Revised pitch one: I believe that one suitable area of research would be looking into effective means of preventing AI chatbots from being negatively impacted by bad data during training. In this case, particularly I would like to focus on excluding responses that are offensive/obscene. Such example responses are known to lead AI that uses machine learning to see similarly offensive/obscene responses as acceptable. Finding an effective and efficient way to automatically reject these data points from a training set should remedy the problem of the chat AI training in bad responses. This is especially important because many machine learning chatbots train based on the wellspring of logged conversations on the internet and from the interactions they have with users. As for potential paths to finding a solution to this problem, I plan to compare existing data from several existing studies on automatically filtering large data sets. Should there be insufficient data available I plan to utilize open source chat AIs that use machine learning and experiment with multiple approaches to filtering out bad data from the AI's training. Ideally, any data sets used in training these bots sets would already be sorted into acceptable and unacceptable responses. To elaborate on what is considered offensive or obscene any response that uses excessive profanity or any level of hate speech would be considered unacceptable.

Revised pitch two: Another worthwhile area of research would be investigating effective means to censor any undesirable responses an AI chatbot might produce. This would serve a similar purpose to the first avenue of research but rather than keeping bad data out this approach would center around keeping bad responses in. Training an AI on large data sets of logged conversations from the internet will inevitably lead to the chatbot training in some responses that are offensive/obscene. Filtering and moterating large data sets may prove to be cost prohibitive so it may instead be fruitful to instead make efforts to moderate the AI's responses. It may very well be easier to find a way to detect that a generated response is undesirable and prevent it from being displayed at all. There are a variety of techniques suitable for this end, black listing words that should never be used in any context, developing an algorithm to score the tone of a response and making sure it remains within acceptable parameters, and finding a way to look for patterns in offensive sentence structure. I suspect that the black list of offensive words could be drawn from existing lists of offensive words. This study would ideally be done by reviewing and comparing data produced by existing studies censoring obscenity produced by chatbots. Should I fail to find sufficient published data the study could be conducted using open source chat AIs that use machine learning. To elaborate on what responses are considered offensive or obscene any response that uses excessive profanity or any level of hate speech would be unacceptable.

Revised pitch three: Finally another way to study machine learning would be to try and find the most effective way to parse the data on screen into information about a task that an AI could use to complete that task. For the sake of this study, I would use an open-source implementation of Tetris as the novel task. Tetris is a suitable task for this study as it is both simple and includes a random element that will prevent the AI from simply finding an optimal series of keys to press rather than finding a method to play the game. An intuitive way to conduct this study would be to find a suitable open-source machine learning AI and refine methods to turn the current screen into workable data that the AI can understand. It would however be more effective to gather the needed data from already published methods of automatically parsing the screen into usable data. For that reason, I will seek out published studies that could provide all the needed data or at least insight into how to begin the process of parsing screen data.