Faculty of Economics and Social Sciences

of the University of Tuebingen

Master thesis

Cognitive Biases in Large Language Models: An empirical analysis of state-of-the-art models

Supervisor: Prof. Dr. Dominik Papies

Winter term 2024/25

Max Mohr
Grünwalder Straße 14
81547 Munich

max.mohr@student.uni-tuebingen.de

M.Sc. Data Science in Business and Economics
5th semester
Matriculation number: 6304784

Date of submission: January 29, 2025

Abstract

Xxx

Contents

List of Abbreviations

XXX	Xxx
XXX	Xxx

List of Tables

List of Figures

1 Introduction

Xxx

2 Theoretical background

2.1 Past studies on human behavioral effects

Humans are constantly exposed to decision making. Decisions can vary between very simple and complex ones. In studying the decision processes of humans, researchers started seeing the human species as a rational species that makes decisions based on logic and reasoning (Juárez Ramos, 2018). However, gaps in these theories such as missing information access were identified quickly. This led to the development of the bounded rationality theory by Herbert Simon simon1955behavioral. The theory suggests that humans are not always rational and that they make decisions based on the information available to them. This theory was further developed by Daniel Kahneman and Amos Tversky, who introduced the concept of cognitive biases kahneman1974judgment. Cognitive biases are systematic errors in thinking that affect the decisions and judgments that people make. It has been estimated that 70% of all decisions by humans are affected by cognitive biases (Juárez Ramos, 2018).

2.2 Leveraging large language models to simulate human behavior

A) Recent developments in large language model

B) The exposure of human behavioral patterns in the models

How could models pick up biases? A) Data (texts of humans e.g.), B) Training and Learning (RLHF)

2.3 Meta analysis techniques

3 Technical framework

Xxx

3.1 Bias selections

3.1.1 Characterizations of cognitive biases

Xxx

3.1.2 Concise and balanced sample

Hier eher auf Definitionen eingehen (nicht den Studienteil)

A) Loss Aversion

Thaler, Richard (2015), "Misbehaving"

"Problem 1. Assume yourself richer..."

B) Sunk Cost Fallacy

Arkes, H.R., and C. Blumer. 1985. The psychology of sunk cost. Organizational Behavior and Human Decision Processes 35(1): 124–140.

3.2 Model selections

Ollama for models, larger models ran on cluster

4 Methodology

Xxx

4.1 Experiments

4.1.1 Studies

Some are choice experiments, some expect a number and then compare the answers between different questioning types.

4.1.2 Scenarios

- Normal (replication of original study)

- Random values

- Explicitly prompt to behave humanlike

Also describe how the normal prompt is structured.

4.2 Response analysis

Somewhere describe what the expected output of the models should look like and what I do if it is different.

4.2.1 Replicability analysis

Original studies and compare results. Perhaps a "bias detected" number. If always 100 experiment runs, bias detected is the percentage of runs where the model acted biased. (Between 0 and 1, perhaps normalize)

4.2.2 Randomized controlled trials

RCTS on randomized values as well as explicitely prompting humanlike behavior. The controll group is the normal prompt.

4.2.3 Model comparisons

Are there any trends between newer/larger models?

5 Results

Xxx

5.1 Replicability analysis

Xxx

5.2 Randomized controlled trail on randomized values

Xxx

5.3 Randomized controlled trail on humanlike behavior

Xxx

5.4 Model comparisons

Xxx

6 Discussion and outlook

Xxx

References

Juárez Ramos, V. (2018). *Analyzing the role of cognitive biases in the decision-making process.* IGI Global.

Formal declaration

I hereby declare that I have written this thesis independently, did not use any sources or resources other than those cited and that the thesis has not been submitted as a whole or in any significant part as part of any other examination process. All information taken from other works - either verbatim or paraphrased - has been clearly indicated. The copy submitted in electronic form is identical in content to the bound copies submitted.

Munich, January 29, 2025

_____

Max Mohr