KLEX
Klusemann Extern
Marschallgasse 19-21

A-8020 Graz

AI

and its implications

Vorwissenschaftliche Arbeit
vorgelegt von Adriel Ondas

BetreuerIn: Manuel Menzinger

Schuljahr 2020-2022

Contents

1 Introduction

1.1 Why is this topic important?

In the last few years, AI has been the topic of many discussions, often without a very realistic view on AI.

While many see AI as these super-strong real-life robots, many do not get just how broad one can define AI in a sense of intelligent agents, and just how close it is to all of our lives, even through we are currently not using as versatile agents, as Hollywood would like us to believe.

The most obvious examples of use of AI are found around computers, more closely around social media, and almost everywhere those mysterious "Algorithms" are used, for example to help advertisers advertise [1] or to sort content [3].

1.2 Questions this thesis seeks to answer:

The Author wants to bring this topic in broader view of the public, and wants to answer the questions of:

- What AI is (in a very abstract sense).
- How AI affects society.
- What previously impossible to solve tasks AI can help to solve.
- What new problems arise.
- How we can build aligned and save AI, to help mitigate those problems.

1.3 How was this thesis made?

Using mainly content found in the book "Artificial Intelligence, an Introduction" [4], the author would like to answer these questions, using other works to back it up in very specific topics not discussed as extensively in this book.

2 What is AI?

2.1 Definition

Defining AI is quite hard, but according to [4], artificial intelligence can be defined very broadly, and as every agent acting on an environment intelligently, and rationally.

An Agent is defined as an entity being able to perceive inputs, and act according to its percept sequence.
An environment can be almost everything: e.g. a website, the real world, or something abstract like a map.
So in contrast to common belief, AI isn't just machine learning and node networks, but rather a much bigger field, containing many subfields.

2.2 Similarities with functions

One might ask why neural nets are this fitting for the task, The general challenge in this field is finding a function to map these percept sequences to actions the agent can take. A possible method for this are neural nets, which can be described as complex functions. Actually there are many more possible functions, which all are possible, but as there are very practical advantages, discussed in a further chapter, neural nets are quite interesting.

2.3 Four possible approaches to AI

There are a few different approaches to AI, but they are generally categorizable into four sectors:

- Thinking humanly
- Acting humanly
- Thinking rationally
- Acting rationally

2.3.1 Thinking humanly

According to [4, page 3], the first approach has many weaknesses, mainly not knowing how a human thinks.

2.3.2 Acting humanly

The second approach raises the question whether it actually would be desirable to model an AI to a human, as "The quest for "artificial flight" succeeded when the Wright brothers and others stopped imitating birds and started using wind tunnels and learning about aerodynamics." [4, page 3].

2.3.3 Thinking rationally

To think rationally means thinking according to the rules of logic, which, in theory is easy, but comes with the problem of defining the starting conditions to an accuracy of 100%

2.3.4 Acting rationally

3 How is it used, and what effects do those use cases have?

There are many use cases for AI, but the ones AI is used in most are very repetitive tasks with loads of data, as those are harder to grasp for humans. As with any tool, one can use AI in both malicious and well-meant applications.

3.1 Usage for malice

Applications for individual benefit of (mostly companies) include (but are not limited to) Facebook [1] and [2], both companies using AI for advertisements, and not benefiting the end-user.

3.2 Usage for the good of everyone

Other use cases through, are benefiting more of the ones using the service. The author, for example, would put google's search-algorithm [3] into this category, as it not only makes searching the web faster, but actually enables it, as without a usable search-algorithm using the web as we are now would be impossible.

4 Aligned AI

4.1 What is aligned AI

Aligned AI would be an AI having the same values as humans would have, and act accordingly. The topic of aligned AI is a very high-level one, mostly not bothering about the technical implementation.

4.2 Problems that arise when trying to implement aligned AI

But therein lies the first problem: What can we say are values, humans have? Some values we, as a species, can agree on, but in most topics there are at least two opposing opinions. To formalize valid rules and values to AI, we would first have to "solve" important philosophical questions, and as we haven't until now, we will have to live with AI's at most being aligned with a group of people.

4.3 Importance of implementing aligned AI

The importance of aligning AI cannot be stressed enough, as aligned AI is vital for the safety of any AI.

List of Figures

Bibliography

[1] facebook. Werbung auf instagram | instagram business.

[2] google. Onlinewerbung leicht gemacht – einfach mehr kunden mit google ads.

[3] google. So funktioniert die google-suche | suchalgorithmen.

[4] Stuart Russell and Peter Norvig. *Artificial Intelligence A Modern Approach*. Pearson Education, Inc., 3 edition, 2010.