Diachronic NLP Analysis of "Therapy-Speak" in General Web Discourse: ADHD and Autism in Common Crawl

Jakob Lütkemeier

MSc Dissertation submitted in partial fulfilment of the requirements for the degree of

MSc in Applied Social Data Science

School of Social Sciences and Philosophy

Department of Political Science

Supervisor: Dr Tom Paskhalis

Course Directors: Dr Jeffrey Ziegler; Dr Tom Paskhalis

Trinity College Dublin

10 August 2026

Note: Student ID is intentionally not printed (university guidance).

Declaration

I hereby declare that this MSc Dissertation is entirely my own work and that it has not been submitted as an exercise for a degree at this or any other university.

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar.

I have also completed the Online Tutorial on avoiding plagiarism "Ready Steady Write", located at http://tcd-ie.libguides.com/plagiarism/ready-steady-write.

Signed: _____   Date: _____

Abstract

This dissertation investigates whether "therapy-speak" related to ADHD and autism has increased in general web discourse over the last decade, and how the surrounding language and framing may have shifted over time. Using Common Crawl as a large-scale web archive, I implement a reproducible pipeline that samples one crawl per year and extracts plaintext (WET) documents for trend estimation and a smaller, deeper corpus for downstream NLP analysis. The study emphasises methodological feasibility: principled sampling, robust filtering, contextual disambiguation, and domain caps to prevent large publishers from dominating the dataset.

Acknowledgements

I would like to thank my supervisor, Dr Tom Paskhalis, for guidance throughout this dissertation, and the MSc in Applied Social Data Science teaching team for methodological training and support.

Contents

List of Figures

List of Tables

1 | Introduction

Research Questions

1. How has the prevalence of ADHD- and autism-related terminology in general web discourse changed over approximately the last 10 years?

2. How has the surrounding context and framing of these terms shifted over time (e.g., clinical vs. everyday usage; self-identification; moral/emotional framing)?

2 | Literature Review

3 | Method

No-copy–paste outputs

Figures are generated by code into `../reports/figures/` and included via `\includegraphics`.
Tables are generated by code into `../reports/tables/` and included via `\input`.

4 | Results

5 | Discussion

6 | Conclusion

Appendices