Shallow Green

From adversarial AI and non-robust features to a witty checkers robot.

By Mees Gelein (s1117378), Jordy van Miltenburg (s2336553) and Jeroen Offerijns (s1738291)
Statement to Experience Course, Media Technology MSc. Leiden University

1. Introduction

Artificial intelligence is more and more becoming a part of our daily lives. While some fear the possibility that AI will take over the world, others believe we, as humans, will be able to stay competitive and use the same technology to enhance our own abilities (Schmidt, 2018).

Whatever side of this discussion you are on, it is hard to deny that AI has the ability to far outclass human intelligence in many tasks. The artificial mind is extraordinarily good at pattern recognition and sometimes even sees patterns that humans are not able to perceive (Ilyas, Santurkar, Tsipras, Engstrom, Tran, & Madry 2019).

In this paper we will describe how we got from this scientific insight to our statement, which ultimately led us to create Shallow Green, our 'unbeatable' checkers AI. We will try to detail our process as well as the problems we encountered along the way, explaining why we arrived at certain decisions or created specific workarounds.

2. Explaining Our Insight and Statement

Finding our insight was a lengthy process; our minds were all over the place and we considered a multitude of scientific sources, ranging from psychological explanations of the *Wisdom of the Crowd*[1] effect to the natural laws of physics. As the initial phase started to come to a close, we decided to focus on Adversarial Examples Are Not Bugs, They Are Features (Ilyas, et. al., 2019) and the insight found within.

[1] For reference this amazing wiki page (last checked on 2-12-19):
https://en.wikipedia.org/wiki/Wisdom_of_the_crowd

In the paper, the example of image recognition is used, but the idea is not limited to just the visual domain. We all know about the existence of image classifiers: neural networks that can classify the object that is displayed in a certain image. Where this becomes novel and interesting is with the introduction of so-called adversarial AI, an intelligence that is trained to disrupt normal AI. In our image classifier example, this would mean that an adversarial AI would slightly alter the input image in such a way that the image classifier makes an erroneous classification. These patterns should be as hard to detect as possible for the human eye, and this is what the adversarial network is trained on.

The features we as humans use to classify the image are dubbed ‘robust features’ by the authors, while the features the image classifier uses to (wrongly) classify the picture are called ‘non-robust features’. Now, we are arriving at the heart of our scientific insight. The authors claim that these non-robust features are not bugs, but features. The image classifier understands something about these images that we, as humans, are unable to see or understand. These altered images contain some patterns that are only visible to an AI.

It is through this final insight that we arrived at our statement: “The way AI experiences the world is the best way.” This statement is meant to be provoking, we would like people to feel challenged by the notion that an AI is superior to humans and ultimately we would love it if a human being would be able to prove us wrong.

3. Our Creative Process

During the early stages of our concept development, we soon realised that we wanted to create a form of competition between an AI and a human. Many concepts were discussed, ranging from seeing if text was written by an AI or a human to an AI finishing other people's sentences.

Before we could continue we needed to think about what we wanted to communicate through this installation, and we made up the following list of requirements:

- The installation needs to show that AI can do things that humans don’t expect.
- The installation needs to convey that AI has a superior understanding of the world.
- The installation should, at the same time, question the above notion.
- The installation needs to be eye-catching or interesting to gain more interaction.
- The interaction needs to be well known or easily teachable.

We soon realized we wanted people to play a game with the AI. This game should be easily teachable and preferably something people don't expect an AI to be able to do. We then decided to go for the physical game of checkers. The game is fairly well known around the world, but at the same time we assume that few people have played a physical game of checkers against an AI.

Thus we decided to make a robot that would play checkers. The robot arm would be the eye-catching thing we needed to gain more interaction, while at the same time fulfilling our requirement of doing something people don't expect AI to be able to do. Our implementation of a checkers AI would then need to show how its understanding of the game was far above and beyond that of any normal human. To put it bluntly: It would need to beat the crap out of any human. However, at the same time we wanted to question this same idea. Therefore we decided to go for scalable difficulty, allowing us to adjust the difficulty on the fly. One moment letting people win if they are alone and when they come back with friends, or when more people are present, really showing the full capabilities of our new baby: Shallow Green[2].

4. The Road to the Final Product

To recreate the intended experience, we needed a lot of complex components spanning different domains of knowledge and expertise and we would need to make all of these components work with each other. Luckily as a group we have a diverse set of skills that allowed everyone to contribute to most parts while also having experts on each of the components. Shallow Green is made up of the following subsystems:

1. A checkers AI (minimax) that runs on the NodeJS web server.
2. A voice-service that allows Shallow Green to speak to its opponents, runs on the same Node JS server.
3. Processing based interface to play a digital version of checkers against the AI, allowing us to test all digital subsystems without needing to be near the physical systems.
4. A custom hardware robot arm with four degrees of freedom. It picks up the pieces using an electromagnet.
5. Processing based GUI to set board coordinates and allow manual control of the robot arm to speed up development times.
6. The firmware that powers the robot arm via an Arduino Nano and receives commands over the Serial connection.

[2] No, this is not a reference to a certain well known chess AI, or is it?

7. A command service that translates AI moves to commands the Arduino can interpret and also stores board coordinates. This runs on the NodeJS server as well.
8. Python based (OpenCV and Tensorflow) board state recognition. This system parses which piece is on which cell and forwards this knowledge to the AI server.
9. Web application to control the AI strength, and toggle other settings, from our smartphones.

It is clear from this list that we did not pick an easy project for ourselves, but we did this with a reason: any technical complexity was necessary to get to our end goal of letting people play checkers against an AI in a novel, unexpected way. During development we constantly had the option of losing the arm in the back of our minds, which would have been possible if its development had been prohibitively difficult. However, we firmly believe we would have lost an integral part of our installation and experience if it would have come to that, not to mention the fact that our installation would have been far less eye-catching.

While it is tempting to write more about the technical process, the arm construction or the programming that has gone into this, we think it's best to continue with how we want Shallow Green to behave at the exhibition. For any more technical details you are always welcome to contact any of the authors, or visit our Github page for this project to see all of the code and some of the technical documentation: https://github.com/MGelein/statement_to_experience

5. The Intended Interaction

When approaching Shallow Green, we want people to be immediately drawn towards it. It should draw attention to itself both through auditory and physical cues. Preferably someone is playing a game at that moment, which would naturally draw some spectators. If no-one is playing, Shallow Green will try to get the attention of random people by calling them out or challenging them.

Playing a game against Shallow Green can be anything along a whole scale of emotional experiences. Because we, the authors, have control of how well the AI plays, we can adjust it on the fly to match the intended experience. We want people to know how little chance they actually have at winning by mercilessly beating them at checkers, but the next game we might want to give them hope and turn down the difficulty. This should hopefully spark some discussion: why does this unbeatable checkers AI sometimes let people win, and what does it mean? Is it even true that sometimes people beat it, or are they lying? Why did Shallow Green almost let me win, only to completely own me in the next few moves? All these questions are even more apt considering the theme of our exhibition. Shallow Green cannot only play checkers, it can also play mind games.

Bibliography

- Ilyas, A., Santurkar, S., Tsipras, D., Engstrom, L., Tran, B., & Madry, A. (2019). Adversarial examples are not bugs, they are features. *arXiv preprint arXiv:*1905.02175.
- Schmidt, A. (2018, September). Is artificial intelligence (AI) taking over the world?: Are humans losing out to AI in the workplace?. In *Proceedings of the 5th international Workshop on Sensor-based Activity Recognition and Interaction* (p. 1). ACM.