With our world becoming more digital by the day, countless fields and industries are utilizing artificial intelligence to develop and advance. Mental health is one such field adopting applications of artificial intelligence, specifically the use of chatbots for therapy. Since chatbots providing mental health care is a fairly new concept, the impact of these therapy bots remains largely unexplored. For this reason, our group seeks to answer this research question: How effective are machine learning chatbots as a tool for therapy?

Although our group members hail from seemingly different disciplines—psychology, computer science and mathematics—this research question connects us all and cannot be answered fully without each member's unique perspective from his or her respective field. From the math perspective we gained insight into the logical reasoning abilities of artificial intelligence; computer science elaborated on the performance of the algorithms in chatbots; and the psychology perspective examined how the chatbots perform specifically in social therapeutic environments.

The first source from the psychology world discusses the ethics of anthropomorphizing robots. Although the discussion of negative implications from this tendency humans have to project human traits onto inanimate objects is both relevant and interesting, the main application for our research question concerns the portion of the paper that details situations where anthropomorphizing could be potentially beneficial. The source comparing human conversations to robot conversations gave a psycho-linguistic view on whether a person knowing they are chatting with a robot impacts the content or style of the conversation. A separate article expands on this idea and investigates self disclosure and whether discussing specifically personal matters to robots vs other people has an impact on the resulting benefits. All of this research is important to answering our research question because if people cannot discuss certain topics or personal

information with robots, or if they do not have the same benefits from sharing with a robot, then of course the efficacy is impacted. Finally, the articles about Socially Assistive Robots and Woebot, a chatbot that mimics cognitive behavioral therapy, both serve to help understand the direct application of these social robots in therapy, and how and in what way they can be effective for different populations.

As mathematics is characterized by logical thinking and reasoning, the sources from the mathematics discipline focus on the effectiveness of the reasoning systems of artificial intelligence. One of the main issues with chatbots is the technology itself. While humans are capable of logical thinking, computers do not have intuition or understand context. When evaluating therapy chatbots, one must consider the chatbots' ability to comprehend the logical context for a conversation. After all, how can chatbots provide mental health care if they are not equipped with the logic to understand the context of the conversation? In addition to the articles about reasoning systems, one of the math sources discusses whether machines are capable of thinking abstractly like humans. Without the ability to think abstractly, therapy bots cannot "read between the lines," or relate distinct concepts. In answering the research question, one should think about whether this trait affects the overall performance of chatbots.

Chatbots themselves have several key characteristics derived directly from Computer Science. Modern chatbots, in particular, are often implemented using some form of "machine learning," a branch of Artificial Intelligence that seeks to simulate computers learning things. Some such algorithms rely on making connections between words and ideas, like humans do, while others aim to simulate a brain, or "neural network," to control the bot. Many sources from the field of Computer Science involve studying the efficacy and success of such algorithms in creating realistic and helpful chatbots.

Chatbots One such source that was chosen for this research involved a case study of Microsoft's "Tay" twitter bot. Unfortunately, this highlighted a major problem: unguided learning can lead to unpredicted results. Tay was targeted by Internet groups purely for entertainment and was taken down for learning to constantly tweet bigoted and hateful views. Overall, Computer Science papers aim to tackle the problems that have been discovered since the conception of chatbots. Two such sources discuss more complex algorithms that blend together a few different approaches to find a healthy medium. It is important to recognize that when humans learn something new, we have a wealth of knowledge we apply to the topic even if we do not realize it. Simulating that in chatbots has proven to be difficult, but advanced algorithms applied with a more controlled dataset and a specific goal in mind (e.g. assistive therapy) can produce chatbots notably more effective than before.

Whether it is for better or for worse, artificial intelligence is inevitably shaping our world. Since a multiplicity of factors determine the efficacy of therapy chatbots, an interdisciplinary approach drawing from psychology, mathematics, and computer science is critical. Although each discipline approaches the question from a different direction, the collection of sources work together to suggest that chatbot therapists are an effective supplement for therapy, but they are not a full substitute for primary types of therapy.

HAVE YOU HEARD A-BOT ROBO THERAPY?

Crossman, M. K., Kazdin, A. E., & Kitt, E. R. (2018). The influence of a socially assistive robot on mood, anxiety, and arousal in children. *Professional Psychology: Research and Practice, 49*(1), 48-56. Doi: http://dx.doi.org.ezproxy.neu.edu/10.1037/pro0000177

This study looks at how socially assistive robots (SARs) can be used to alleviate children's suffering and help with mental health. After either waiting quietly, interacting with a SAR, or interacting with a SAR turned off, children who interacted with the SAR showed a significant increase in positive emotion, showing that it was the interaction with the SAR that improved mood rather than physical aspects of the SAR. However, interacting with the SAR had no effect on anxiety, arousal, or negative mood. This suggests that SARs may be most useful in helping mental health by boosting mood rather than reducing anxiety or negative moods, which is valuable information for our research question. The other useful portion of the article is the discussion on why robots may be potentially helpful with mental health in overall.

D'Alfonso, S., Santesteban-Echarri, O., Rice, S., Wadley, G., Lederman, R., Miles, C., … Alvarez-Jimenez, M. (2017). Artificial intelligence-assisted online social therapy for youth mental health. *Frontiers in Psychology, 8*, 796.

On the more human-computer interaction side of Computer Science, this study focuses on developing a chatbot and respective interface for communicating with said chatbot to assist with social therapy for the youth in an attempt to treat mental illness / improve mental health. Specifically, the authors discuss the methodology behind developing the "moderated online social therapy" (MOST) web application. The app is designed to be interactive by providing feedback, suggestions, and comments for its users. The authors also

go into detail regarding several of the algorithms used in designing various aspects of their software. Keeping track of emotion scores, sentiment, and keyword analysis provides them with data that can be used to make the best suggestions to a user. For the authors, the chatbot foundation of their app is involved in making these suggestions. It provides a more conversational, even human, feel to interactions with online therapy and allows for more individualized discussions with prospective patients.

Darling, K. (2017). "Who's Johnny?" Anthropomorphic Framing in Human–Robot Interaction, Integration, and Policy. *Oxford Scholarship Online*,1-21. doi:10.1093/oso/9780190652951.003.0012

In this source, Darling discussed the ethics of use and policy of robots in human-robot interactions, considering how we lean towards humanizing technology despite knowing it is not alive. She examines how framing robots in human terms can lead to issues in areas from emotional manipulation to privacy. This relates to our research, since in terms of therapy robots, privacy of personal data is a particularly delicate issue and the ethics need to be considered. In addition, she discusses potential benefits to anthropomorphizing robots and how many concerns about robots seem to ignore the potential for important supplemental relationships, as a chatbot is engineered to be, and instead regard robots as trying to replace human contact. Although this source is not peer reviewed research, it was published by the Oxford University Press, and is an extensively cited and reliable source of information.

Dutta, S., Joyce, G., & Brewer, J. (2017). Utilizing chatbots to increase the efficacy of information security practitioners. In *Advances in Human Factors in Cybersecurity* (pp. 237–243). Springer, Cham.

While not peer reviewed, this conference paper provides a unique look on the wide range

of possible uses for chatbots. The authors explore various ways in which chatbots can be utilized to enhance cybersecurity, primarily concerning themselves with analyzing malicious behavior and assisting with penetration testing (an approved cyber attack performed to assess weaknesses in a computer system). The study emphasizes the potential benefits of chatbots to simulate the insight provided from colleagues, but immediately. Security practitioners could stay up to date with modern ideas and the current threat landscape without constantly needing to check in with other professionals. This advantage would be paralleled in our topic, where patients / therapists could use trust in chatbots to provide relatively accurate responses to more common / mundane issues. Chatbots also have an advantage when analyzing data. The authors explain how chatbots can tap into their computer side and go over (in this case cyber information) data, providing analysis for a user. While a professional's take on data is always important, the ability to interface with a machine capable of providing base insights almost instantaneously is invaluable.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health,4*(2), 1-19. doi:10.2196/mental.7785

This study researched the efficacy of a conversational agent called Woebot in treating college students with anxiety and depression by modeling cognitive behavioral therapy (CBT) principles. In comparison to the control group who only received a mental health e-book to self-educate, participants in both the control and experimental group had a significant decrease in anxiety after 2 weeks, but in terms for depression, only the Woebot group had a significant reduction of symptoms. In addition, the Woebot users had a higher

levels of general satisfaction, and a greater amount of emotional awareness suggesting that the Woebot is an effective and easy way to receive CBT. This study directly relates to answering our research question in determining efficacy of chatbot therapy, since Woebot is a form chatbot therapy.

Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior, 49*, 245-250. doi:10.1016/j.chb.2015.02.026

This study examined the differences in communication when people converse with an intelligent agent versus when people converse with another human. The study collected data from 100 conversations with a chatbot called Cleverbot and 100 instant messaging conversations for comparison. Although there was a notable difference in conversations, they indicate that people are simply altering their messaging style to mimic the bot's style of communication. When people communicated with a chatbot, they tended to send shorter messages, but would send more twice as more messages to chatbots than to other people, disproving the misconception that people would be more uncomfortable communicating with chatbots. We can apply these findings to our research question of chatbot's efficacy, knowing that despite people changing their communication style in chatbot therapy, it does not impact the actual content or quality of communication.

Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication, 68*(4), 712-733. doi:10.1093/joc/jqy026

This study compared the psychological effects of self-disclosure after a conversation with either another person or a computer. To collect data, the authors recruited 128 college

students to participate in the study. Although in both conditions the participant would talk with a person, the perceived identity was manipulated to either be a computer or human. The study found that there was not a notable difference between the emotional, relational, and psychological benefits of disclosing to a chatbot and disclosing to a human partner. Participants felt "significantly better" after emotional disclosure, regardless of partner. This study took many measures, including multiple manipulation checks, in order to assure their results were valid. In therapy, self-disclosure is a very important aspect, so this research can be used to support the efficacy of chatbots in therapy.

Kulikov, V. (2017, July 13). Can a machine ever learn to think abstractly like humans? Retrieved February 19, 2019 from https://www.helsinki.fi/en/news/data-science/can-a-machine-ever-learn-to-think-abstractly-like-humans

This article explores the idea of computers being capable of abstract thought. Although not peer-reviewed, this article cites many studies and experiments to support its claims. Additionally, the author is a postdoctoral researcher in mathematics, which suggests this source is credible. In the article, the author explains that humans have an innate understanding of some concepts, especially mathematical concepts such as real numbers and functions. The author argues that in order to build a computer capable of abstract thought and to make a "machine genuinely understand something," we must first understand the nature of abstract and consciousness in humans. This source provides evidence that despite machines' intelligence, humans have a much deeper understanding of concepts than machines. This source is necessary to answer the research question because a chatbot's ability to understand concepts is a factor of its effectiveness.

Neff, G., & Nagy, P. (2016). Automation, algorithms, and politics| Talking to bots: Symbiotic

agency and the case of tay. *International Journal of Communication Systems*, *10*(0), 4915-4931.

The authors establish what is effectively a case study on the emergence and interactions of / with agency in regards to artificial intelligence. Analyzing Microsoft's Twitter based chatbot, Tay, the study focuses on the public's perceptions of its (her) behavior and apparent personality as well as interactions between twitter users and the bot itself. The authors general argument hinges on the idea of "symbiotic agency", emphasizing the importance of the human interaction with Tay that she learned from as well as the responses that she evoked. In this case, the authors argue, the agency that Tay develops is *not* autonomous. This has important implications for our topic, as chatbots developed for therapy would likely "feed" off of their patients. It is crucial that this symbiotic relationship is monitored / well developed to prevent therapists devolving into a nonsense bot spewing hate like Tay did. The connection between users and bots is an important one that the authors conclude on. Researches / developers cannot view chatbots as purely independent tools if they wish to use them effectively and responsibly.

Smith-Miles, K., & Hemert, J. V. (2011). Discovering the suitability of optimisation algorithms by learning from evolved instances. *Annals of Mathematics and Artificial Intelligence,61*(2), 87-104. doi:10.1007/s10472-011-9230-5

This study seeks to understand the strengths and weaknesses of different algorithms in order to find the ideal conditions for an algorithm to excel. Knowledge of these conditions improves algorithm design and algorithm performance. The authors study each algorithm's strengths and weaknesses by using an "evolutionary algorithm to evolve instances that are uniquely easy or hard for each algorithm." The paper discusses the

features that characterize the ideal conditions of an algorithm and proposes criteria and methodology to assess these features. The results show that the proposed methodology can find the best performing algorithm for an instance. While this paper focuses on the performance of algorithms, it suggests that some algorithms perform better than others. This source helps answer the research question since therapy bots can only be as effective as their algorithms allow them to be.

Walicki, M., Bezem, M., & Szajnkenig, W. (2008). Developing bounded reasoning. *Journal of Logic, Language and Information,18*(1), 97-129. doi:10.1007/s10849-008-9070-9

This article examines the reasoning systems in three different features in agents. These agents consist of agents that do not have any restrictions in their reasoning systems, agents who are bounded, and agents who perform reasoning in time. The authors treat these three aspects independent of each other, and they provide a general model for each one of them. As the paper takes a logical approach, the authors go in depth about the logical sequences in systems, and the math behind the sequences. Although the paper is dense and requires the reader to have a great understanding of mathematical concepts, it demonstrates how agents are capable of having internal reasoning mechanisms.

Wan, F., & Song, C. (2018). A neural network with logical reasoning based on auxiliary inputs. *Frontiers in Robotics and AI,5*, 1-12. doi:10.3389/frobt.2018.00086

This paper aims to mimic the human behavior of deductive reasoning in artificial neural networks. The authors propose a neural network design that is able to reason logically by using auxiliary inputs as hints. The authors argue that the introduction of the proposed auxiliary input in neural networks would reduce uncertainties in data by using logical deduction. The different sets of auxiliary inputs use different ways of logical

deduction to explain the predicted result. The authors prove their concept of logical learning by testing the network with several datasets. The experiment results show that the network improves the prediction accuracy. This article uses scientific language that non-practitioners would most likely not understand. Based on mathematics, computer science, and machine learning, the concepts are technical. However, the paper shows that computers are capable of logical reasoning, and they are on their way of becoming more human-like with deductive reasoning.

Zain, J., & Lokman, A. (2010). Extension and Prerequisite: An Algorithm to Enable Relations Between Responses in Chatbot Technology. *Journal of Computer Science*, *6*(10), 1212-1218. doi:10.3844/jcssp.2010.1212.1218

This study focuses on how to improve a chatbot's responses based on the conversation as it has developed with a potential user. Often, chatbots that can be found online function similarly to a search engine, as the authors explain, producing an output based on the given piece of a whole conversation. To make chatbots more believable and personable, even human-like, the authors argue for the use of algorithms that would allow a chatbot to analyze the previous statements and its respective responses in order to give the most wholistic response in the moment. The authors propose the algorithm referred to as "Extension and Prerequisite" and develop a chatbot using this tool in conjunction with relational databases to allow for a chatbot query to produce different results tied to the flow of the conversation as a whole. The authors emphasize their point through a case study of the Virtual Diabetes Physician chatbot. With the use of their technology, ViDi can maintain a specific issue conversation with a user, allowing it to be more effective in real interactions developers can expect their bots to face.