Topics in the Foundations of Artificial Intelligence

A Reader

Draft from April 12, 2024.
Please don't cite or distribute without permission.
Comments welcome!

Levin Hornischer

Levin.Hornischer@lmu.de

https://levinhornischer.github.io/FoundAI/

Contents

Preface

This is the reader for the course *Advanced Topics in the Foundations of AI* given during the summer semester 2024 at *LMU Munich* as part of the *Master in Logic and Philosophy of Science*. The reader is written as the course progresses. A website for the course is found at

> https://levinhornischer.github.io/FoundAI/.

Comments   I'm happy about any comments: spotting typos, finding mistakes, pointing out confusing parts, or simply questions triggered by the material. Just send an informal email to Levin.Hornischer@lmu.de.

Content   In recent years, artificial intelligence and, in particular, machine learning made great—but also disconcerting—progress. However, their foundations are, unlike other areas of computer science, less well understood. This situation is sometimes compared to being able to build steam engines without having a theory of thermodynamics.

This seminar is about the mathematical foundation of AI. After a review of the classical theory (Computability Theory, No-Free-Lunch Theorem, Universal Approximation Theorem, etc.), we read some recent research papers to get an overview of some current approaches to the foundations of AI.

Objectives   In terms of content, the course aims to convey and overview of the foundations of AI—including both classic material and cutting-edge research. In terms of skills, the course aims to teach the ability to both mathematically and philosophically assess the different approaches to the foundations of AI.

Prerequisites   In order to appreciate the literature, the course requires basic familiarity with mathematics (calculus, linear algebra, probability theory), logic (including, ideally, computability theory), and AI (neural networks). Some papers also use more advanced concepts from topology, probability theory, or category theory, so you should also be prepared to read up on those. But they are not assumed: the seminar sessions are,

among others, meant to get clearer on these concepts. Programming skills will of course be useful, but will not be assumed.

Schedule and organization   The course is organized as a seminar. Hence, for each session, we have assigned readings, which we then discuss during the session. The reading for each week is announced in the schedule on the course's website. The readings are roughly organized by topic, forming the chapters of this reader.

Layout   These notes are informal and partially still under construction. For example, there are margin notes to convey more casual comments *This is a margin note.* that you'd rather find in a lecture but usually not in a book. Todo notes indicate, well, that something needs to be done. References are found at This is a todo note the end.

Notation   Throughout, 'iff' abbreviates 'if and only if'.

1 Introduction

to be added