

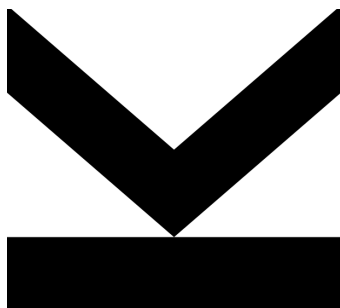
Submitted by
Viktor Maximilian Loreth
k12006268

Submitted at
Institute of
Computational
Perception

Supervisor
Katharina Hoedt, PHD

August 21, 2023

INTPRETABILITY METHODS AND THEIR EVALUATION: AN IN-DEPTH LOOK



Bachelor Thesis
to obtain the academic degree of
Bachelor of Science
in the Bachelor's Program
Artificial Intelligence

Sworn Declaration

I hereby declare under oath that the submitted Bachelor Thesis has been written solely by me without any third-party assistance, information other than provided sources or aids have not been used and those used have been fully documented. Sources for literal, paraphrased and cited quotes have been accurately credited.

The submitted document here present is identical to the electronically submitted text document.

Linz, August 21, 2023

Abstract

- Motivation zum Thema [Wieso brauchen wir das Thema? Welche Rolle spielt es in der Welt?]
- Welche Rolle spielt AI in der Welt? Welche Rolle spielen Interpretability Methoden in der Welt?
Wieso wird es erforscht?
- Worüber geht es in der Thesis [Interpretability of AI models,]
- Die Thesis fasst zusammen, was sind Interpretability Methoden, was für Arten gibt es, wie werden Sie angewandt?

Contents

1	Introduction	1
2	Fundamental theory	2
2.1	abc	2
2.1.1	abc	2
3	Project work	3

List of Figures

1 Introduction

Artificial Intelligence is playing an increasing role in our world. Automotive driving, [Various other technologies] are being developed continuously.

1.1 Structure of the thesis

1. Fundamental Theory

What is Interpretability of AI What methods do exist?

1.1. Ad-hoc methods - Worüber geht es in der Thesis [Interpretability of AI models,]

- Die Thesis fasst zusammen, was sind Interpretability Methoden, was für Arten gibt es, wie werden Sie angewandt?

Struktur erklären [Was wird in dieser Bachelorarbeit besprochen?]

2 Fundamental theory

2.1 Different types of AI-Interpretability Methods

2.2 Evaluation of Interpretation methods

2.3 RoAR

2.4 Related Work

3 Project work