

AI
and its implications

Adriel Ondas

2020-2022

Contents

1	Introduction	5
2	What is AI?	7
2.1	Definition	7
2.2	Similarities with functions	7
2.3	Four possible approaches to AI	7
2.3.1	Thinking humanly	8
2.3.2	Acting humanly	8
2.3.3	Thinking rationally	8
2.3.4	Acting rationally	8
3	What can you do with it?	9
3.1	Malice	9
3.2	Greater good	9
4	How does AI affect us?	11
5	Effects of AI	13
5.1	preface	13
6	Conclusion	15
7	Aligned AI	17
7.1	What is that?	17
7.2	Problem	17
7.3	Importance	17
	Bibliography	21

1 Introduction

Why is this topic important? In the last few years, AI has been the topic of many discussions, even though many do not really understand what Artificial Intelligence is. Even the Author himself needed to figure some things out first, as he started to write this work.

While many see AI as these super-strong real-life robots, many do not get just how broad one can define intelligent agents, and just how close it is to all of our lives, even though 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 “Algorithms” are used, for example to help advertisers advertise [1] or to sort content [3].

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...

- whether AI is just a tool
- whether you can call an intelligent agent a tool
- what previously impossible to solve tasks can be solved using AI
- what new problems arise
- what effect the use of AI has on society
- how AI should be used

Maybe the author will focus on “aligned” and “safe” AI as well.

How was this thesis made? Having read through the Introduction of the book “Artificial Intelligence, an Introduction” [4], the author would like to answer these (and more) questions using mainly this book, using other works to back it up in very specific topics not discussed as extensively in this book.

The Author would like to start by means of an overview about intelligent agents, followed by a chapter concerning the current and further use of AI in our everyday life.

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 What can you do with it?

There are many use cases for AI, but the most popular ones are repetitive tasks. As with any tool, one can use AI in both malicious and well-meant applications.

3.1 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 Greater good

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 How does AI affect us?

How does it affect

5 Effects of AI

5.1 preface

es war einmal, vor langer Zeit, da gab es was

6 Conclusion

the end.

7 Aligned AI

7.1 What is that?

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.

7.2 Problem

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.

7.3 Importance

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