

## Section V: Conclusion

Future research in this area may include exploring which variants of black box and white box AI are able to work in a team most successfully. Research into the different types of AI and how they work together will help the effort to make an interpretable and accurate AI. Future studies that build on this research will become essential to allowing for AI algorithms to be more transparent in the future. Currently, the use of AI algorithms are on the rise while little effort is being made to ensure that these algorithms are used in a responsible manner. It is currently very difficult to know if a certain AI model is biased or not. There have been rising concerns about how trustworthy Black AI is (von Eschenbach, 2021). The creation of a fully functional ensemble gray box AI would begin to address many of the issues that arise when using a black box AI in scenarios with serious consequences for bias. Bias in AI algorithms is especially dangerous because it is very easily scalable. It is very difficult for a biased interviewer to affect millions, but for an automated decision process that can exist in many places, it is scarily possible.

The tools to properly analyze and understand decisions that AI agents are making need to be developed soon. Effects of biases in AI models AI have already been seen on large scales (O'Neil, 2016). It is crucial that the tools to better understand what processes AI models take become available.

An ensemble gray box AI that is fully interpretable and is able to rival the accuracy rates of powerful black box AI would allow for the safe and ethical use of AI on a large scale. Giving the black box AI the explainability it needs to be used safely in high-stakes fields such as law

and medicine would revolutionize AI technology. This would allow biases in AI systems to be discovered more easily and a healthier and safer way AI to continue to grow.