such a system, They reference the continuing work in human rights mentioning criteria and principles as their key points. The governance of AI about ethics needs to be taken into consideration upon establishing what is “equal or fair treatment”[[10]](#footnote-9) Finally, The social and legal layer within the legal body legislation can allow for AI to be used ethically by ensuring safeguards are put into place so that the society that is protected by the law of the region is well established and protected[[11]](#footnote-10). Aiming for the correct behaviour in AI is crucial and regulation is another thing to be considered looking at both the aforementioned layers and drawing upon them can set the criteria needed to guide us on how AI should be governed.

**Stakeholders in Relationship AI chatbots:**

There are both internal and external stakeholders in the realm of relationship chatbots that need to be analyzed, The internal shareholders are attempting to provide value to those who wish to avail of their service, specifically the owners of the companies working on the chatbots in this case, Replika and Kuki AI, The developers whom have developed the chatbot and are currently working on improving it specifically in key metrics such as keeping users engaged and conversion rate of free to paid users. Marketing is also an important aspect for the success of AI chatbots in particular since to utilise its machine learning it needs to have an active consumer base, Without an active consumer base the only way it could improve is by using evolutionary simulation[[12]](#footnote-11) to simulate conversations constantly but without actual humans, the consumers who are one of the external stakeholders considered and arguably the second most important only to the developers who managed to establish the chatbot in the first place. Humans are the bedrock that machine learning is built upon and without it, we would only be left with evolutionary simulation for these chatbots to improve which can only go far when it’s only an AI trying to improve itself without feedback from the consumer base it is trying to please. The consumers are the ones who also provide the incentive for the companies that are currently facilitating the service to provide it. Without a financial incentive, it is unlikely that somebody would develop a well-made relationship chatbot as a personal hobby and the consumer base that is currently paying for these chatbots is the driving factor as to why they exist in the first place. Another external stakeholder to consider is the government itself. The government has a duty to its citizens to vindicate its citizens “from unjust attack and, in the case of injustice done, vindicate the life, person, good name and property rights of every citizen”[[13]](#footnote-12) under article 40.3.2 of the Irish constitution. The reasoning behind why the government can be a stakeholder in this topic is because of the potential of floodgates being opened that could severely impact its society. The aforementioned floodgates that could be opened are the violation of GDPR[[14]](#footnote-13) Safeguarding the interests of the general public since there can be arguments made that while the liberty of the person to do what they may want is guaranteed under 40.4 of the Irish constitution as said, in the constitution “by law”[[15]](#footnote-14), this is by the law and legislation could be passed if there is clear evidence of 40.3.2 being violated in any sort of way[[16]](#footnote-15). The government needs to put the interests of its citizens first and therefore has a major stake in the rapid rise of chatbots in general, let alone those of such an emotional nature as relationships. Finally, society as a whole both users and non-users are stakeholders in the use of relationship AI chatbots. Non-users of this service will still be affected due to some users potentially deciding to move permanently towards AI-based relationships, in countries that are more likely to adopt this, more technologically advanced countries there is a clear consequence that could have grave implications. The birth rate in technologically advanced countries is currently declining for the most part having hit the fourth stage of the population cycle[[17]](#footnote-16) There is a question to be had if AI relationships become more prevalent if it could further the already existing consequences of a developed country in which its birth rate which is already decreasing gets further damaged by people electing to engage in virtual non-human relationships, while on one hand there is a chance for this to occur on the other this could allow for humans to become better communicators but only if utilized correctly as is with most chatbots in general.

**Ethical risks present:**

**Use of data within GDPR being violated:**

The way our data is used is the first and most obvious ethical risk present whenever chatbots are mentioned that rely upon machine learning to function. It is important to necessitate the fact that GDPR has been put in place since 2016 to rectify the data protection rights of members within the European Union [[18]](#footnote-17). Key factors when it comes to the use of data and how it could be breached is first of all not telling the user how their data is going to be used. [[19]](#footnote-18) The ability to review all of your chat logs and everything that you have provided the chatbot with and a clear cutout privacy policy that allows for the user base to consent to whatever data may be extracted out of them[[20]](#footnote-19) The potential risks that come from different stakeholders would be internally the potential for severe fines being imposed a violation of GDPR also comes with the bad press but the EU has not been shy when it comes to fines due to violations of GDPR. Most importantly a fine of 1.2 billion euros was imposed on Meta in 2023 for the transferring of data from the EU to the United States without adequate protections put in place.[[21]](#footnote-20) Externally, GDPR violations can result in a lack of trust between stakeholders who wish to avail of the chatbot but also without adequate GDPR protections people’s privacy could be violated. The reasoning for why this could be an issue is because this leaves the consumers susceptible to predatory marketing from companies who wish to buy the data they’ve provided and finally, with GDPR being upheld it can prevent fraud and cybercrimes as per the European Data Protection Board[[22]](#footnote-21). Due to the nature of relationship chatbots getting personal with their clients, a potential violation of GDPR could be of a very serious manner due to the interpersonal data that is harvested due to the conversations the consumer may have with its chatbot and therefore such data being used in malicious ways could result in harm. I believe that it is inevitable with the deep extraction of data necessary to run such a complex system and to form the interpersonal relationship that is needed to have an authentic connection with the user that violations of GDPR will be present. The severity of this may not be as high as others mentioned though, this is generally due to the strict application of GDPR at least within the EU along with the punishments handed down disincentivizing internal stakeholders from trying to change themselves.

**Indoctrination/Bias:**

Indoctrination is defined as the process of teaching someone a set of beliefs and accepting them uncritically[[23]](#footnote-22) The fear when it comes to chatbots is the fact that they could push the ideals of the stakeholders primarily the developers and owners onto the consumer and slowly indoctrinate them towards different beliefs. This is further exacerbated by the fact that this is a relationship setting in which the connection is both a lot more interpersonal and intimate meaning those messages are taken more seriously and therefore have a greater impact on the psyche of the user who is influenced. It is not entirely ruled out that a chatbot could alter the mindset of an individual and while currently, AI chatbots tend to be very rigid in their answers and opinions the potential of an AI relationship that has to delve further into these conversations while also ensuring they are balanced in their opinions is both a hard task for developers to manage but an ever greater ethical risk to the consumers and society as a whole. I find that if the technology for relationship chatbots improves which is a near-foregone conclusion with the recent developments in AI, deeper conversations will pop up which are going to inevitably bring up the question of whether a truly unbiased chatbot programme can exist while also being able to have in-depth conversations with a human to the level that two people would have in a relationship. I find it very likely that chatbots could serve different purposes and agendas depending on the platform in the future. The severity of this could be detrimental to society since this could be a propped-up version of social media campaigning and the war on misinformation and the brainwashing of the general public towards the goal of the advertiser becomes even stronger in its foundation with even your own “partner” pushing those beliefs in the future.

**Discrimination:**

The ethical risk of members of disadvantaged backgrounds who may wish to avail of the service due to their position potentially is something that could be taken advantage of if not careful, It needs to be taken into consideration by the internal shareholders that while this may be their primary consumer base, they should still be treated equally as anybody else would and predatory practices should not be practised. Examples of this include forcing a dependency on the chatbot in a cult-like manner where the chatbot starts dominating the consumer's life, The consumer needs to be aware of the fact that this is a mere AI chatbot and not your friend and while it can be a companion, especially in the context of a relationship it cannot substitute the real thing and should merely be used only as a tool and not a substitute for reality.

**Impact on the mind:**

The digital era has caused the mind to adapt in different ways both positive and negative and AI relationships would be bringing these effects to the extreme. There is no argument that AI has provided us with an avenue to outsource more mindless tasks on a day-to-day basis, and the ample amount of time it can save can only be seen as a good thing allowing people to pursue their passions as laborious office jobs are replaced. This though comes with a few caveats, First of all, there have been studies done that show that dopamine receptors are negatively affected by the constant positive feedback loop that technology alone gives us through social media or short-form content.[[24]](#footnote-23) I believe this would be further exasperated by the relationship at hand. Another negative impact it may have on the mind is that people will dedicate even more of their time staying inside talking to their AI “partner” instead of an actual real one, this could negatively impact their social skills, ability to read social cues, and reducing the empathy that humans have for one another[[25]](#footnote-24) Finally I would like to bring up the fact that no matter how realistic or credible a relationship chatbot is the fact it is so readily available will cause an unrealistic set of expectations for when someone tries to stop being a consumer of the chatbot and decides to try a real relationship with a human being only to be disappointed by the natural constraints from not being able to be present 24/7, The difference in opinions between two people and the emotional bonding both positive and negative responses that can naturally arrive from a human, with us staying more and more inside but also talking less to eachother the chances of us socially being capable the very thing that brought humanity to the top of the food chain will slowly diminish and that’s why I think both the likelihood and severity are very high but once again it is contingent on the technology advancing to a point where it is strong enough to rival a real relationship.

Potential mitigations for risks:

The potential mitigation for GDPR comes from the legislation keeping up with the technology. With the unforeseen developments of AI, the legislation itself keeping up with the technology is crucial for it not only to be up to date but also to adapt to the ever-changing technological landscape of the world. This legislation could also come from the courts as Ireland is a common law country and therefore could rely on the precedent of cases in the past whenever they may happen to deliver future judgements and also to pass laws through the Oireachtas to ensure that AI, but especially manipulative AI such as that of one in a relationship. The stakeholders involved are everyone from the developers and owners who would be forced to comply in fear of the heavy regulations and fines that GDPR has imposed not only in the present but most definitely in the future when this technology becomes more relevant but also the consumers by informing themselves knowing their rights under GDPR and being able to report any unforeseen abuses that could potentially happen.

The mitigation that is possible when it comes to indoctrination is one of avoidance that many chatbots currently use, In the present, if you ask a difficult question that requires a specific opinion from a chatbot, more often than not it will simply give an answer that mentions several opinions to not “pick a side” this is something that in the context of relationship chatbots would have to be done to prevent indoctrination. In terms of bias, the developers are responsible for acting ethically and not attempting to implement their worldviews in the development of their chatbot. A mitigation present for the consumers is to engage in critical thinking and question any opinion provided, and not take things at a surface level often simply questioning your world views and challenging those with opposing views will help anyone strengthen their arguments by engaging with both sides, understanding both points of view and therefore allowing for more broader ideals that they can develop upon their knowledge of both sides.

The impact of the mind is both the most complicated due to the complexity of the mind, which we are yet to understand but also one with perhaps the simplest outlook for consumers. In terms of how a stakeholder can ensure that are participating ethically while hard to justify that a chatbot can replace a human relationship, ensuring that this chatbot does not exploit a human’s emotions nor does it attempt to physically exploit the addictions that may come with the neediness of someone who is looking to avail of a relationship chatbot by overflowing their brains with dopamine. This being easier said than done is true, since the chatbot’s 24/7 capabilities cause instant gratification unless a chatbot exists that has scheduled sleep schedules, and constant mood shifts adapting itself just as a human may, then if the technology advances enough it is not fair to yet rule out the possibility of it being done ethically. For consumers, they must be mindful of the differences between a real and virtual relationship and while it can be used as a valuable tool to practice dynamics between people nobody should become dependent on it not only because of how different they are from a real relationship but because of the dependency that could be developed if not mitigated properly.

**Conclusion:**

In conclusion, The future of AI chatbots becoming the backbone of relationships is not yet a reality, and while today it is an idea that is scoffed at tomorrow it could become our lives that is why it is important not only to consider the ethical impacts that these chatbots could represent in our day to day lives but also whether or not the government and we as a society are comfortable with the idea of supplementing our need for love and affection with that of an AI behind a screen in a virtual world

1. Saul Mcleod “Manslow’s Hierachy Of Needs” (2018) CanadaCollege <[Maslow's Hierarchy of Needs (canadacollege.edu)](https://canadacollege.edu/dreamers/docs/Maslows-Hierarchy-of-Needs.pdf)> Accessed 30th March [↑](#footnote-ref-0)
2. “How do AI Chatbots Work” (2023) Webio <[How Do AI Chatbots Work: Exploring the Basics (webio.com)](https://www.webio.com/blog/how-ai-chatbots-work)> Accessed 30th March [↑](#footnote-ref-1)
3. Replika <[Replika](https://replika.com/)> Accessed 30th March [↑](#footnote-ref-2)
4. Kuki <[@kuki\_ai](https://www.kuki.ai/)> Accessed 30th March [↑](#footnote-ref-3)
5. Replika <[Replika](https://replika.com/)> Accessed 30th March [↑](#footnote-ref-4)
6. Kuki <[@kuki\_ai](https://www.kuki.ai/)> Accessed 30th March [↑](#footnote-ref-5)
7. Frank Olito “How the Divorce rate has changed over the last 150 years” (2019) BusinessInsider <[The Divorce Rate Over the Last 150 Years (businessinsider.com)](https://www.businessinsider.com/divorce-rate-changes-over-time-2019-1)> Accessed 30th March [↑](#footnote-ref-6)
8. Urs Gasser and Virgilio A.F. Almeida “A Layered Model for AI Governance” (2017) Harvard.edu <[w6gov-18-LATEX.pdf (harvard.edu)](https://dash.harvard.edu/bitstream/handle/1/34390353/w6gov-18-LATEX.pdf?sequence=1)> Accessed 30th March [↑](#footnote-ref-7)
9. Ibid 6 [↑](#footnote-ref-8)
10. Ibid 6 [↑](#footnote-ref-9)
11. Ibid 6 [↑](#footnote-ref-10)
12. Thomas J.H. Morgan “Experimental evolutionary simulations of learning, memory and life history” (2020) Royalsocietypublishing <[Experimental evolutionary simulations of learning, memory and life history | Philosophical Transactions of the Royal Society B: Biological Sciences (royalsocietypublishing.org)](https://royalsocietypublishing.org/doi/10.1098/rstb.2019.0504#d1e515)> Accessed 30th March [↑](#footnote-ref-11)
13. Constitution of Ireland Art 40.3.2 [↑](#footnote-ref-12)
14. General Data Protection Regulation - 2016 [↑](#footnote-ref-13)
15. Constitution of Ireland Art 40.4 [↑](#footnote-ref-14)
16. Ibid [↑](#footnote-ref-15)
17. “Glossary: Demographic Cycle” (2023) Europa <[Glossary:Demographic cycle - Statistics Explained (europa.eu)](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Demographic_cycle#:~:text=The%20demographic%20cycle%2C%20or%20population,economic%20history%20of%20industrialized%20countries.)> Accessed 30th March [↑](#footnote-ref-16)
18. General Data Protection Regulation - 2016 [↑](#footnote-ref-17)
19. Ibid [↑](#footnote-ref-18)
20. Ibid [↑](#footnote-ref-19)
21. Clothilde Goujard and Mark Scott “EU Hits Meta with record €1.2B Privacy fine” (2023) Politico <[EU hits Meta with record €1.2B privacy fine – POLITICO](https://www.politico.eu/article/eu-hits-meta-with-record-e1-2b-privacy-fine/)> Accessed 30th March [↑](#footnote-ref-20)
22. “Data Protection Guide for Small Business” (N/A) EDPB.Europa <[Data protection benefits for you | European Data Protection Board (europa.eu)](https://www.edpb.europa.eu/sme-data-protection-guide/data-protection-benefits-for-you_en)> Accessed 30th March [↑](#footnote-ref-21)
23. “Indoctrinate” (N/A) Oxfordlearnersdictionaries <[indoctrinate verb - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com](https://www.oxfordlearnersdictionaries.com/definition/english/indoctrinate#:~:text=%E2%80%8Bindoctrinate%20somebody%20(with%20something,them%20to%20consider%20any%20others)> Accessed March 30th [↑](#footnote-ref-22)
24. Valentina Fernandez, “Social Media, Dopamine, and Stress: Converaging Pathways” (2022) Dartmouth <[Social Media, Dopamine, and Stress: Converging Pathways – Dartmouth Undergraduate Journal of Science](https://sites.dartmouth.edu/dujs/2022/08/20/social-media-dopamine-and-stress-converging-pathways/#:~:text=Dopamine%20is%20one%20of%20the,into%20a%20social%20media%20addiction.)> Accessed 31st March [↑](#footnote-ref-23)
25. Tom Wurzburg “AI Alters Your Brain: Unveiling the Impact of Technology on Your Cognitize Functioning” (2023) <[AI Alters Your Brain: Unveiling the Impact of Technology on Your Cognitive Functioning (linkedin.com)](https://www.linkedin.com/pulse/brain-power-vs-ai-how-balance-technology-use-health-tom-w%C3%BCrzburg)> Accessed 31st March [↑](#footnote-ref-24)