# Bachelor Project Outline:

**Name:**

Analyzing and Mitigating Information Pollution and Hallucinations in Children's Online Search Behavior: A Comparative Study between Google and ChatGPT

**Project Objective:**

The objective of this project is to investigate and classify the occurrences of information pollution and hallucinations in children's online search behavior. The project aims to analyze search results retrieved from Google APIs and compare them with responses generated from conversational AI models like ChatGPT. By doing so, the project seeks to identify and classify the information pollution and hallucination patterns in children's online search behavior and propose possible solutions to mitigate them. The project also aims to investigate whether ChatGPT yields more pertinent outcomes compared to conventional methods of browsing, such as using Google. For the comparison, tools would be developed which would make it easier to come up with results.

**Likelihood Function Implementation:**

Jaccard Similarity Coefficient:

It is a measure of similarity between two sets of data. It is calculated by dividing the size of intersection of two sets by the size of their union. It measures the similarity between the two sets or objects by comparing their common elements to their total elements. It gives out a coefficient which is between 0 and 1, where 0 means that there is no similarity while 1 means that they are identical.

**Objectives:**

Week 1-2: Literature Review and Definition of Key Concepts:

* Review Literature on information pollution and hallucinations in the context of only search behavior of children.
* Define the key concepts related to information and hallucinations.
* Going over the material provided on Information Pollution

Week 3-4: Collecting Data:

* Collect responses from conversations from ChatGPT using the tested queries as prompts.
* Collect a dataset of search results from Google API using the same children queries.
* Update the ChatGPT results from the queries manually.

Week 5-6: Data Preprocessing and Analysis:

* Clean and preprocess the collected data.
* Create an API to collect results from ChatGPT (queries same ones used for Google API)
* Analyze and compare the data from ChatGPT and search results using different metrics.

Week 7-8: Identification and Classification of Information Pollution and Hallucinations:

* Identify and classify the patterns of information pollution and hallucinations in children's online search behavior based on the data analysis.
* Propose a classification model for information pollution and hallucinations in children's online search behavior.

Week 9-10: Development of Possible Solutions along with Tools:

* Propose possible solutions to mitigate the identified patterns of information pollution and hallucinations in children's online search behavior.
* Evaluate the effectiveness of the proposed solutions.

Week 11-12: Conclusion and Report Writing:

* Summarize the findings of the project.
* Write the final report and prepare the presentation.

**Thesis Outline:**

1. Introduction:
   * Background and Motivation
     + Information Pollution
     + Hallucination
   * Research Question and objectives
     + Effects of Information Pollution and Hallucinations in Children's Online Search Behavior
     + Study between Google and ChatGPT in context of Children’s Online Search Behavior
   * Outline of the Thesis
2. Problem:
   * Identification and Description of the Problem
     + Information Pollution
     + Hallucination
     + Children’s Online Search Behavior
     + Chat GPT vs Google
   * Literature review on related work
3. Proposed Method of Comparison:
   * Explaining the Queries
   * Explanation of the Methodology
     + Explaining the Tools Developed for the collection of the Data.
     + Explaining the Comparison Tools developed
4. Data Analysis
   * Observing of the Patterns
   * Likelihood Function:
     + Jaccard Similarity Coefficient
   * Analyzation of Patterns
   * Stating the Findings
5. Solution:
   * Proposed solution to address the problem.
   * Description of the solution architecture and implementation
   * Technical details of the solution
6. Validation:
   * Summary of the main contributions
   * Implications of the solutions
   * Limitations and Future Work
7. Bibliography:
   * List of sources cited in the Thesis.