



Investigating Information Pollution and Hallucinations in Children's Online Search Behavior

A Comparative Study between Google and Conversational AI Models.

Harkeerat Singh Sawhney

Abstract

The ubiquitous nature of the internet has led to a surge in online activity among young children, making it a go-to source for information. However, the reliability and accuracy of online information tends to be often questioned, particularly for children who are unaware of its veracity. Recently, there has been a much more user-friendly approach to accessing online information in the form of Conversational AI Models. This thesis project aims to comprehensively investigate and classify instances of Information and Hallucinations in children's online search behavior. The research would draw a comparative analysis between Google Search and Conversational AI models such as ChatGPT, to determine the more effective option in the context of Children's online search behavior. The study would utilize innovative tools to facilitate this comparison, including the development of modules to predict the relevance of query results. Through identifying and categorizing patterns of information pollution and hallucinations in children's online search behavior, this research aims to propose effective solutions to mitigate these issues. With the goal of improving the quality of information available to young children, this study is contributing to the field of online search behavior research.

Advisor
Prof. Landoni Monica

Advisor's approval (Prof. Landoni Monica):

Date:

Contents

1	Introduction	2
1.1	Background and Motivation	2
1.1.1	Information Pollution	2
1.1.2	Hallucinations	2
1.2	Research Questions and Objectives	2
1.3	Structure of the Thesis	2
2	Problem Statement	2
2.1	Children’s Search Behavior	2
2.2	ChatGPT vs Google Search	2
2.3	Literature Review on related work	2
3	Methodology	2
3.1	Data Collection	2
3.1.1	Google Search	2
3.1.2	ChatGPT	2
3.2	Data Preprocessing	2
3.3	Data Analysis	2
3.3.1	Similarity Analysis	2
3.3.2	Sentiment Analysis	2
3.3.3	Topic Modeling	2
4	Proposed Solution	2
4.1	Machine Learning Models	2
4.1.1	Creation of ML Models	2
4.1.2	Evaluation of ML Models	2
5	Validation	2
5.1	Implications	2
5.2	Limitations	2
5.3	Future Work	2
6	Bibliography	2

1 Introduction

1.1 Background and Motivation

1.1.1 Information Pollution

1.1.2 Hallucinations

1.2 Research Questions and Objectives

1.3 Structure of the Thesis

2 Problem Statement

2.1 Children's Search Behavior

2.2 ChatGPT vs Google Search

2.3 Literature Review on related work

3 Methodology

3.1 Data Collection

3.1.1 Google Search

3.1.2 ChatGPT

3.2 Data Preprocessing

3.3 Data Analysis

3.3.1 Similarity Analysis

3.3.2 Sentiment Analysis

3.3.3 Topic Modeling

4 Proposed Solution

4.1 Machine Learning Models

4.1.1 Creation of ML Models

4.1.2 Evaluation of ML Models

5 Validation

5.1 Implications

5.2 Limitations

5.3 Future Work

6 Bibliography