Master's Project

Bias and Fairness in Digital Archival System

Rafael Biehler

Supervised by:
Prof. Dr. Heiko Schuldt
Fynn Farouz

Abstract

Todo

Contents

Center for
Innovative Finance

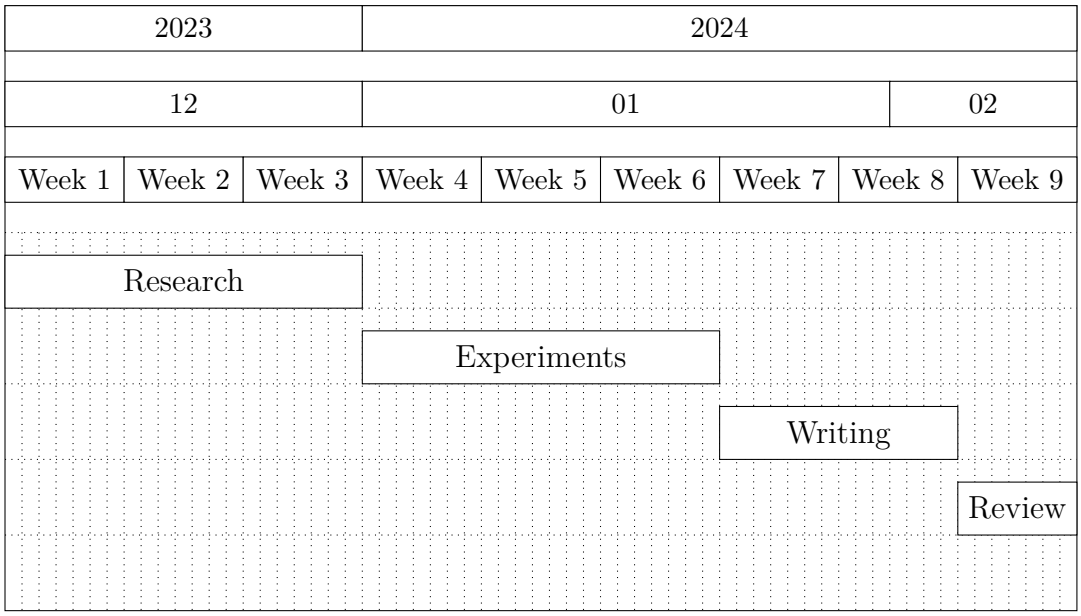
Plagiatserklärung

Ich bezeuge mit meiner Unterschrift, dass meine Angaben über die bei der Abfassung meiner Arbeit benutzten Hilfsmittel sowie über die mir zuteil gewordene Hilfe in jeder Hinsicht der Wahrheit entsprechen und vollständig sind. Ich habe das Merkblatt zu Plagiat und Betrug vom 22. Februar 2011 gelesen und bin mir der Konsequenzen eines solchen Handelns bewusst.

Rafael Biehler

1 Time Management

The following chart depicts my schedule from today (15.12.2023) to the submission date (12.02.2024). Research and Writing are spread around to each three weeks, because in this time I will be studying and working on other lectures as well. Writing and the final review process are relatively short, as I will be able to work full-time.



1.1 Milestones

1. **Research:** Be familiar with the basics of bias and fairness metrics. Know about the most recent methods of evaluating bias and fairness in similar contexts, like generative AI models. At the end of this block, I should have a clear plan of how I should structure my experiments.

2. **Experiments:** Write code to test my plan in action. Calculate numbers which could indiciate performance. Run more tests to challenge my resulsts, like a trial against randomly generated content.

3. **Writing:** Write the report for the search.

4. **Review:** Review the report together with Fynn to make sure it is acceptable for handing in.

2 Introduction

In the context of Galleries, Libraries, Archives, and Museums (GLAM), a prevalent challenge is the imbalance between the volume of work and the available staffing. This often leads to a substantial amount of image data being left unlabeled or uncategorized. To mitigate this issue, there is increasing interest in the use of multimedia retrieval systems that enable querying image data by using natural language descriptions.

Feature Extraction Models are used to represent image and text data in a common format. This process may introduce biases and fairness issues, which can impact the results in unintended and often undesired ways. It is not yet clear how to best detect unfairness issues for classification tasks.

We will introduce a new method that tries to be able to accurately detect Bias and Fairness issues. We will be working with a GLAM dataset that contains labeled photographs, which we will use to accurately determine potential biases against First Nations people.

3 Background

Todo

4 Methods

Todo

5 Discussion

Todo

References