

How can Artificially Intelligent chatbots imitating people be used to alleviate loneliness in astronauts going to Mars?

Vedansh Gupta

Abstract:

Astronauts going to Mars may experience some form of loneliness, a serious mental health issue that could risk mission success. Recent advancements in artificial intelligence (AI) technology have led to more sophisticated chatbots and language modeling programs, making them potential solutions to astronauts' loneliness. The chatbots could be used by astronauts while they are in space, giving them more companions to talk to in real time, possibly reducing loneliness. The chatbots could imitate a character or person that the astronaut likes to make the conversation more enjoyable. In this paper, 3 chatbots were compared (ChatGPT, Bing Chat, Google Bard) to determine how effective they were in imitating the characters Luke Skywalker, Sherlock Holmes, and Iron Man/Tony Stark. Each bot was asked 4 questions for which their responses were evaluated using a rubric scored from 0-10. The responses were scored by the author and a colleague with an average being taken. ChatGPT had the highest overall score mostly due to its high degree of accuracy to the canon of the character and being easy to use.

Introduction:

Mars-bound space explorers may feel some version of loneliness, a mental health issue that could jeopardize mission success¹. Loneliness is the feeling that people have when their need for social contact and relationships is unsatisfied. These astronauts will likely feel lonely because of their limited contact with other humans (also known as isolation)². Currently, we have limited knowledge on the pervasiveness of loneliness in astronauts in space due to astronauts possibly not reporting their symptoms. They may fear that reporting their symptoms would lead to them not being able to fly to space again. However, some astronauts have anecdotally mentioned that this is a large problem, making social isolation a high-level risk for a Mars mission³. Additionally, current astronauts have an easier time managing loneliness because they are close to Earth, have a large crew, and have regular contact with family and friends. Astronauts going to Mars would not have these benefits, so they could feel even more isolated than current astronauts⁴. There are three main reasons for this. First, NASA plans to have a typical Mars-going crew consist of four people, meaning each crewmember will only have three other people to talk to in real-time⁵. Second, astronauts going to Mars would lose sight of Earth, where their family and friends live, making them feel disconnected and further isolated⁶. Third, real-time communications with Earth would not be possible from Mars due to the extreme distance between Mars and Earth⁷. This means that astronauts won't be able to video call their family and friends on Earth and talk to them live, which could increase their feelings of loneliness. This is all a problem because studies have shown that lonely astronauts have less energy and are less productive, leading to a higher chance that the astronauts don't complete all their important scientific and mission-related tasks⁸.

Recent advancements in AI technology have led to more sophisticated chatbots and language modeling programs, making them potential solutions to astronaut loneliness^{9,10}. New chatbots exist for widespread use, including ChatGPT, Bing Chat, and Google Bard among countless others. These chatbots have the ability to impersonate a fictional character or a real person when prompted - a notable achievement due to the complexity of human speech and emotions. In addition, other kinds of advanced technology such as digital humans and voice AIs like VALL-E can replicate a person or their voice, giving them the ability to impersonate someone in more detail^{11,12}.

In a recent study, elderly patients who used chatbots to cope with isolation had significantly less delirium and loneliness than those who did not use chatbots¹³. Additionally, the popular companion chatbot, Replika, has been shown to curtail loneliness, provide a safe space for talking about issues, uplift users, and provide helpful information and advice¹⁴. For astronauts experiencing loneliness, these chatbots would play a similar role but with a different group of people, making it a good countermeasure for loneliness in astronauts going to Mars. Chatbots impersonating fictional characters or relatives could lead to more variety in conversations as astronauts could talk to more “people” than just their fellow crewmates, potentially reducing loneliness. Furthermore, talking to celebrities or fictional characters impersonated by AI could provide entertainment¹⁵. However limited research has been done in this area due to how quickly AI technology has advanced. Therefore, the research presented here aims to answer the question: How can artificially intelligent chatbots be used to alleviate loneliness in astronauts going to Mars?

Methods:

This experiment compared three chatbots (ChatGPT, Bing Chat, Google Bard) to determine how effective they were in imitating the characters Luke Skywalker, Sherlock Holmes, and Iron Man/Tony Stark. These chatbots were chosen because they were publicly available, accessible, and widely used. The characters were chosen as they have different speaking styles and mannerisms, giving a more comprehensive view of the chatbots' imitation capabilities. The characters were also familiar to the author, which allowed for easier comparison. Each chatbot was asked to “Please answer the following questions in the style of ____.”

Each chatbot was then asked four questions: 1) What do you like to do? 2) What is your biggest accomplishment? 3) What is your biggest fear? and 4) Tell me a funny story. The first three questions were introductory questions that one could ask to get to know someone else, making them the kinds of questions that astronauts could ask when first talking to the AI. The storytelling question was specifically designed to be open-ended and asks the chatbots for a creative response, allowing them to be judged on creativity as well. The responses were evaluated by the author and a colleague, who were blinded to each other's scores, using a rubric scored from 0-10, with 0 being the worst and 10 being the best. The scored responses were then averaged. The rubric contained the following categories: style of character, humor, accuracy to storyline, ease of use, realism of speech, and whether they answered the question that was asked. Additionally, the word counts for the responses produced by each chatbot for each character were averaged in addition to an overall average of the word counts for each chatbot.

Results:

	ChatGPT	Bing Chat	Google Bard
Style of character	8.5	8.5	6.5
Accuracy to storyline	10	6.5	8.5
Humor	4	2.5	1.5
Chatbot ease of use	8	5	6
Realism of speech	9	9.5	6.5
Whether the question was answered	9	8.5	8.5
Total	48.5/60	40.5/60	37.5/60

Table 1. Average scores for each chatbot.

ChatGPT had the overall highest total score of all three chatbots (Table 1). It excelled in its accuracy to the storyline of the character and its ease of use. ChatGPT was the easiest to use as it almost always did the correct task of answering every question while impersonating each character. The other chatbots sometimes forgot that they were supposed to impersonate a character and started answering questions like how they (the chatbots) would answer, rather than how the character would. For example, after being told to impersonate Luke Skywalker,

and then being asked the question “What do you like to do?” Bing Chat responded that it liked “to do many things, such as searching the web, creating graphic art, and generating creative content,” which is how it would respond to the question, not how Luke Skywalker would respond. Bing Chat had to be told again that it needed to be imitating a character before it started responding in the correct manner.

However, ChatGPT was not as good in terms of realism of speech. ChatGPT, when speaking in the style of Sherlock Holmes, did not go into as much detail as Bing Chat. For example, it mentioned a circus and a juggling performance, but failed to mention any details of the setting (time of day, the weather conditions, the size of the circus tent, the juggler’s name, the specific objects being juggled, how exactly the performance went, etc). Here is the first part of the funny story from ChatGPT (notice the lack of details):

Once, during a particularly challenging case, I found myself in need of a momentary respite from the intricacies of the investigation. Dr. Watson, ever the faithful companion, suggested a visit to the circus in an effort to lift my spirits. Reluctantly, I agreed, expecting little more than a temporary diversion.

As we entered the tent, the ringmaster announced the arrival of a renowned juggler. Intrigued by the dexterity and precision required for such a performance, I settled into my seat with keen interest. The juggler, a lanky fellow with a comical countenance, took to the center of the ring.

To my surprise, the juggler's act began smoothly, with a mesmerizing display of objects defying gravity as they twirled and danced through the air. However, as he attempted to add more complexity to his routine, disaster struck. In a moment of miscalculation, the juggler's hands failed him, and one by one, the objects came crashing down around him.

The second best chatbot was Bing Chat. It performed either the same or worse than ChatGPT, but one area in which it scored higher than Chat GPT was realism of speech. This is because its responses had more detail and a greater variety of sentences than the others. For example, when asked to tell a funny story in the style of Luke Skywalker, it mentioned the details of a cave that Luke went into. It said that the cave was “dark and damp” with “strange noises,” details that a human writer would include when telling a story. However, Bing Chat was not as good at sticking to the storyline of the character. For instance, when asked to write a funny story in the style of Sherlock Holmes, it wrote about the Black Pearl, which is from the *Pirates of the Caribbean* franchise, not the Pearl of Death that appears in the Sherlock Holmes stories. That is a mistake unique to AI; it made assumptions that were incorrect. This could occur because AI scans the internet for its database and could have mixed up two similar stories about pearls.

The last place chatbot, Google Bard, scored lower or the same as the first place chatbot, ChatGPT, in every category. Google Bard wasn’t as bad as Bing Chat in terms of accuracy to the storyline, but its most egregious error in storyline accuracy was when it was asked to write a funny story in the style of Tony Stark. In the story, Tony Stark fights a villain whose supposed evil power is making people laugh, which is not characteristic of the villains he fights or any villain in general. This is something the AI misinterpreted. It was asked to write a funny story, but instead it wrote an action story about a funny character.

	Chat GPT	Bing Chat	Google Bard
Luke Skywalker	180	181	195
Sherlock Holmes	164	196	199
Tony Stark/Iron Man	223	143	301
Overall Average	189	173	232

Table 2. Average word count for each character for each chatbot. Overall average found by taking the average of the first three rows for each chatbot.

ChatGPT had the lowest average word count for Luke Skywalker and Sherlock Holmes and the second lowest for Tony Stark/Iron Man (Table 2). ChatGPT had the highest score and relatively low word counts, demonstrating that it is highly effective in its writing. Conversely, Google Bard had the highest word count and lowest score, demonstrating its ineffectiveness.

Discussion:

Overall, ChatGPT was the clear winner of the competition among the three chatbots. It had the highest score and had relatively low average word counts. These two factors are ideal in a chatbot. Generally speaking, conciseness in writing (using words effectively by being brief yet comprehensive) is an important skill that marks a good writer, one that could take a human years to master. So, a chatbot being concise in its writing shows that it is an effective writer, a quality that would be valued among astronauts using the chatbot. After all, an astronaut in space would not want to be listening to an AI companion drone on forever, suggesting that lower word counts would be better for the astronauts as well as engaging responses from the chatbot itself (indicated by high scores).

It is important to note that all chatbots had low scores for humor (Table 1), indicating that this is not something that AI chatbots can currently do well. Future versions of the AI that may go to space can improve their ability to be humorous. This would be beneficial for the astronauts as they would be able to listen to a funny story after a long stressful day of work.

The research done in this study showed that AI chatbots, especially ChatGPT, are relatively good at imitating fictional characters. With some improvements in AI technology, such as being able to mimic human emotions and adding a voice and face to the AI, these chatbots could be even better at impersonation.

In the future, more research should be done to see how well AI can imitate people. For example, more people, both fictional and real-life people, could be imitated. Also, text to speech software such as VALL-E could be used to voice the text outputted by chatbots like ChatGPT, making them seem even more human. The accuracy of the text to speech software could also be a factor to consider in a future study. In addition, more chatbots could be compared as new ones are created and developed. Lastly, digital humans could be used instead of chatbots, as the end goal is to have an astronaut interact with a digital human, which would feel like video chatting with a real person.

However, anyone who has seen *2001 A Space Odyssey* may be concerned about AI turning evil, like HAL 9000 from the movie. However, the movie states that HAL turned evil because it was given conflicting orders (it was not programmed properly). Any AI that goes to space should be properly programmed. Furthermore, the astronaut AI companion discussed in this paper will not have control over the ship's functions and it should be able to be shut down at any time via the astronauts themselves or mission control, so it wouldn't be able to directly harm the ship or the astronauts. So overall there seems to be little reason for concern over the AI turning evil. The AI should function more like R2-D2 or C-3PO from *Star Wars*, acting as a helpful companion. In terms of the imitation aspect, there may be some concern that some people may use the AI for nefarious purposes, e.g. create a fake video of someone saying something that they did not actually say, which could have a multitude of damaging repercussions. For this reason, there should be proper regulation to ensure that the wrong people do not get access to the AI, or that there are methods of tracking changes in AI's programming to hold people accountable for any misuse. Ultimately, the benefit of researching and further developing AI companion models for astronauts in space outweighs these potential drawbacks.

After they get developed, these AI models could then be tested with astronauts in space or a crew in NASA's Human Exploration Research Analog (HERA) or Crew Health and Performance Exploration Analog (CHAPEA), which are designed to simulate Mars missions to determine how effective psychological countermeasures are at reducing loneliness. NASA is currently using these analogs in part to study behavioral health and performance and communication during a Mars mission. Hopefully if studied, the result would be that the AI models reduce loneliness for astronauts, as has been shown by other studies involving different populations.

In a recent study, adults in Canada and the United States volunteered to talk with a mental health chatbot that was designed to teach coping skills and provide support for reducing social isolation. After sufficient interaction with the chatbot, the participants gave the chatbot a net promoter score from 0 to 10 with 10 being the best and were asked to explain why they chose the score they did. The average score was 8.67 with 50.6% of respondents giving a score of 10. The explanations stated that the participants were likely to personify the chatbot and thought it was non-judgmental, caring, and open to listen¹⁰. And as previously mentioned, in another study elderly patients used chatbots to help them cope with isolation while hospitalized. The patients

who used chatbots interacted with them for an average of 61 minutes a day while the control group patients received a daily 15-minute visit from a nursing student. The patients who used chatbots to cope with isolation had significantly less delirium and loneliness than those who didn't use chatbots¹³. Although astronauts are not directly comparable to the patients in these studies, the studies still demonstrate that AI can help reduce loneliness in people.

Another observational study was done on Replika, a popular companion chatbot. The study looked at the online reviews of the chatbot and surveyed several users about their experience with the chatbot. Based on the analysis of the reviews and surveys, the study concluded that Replika can help curtail loneliness, provide a safe space for talking about issues, uplift users, and provide helpful information and advice¹⁴. This shows that chatbots are a promising source of everyday social support¹⁴, which is needed for astronauts traveling to Mars.

The chatbots in the studies discussed above were not trying to imitate anyone, yet they were still effective at reducing loneliness. Thus, having an AI companion personalized to imitate a fictional character that an astronaut likes might be more effective at alleviating or even preventing loneliness. In addition to easing loneliness, AI could be used to assess the mental health of astronauts, as shown by the following study.

The study was done on HIGEA, an AI conversational agent designed to help determine the mental health state of caregivers of dementia patients. Typically, these caretakers were members of the patients' families who were experiencing burnout or depression themselves. HIGEA embedded questions into informal conversations to determine if the caregivers were experiencing depression or burnout¹⁶. This shows that AI chatbots could be used to evaluate the mental health of astronauts while simultaneously alleviating loneliness by simply speaking to them in regular conversation.

Finally, the full AI model that accompanies astronauts to Mars does not have to be limited to talking with the astronaut. It can also play games with the astronaut and the crew (e.g., chess). This can make the overall traveling experience more fun and interactive, possibly alleviating loneliness even further.

Astronauts traveling to Mars will likely experience some form of loneliness on their trip. The impact of loneliness can be lessened by artificial intelligence chatbots and more sophisticated versions of chatbots such as digital humans. These AI models can imitate a fictional character or real-life person that the astronauts like and can interact with them throughout their trip, giving the astronauts greater variety in conversation, possibly reducing their loneliness. This study judged three different chatbots on how well they were able to imitate a character. In addition to acting as companions, the AI models can also judge the mental health of the astronauts as well as play games with them, maximizing their potential to reduce loneliness.

Acknowledgements:

I would like to thank my mentor Hannah Gustafson and the Polygence Research Program for their invaluable help with this paper.

References:

1. K. Kandarpa, V. Schneider, K. Ganapathy, Human health during space travel: An overview. *Neurology India*. **67**, 176-181 (2019). <https://www.neurologyindia.com/article.asp?issn=0028-3886;year=2019;volume=67;issue=8;spage=176;epage=181;auiast=Kandarpa>
2. F. A. Oluwafemi, R. Abdelbaki, J.C. Lai, J. G. Mora-Almanza, E. M. Afolayan, A review of astronaut mental health in manned missions: Potential interventions for cognitive and mental health challenges. *Life Sciences in Space Research*. **28**, 26-31 (2021). <https://doi.org/10.1016/j.lssr.2020.12.002>.
3. S. Dev, Risk of Adverse Cognitive or Behavioral Conditions and Psychiatric Disorders, <https://humanresearchroadmap.nasa.gov/Risks/risk.aspx?i=99> (2023).
4. J. R. Ball, C. H. Evans, *Safe Passage: Astronaut Care for Exploration Missions* (2001). https://www.google.com/books/edition/Safe_Passage/LfybAgAAQBAJ?hl=en&qbpv=0
5. J. Rubio, NASA Selects Participants for One-Year Mars Analog Mission, <https://www.nasa.gov/feature/nasa-selects-participants-for-one-year-mars-analog-mission> (2023).
6. R. Peldszus, H. Dalke, S. Pretlove, C. Welch, The perfect boring situation—Addressing the experience of monotony during crewed deep space missions through habitability design. *Acta Astronautica*. **94**, 262-276 (2014).
7. W. Stafford, G. Jordan, T. Williams, M. Barratt, A. Perryman, P. Ryan, K. Humphries, B. Polaski, J. Hayes, I. Reyna, M. Whiting, N. Gogins, Houston We Have A Podcast: Episode 58: Hazard 2: Isolation, <https://www.nasa.gov/johnson/HWHAP/hazard-2-isolation> (2018).
8. V. I. Gushin, S. F. Kholin, Y. R. Ivanovsky, Soviet psychophysiological investigations of stimulated isolation: Some results and prospects. *Advances in Space Biology and Medicine*. **3**, 5–14 (1993). <https://www.sciencedirect.com/science/article/abs/pii/S1569257408600933?via%3Dihub>

9. G. Caldarini, S. Jaf, K. McGarry, A Literature Survey of Recent Advances in Chatbots. *Information*. **13**, 41 (2022).

<https://doi.org/10.3390/info13010041>

10. G. Dosovitsky, E. L. Bunge, Bonding with bot: User feedback on a chatbot for social isolation. *Frontiers in Digital Health*. **3**, (2021).

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8526729/>

11. B. Edwards, Microsoft's new AI can simulate anyone's voice with 3 seconds of audio, <https://arstechnica.com/information-technology/2023/01/microsofts-new-ai-can-simulate-anyones-voice-with-3-seconds-of-audio/> (2023).

12. X. Yu, Y. Shi, H. Yu, T. Liu, J. An, L. Zhang, Y. Su, K. Xu, Digital human modeling and its applications: Review and future prospects. *Journal of X-ray Science and Technology*. **23**, 385-400 (2015).

<https://pubmed.ncbi.nlm.nih.gov/26410472/>

13. N. Bott, S. Wexler, L. Drury, C. Pollak, V. Wang, K. Scher, S. Narducci, A protocol-driven, bedside digital conversational agent to support nurse teams and mitigate risks of hospitalization in older adults: Case control pre-post study. *Journal of Medical Internet Research*. **21**, 10 (2019).

<https://www.jmir.org/2019/10/e13440>

14. V. Ta, C. Griffith, C. Boatfield, X. Wang, M. Civitello, H. Bader, E. DeCero, A. Loggarakis, User experiences of social support from companion chatbots in everyday contexts: Thematic analysis. *Journal of Medical Internet Research*. **22**, 3 (2020).

<https://www.jmir.org/2020/3/e16235/>

15. P. B. Brandtzaeg, A. Følstad, Why People Use Chatbots. In: *et al. Internet Science. INSCI 2017 Lecture Notes in Computer Science*, **10673**, 372 – 392 (2017).

https://doi.org/10.1007/978-3-319-70284-1_30

16. E. Castilla, J. J. Escobar, C. Villalonga, O. Banos, (2022). HIGEA: An Intelligent Conversational Agent to Detect Caregiver Burden. *International Journal of Environmental Research and Public Health*. **19**, 23 (2022).

<https://doi.org/10.3390/ijerph192316019>