Bachelor Project

Czech Technical University in Prague

F3 Faculty of Electrical Engineering
Department of Computer Science

Bias Detection in Czech News

Tomáš Horych

Supervisor: Ing. Jan Drchal, Ph.D
Field of study: Open Informatics
Subfield: Artificial Intelligence and Computer Science
February 2022

Acknowledgements

We thank the CTU in Prague for being a very good *alma mater*.

Declaration

I declare that this work is all my own work and I have cited all sources I have used in the bibliography.

Prague, February 10, 2022

Prohlašuji, že jsem předloženou práci vypracoval samostatně, a že jsem uvedl veškerou použitou literaturu.

V Praze, 10. února 2022

Abstract

This manual shows how to use the ctuthesis LaTeX class, what are the requirements, etc.

Keywords: bias detection, transformers, text classification

Supervisor: Ing. Jan Drchal, Ph.D

?abstractname?

Tento mánuál představuje LaTeXovou třídu cuthesis, její použití, požadavky na systém atd.

Keywords: bias detekce, transformers, text classification

Contents

Figures   **Tables**

Chapter 1

Introduction

This is introduction to my thesis, motivation.

1.1 About bias

Why bias. types of bias, gender, media bias. bias in text, bias in models.

1.2 Gender bias - SOTA

mdgender = classification. stereoset (and others) = mitigating?

1.3 Media bias - SOTA

1.4 Media bias - some more in depth info?

Part I

Data

Chapter 2

Current media bias related datasets

2.1 Current datasets

2.2 BABE dataset

2.3 SUBJ dataset

2.4 MPQA dataset

2.5 WIKI npov dataset

2.6 and more

Chapter 3

Czech datasets

4

Part II

Experiments

Chapter 4

Experiments

4.1 Transformer architecture

4.2 Czech models

4.2.1 RobeCzech,Czert,FERNET

4.3 Training

RobeCzech [**?**]

4.4 Evaluation

Part III

Application to news corpora

Chapter 5

Czech news corpora

Chapter 6

Interpretability? Ethics?

Chapter 7

Conclusion

7.1 Solutions

7.2 future work