A Context-Aware System for Bias Identification in Job Advertisements using Natural Language Processing (NLP)

Vogt Luise
Guan Zhize
Nasev Veselin
Guo Yonghui

Agenda

- Introduction
  - Context and Motivation
  - Problem Statement and Research Questions
  - Related Work
- Concepts and Approach
- Deliverables
- Planning
  - Milestones
  - Gantt Chart
- Risk Analysis and contingency plans

Context and Motivation

Masculice Coded Language:
Ethnic / Racial Bias

Masculine Coded Language:
Gender Bias

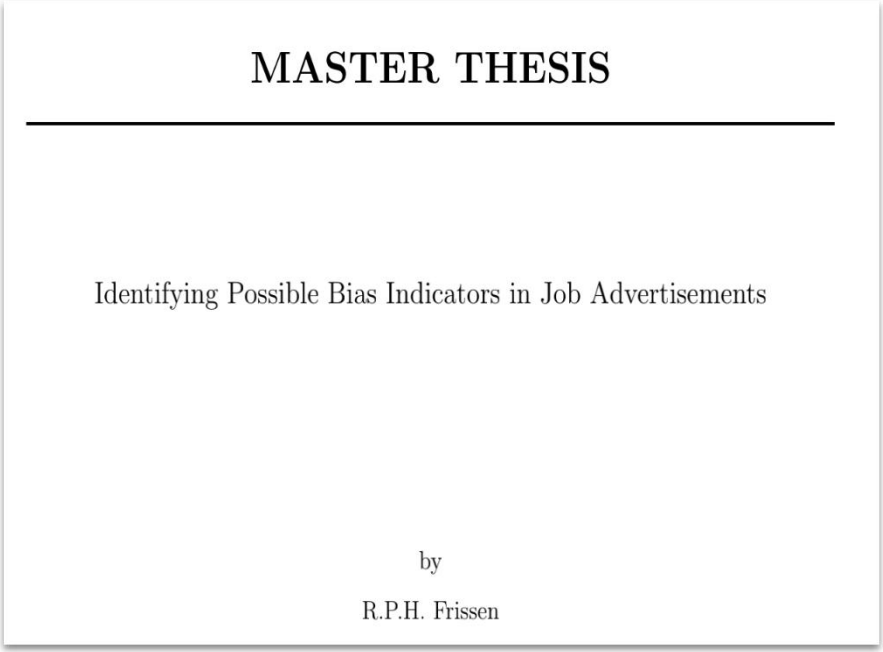
We are looking for a young and driven candidate who can bring innovation to our organization, and is a true team player for the rest within the organization. Are you the master of technology and passionate person we are looking for? We are still only 1% done at Facebook – this team is inventing every day and it takes tenacity, bravery and the ability to see the big opportunities to thrive.

"Independent"
is a neutral word that you can use instead of "bavery"

Problem Statement and Research Questions

1. What kinds of biased language are commonly identified in job advertisements?
2. What words are related to the most common biases?
3. How can a context-aware natural language processing tool be used to classify different types of biases in job advertisements?
4. What are key predictors that explain the prediction for a particular class of bias?
5. How can a NLP machine learning model be used to establish a suggestion system to replace biased language?

Related Work



MASTER THESIS

Identifying Possible Bias Indicators in Job Advertisements

by

R.P.H. Frissen

Figure 1: Thesis cover of Identifying Possible Indicators in Job Advertisments ([1])

Five Aspects to Improve

1. Suggestion system
2. Context-aware model
3. Better quality of annotations
4. The taxonomy of bias types
5. Website to demonstrate our system

Concepts

The Taxonomy of Gender Bias

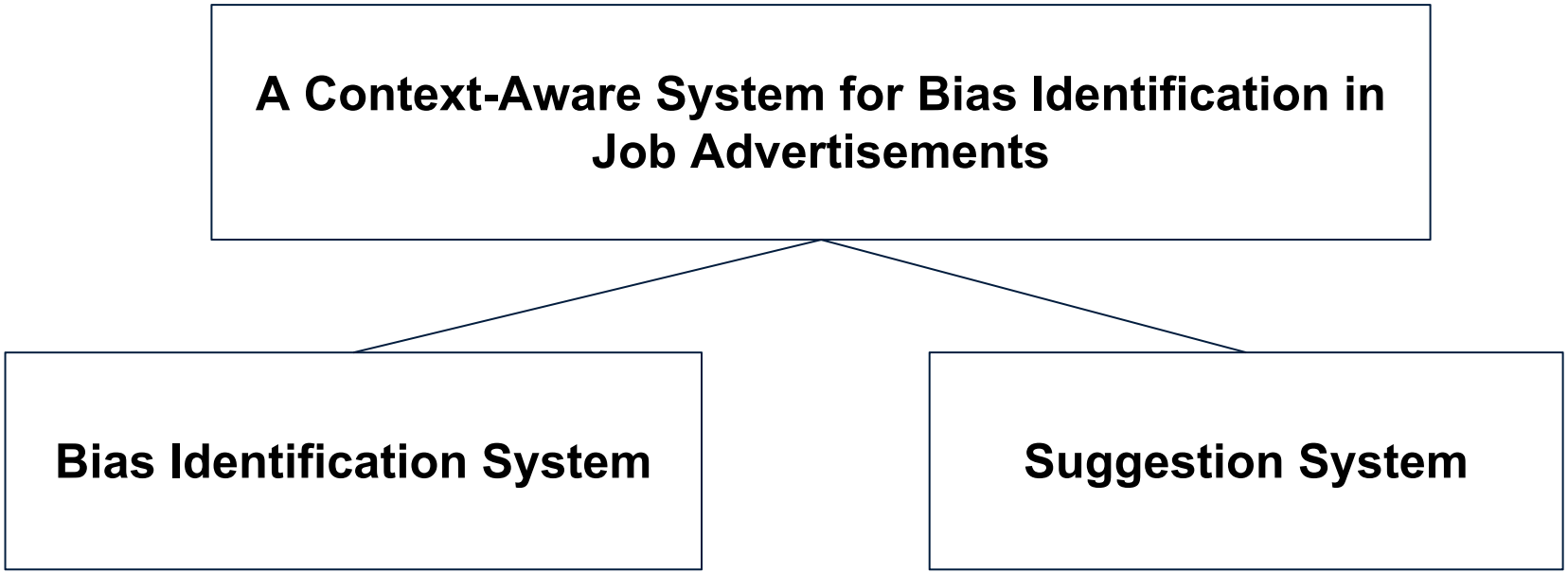
- Five parent bias types
  - Generic Pronouns
  - Stereotyping Bias
  - Sexism
  - Exclusionary Terms
  - Semantic Bias

- Two bias subtypes each

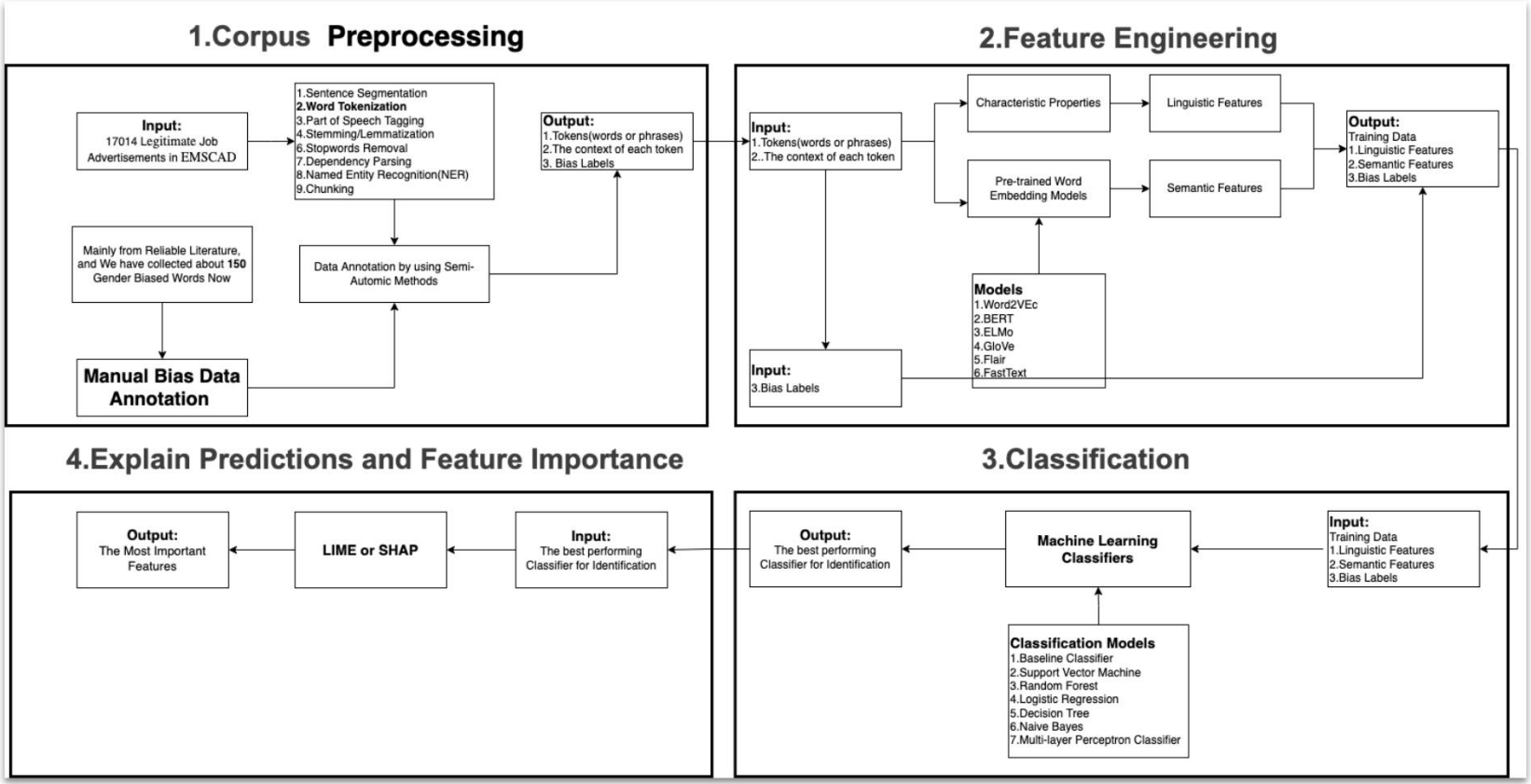
Bias Type	Bias Subtype	Example	Methodology
Generic Pronouns	Generic He	A programmer must carry his laptop with him to work.	Supervised Learning
	Generic She	A nurse should ensure that she gets adequate rest.	Supervised Learning
Stereotyping Bias	Societal Stereotypes	Senators need their wives to support them throughout their campaign.	Supervised Learning
	Behavioural Stereotypes	The event was kid-friendly for all the mothers working in the company.	Supervised Learning
Sexism	Hostile Sexism	Women are incompetent at work.	Supervised Learning
	Benevolent Sexism	They're probably surprised at how smart you are, for a girl.	Supervised Learning
Exclusionary Terms	Explicit Marking of Sex	Chairman, Businessman, Manpower, Cameraman	Lexicon-Based
	Gendered Neologisms	Man-bread, Man-sip, Man-tini	Lexicon-Based
Semantic Bias	Metaphors	"Cookie": lovely woman.	Supervised Learning
	Old Sayings	A woman's tongue three inches long can kill a man six feet high.	Supervised Learning

Figure 2: Overview of the taxonomy and link to detection methodology ([2])

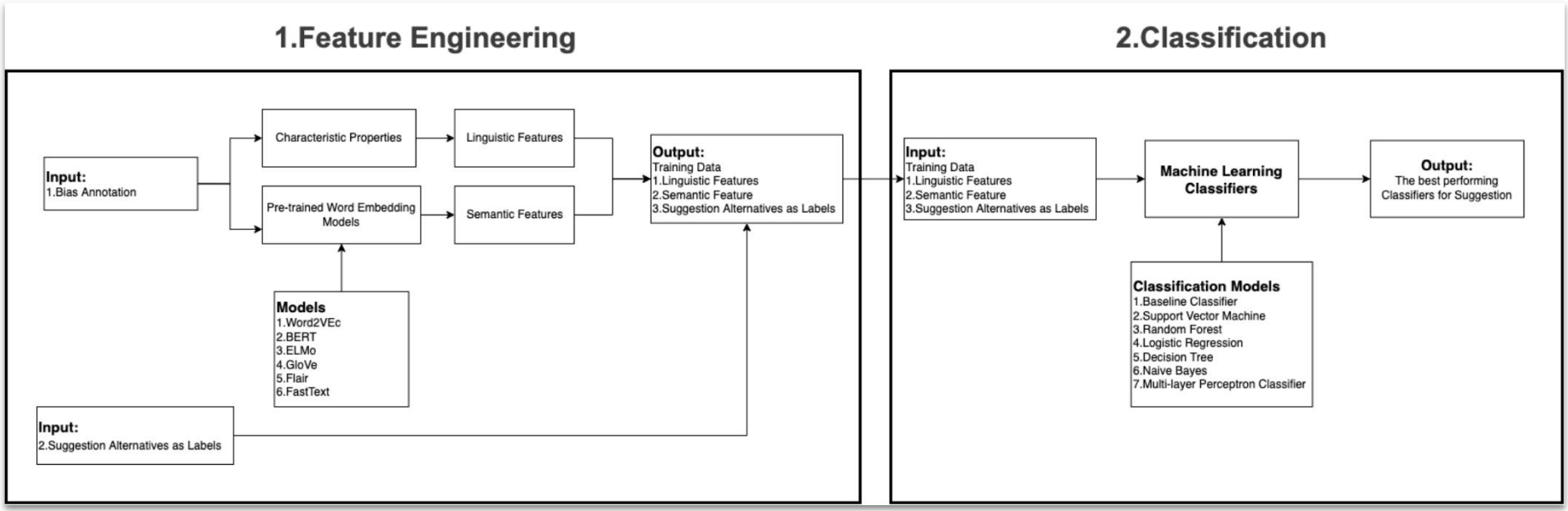
Approach


A Context-Aware System for Bias Identification in Job Advertisements
Bias Identification System
Suggestion System

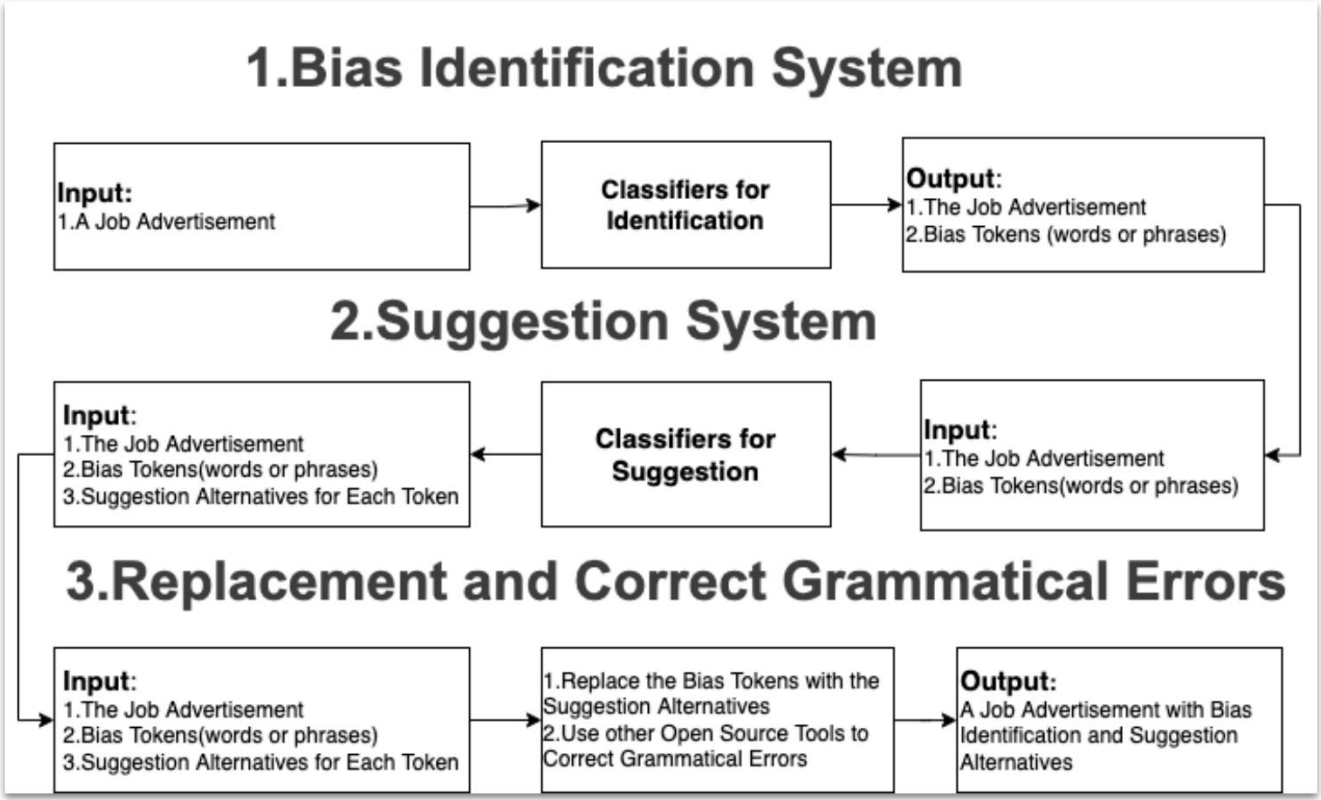
Approach--Bias Identification System



Approach--Suggestion System



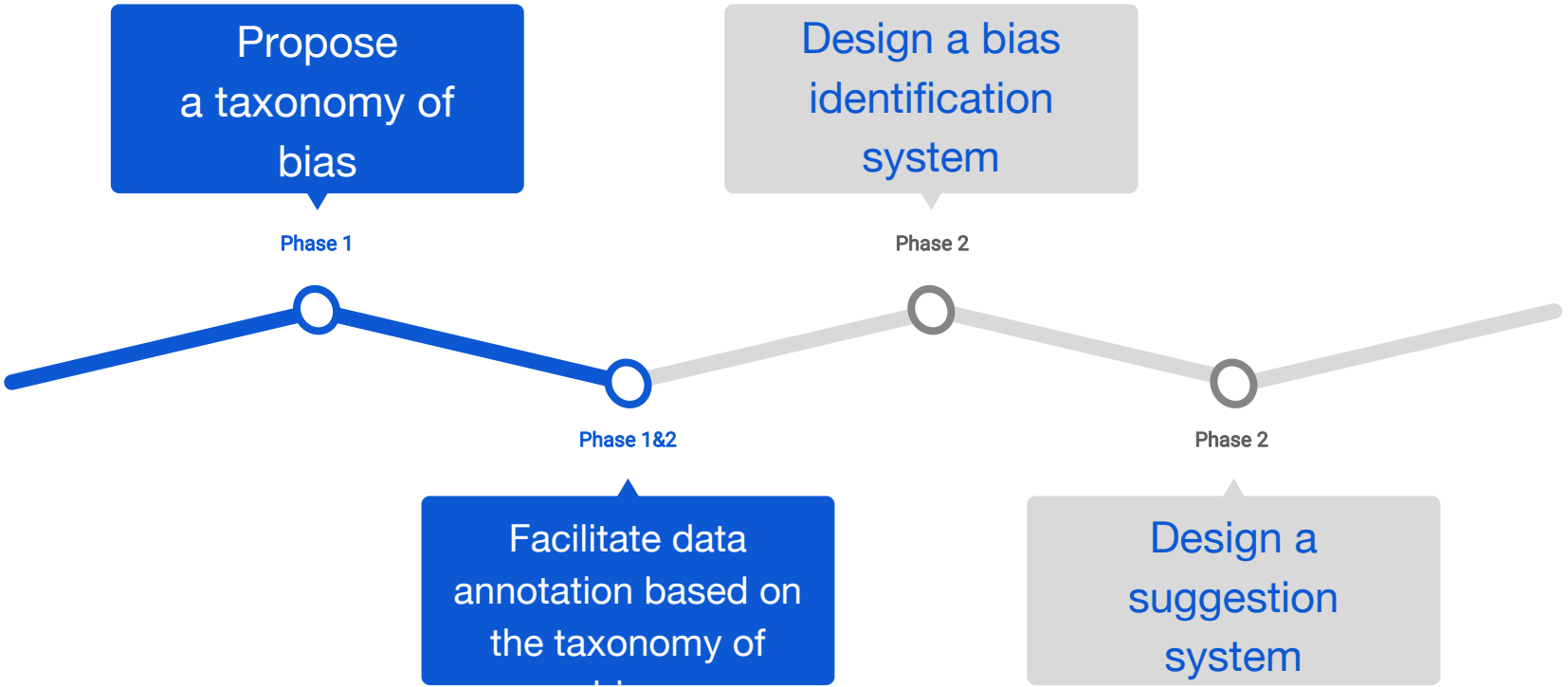
Approach--A Context-Aware System for Bias Identification



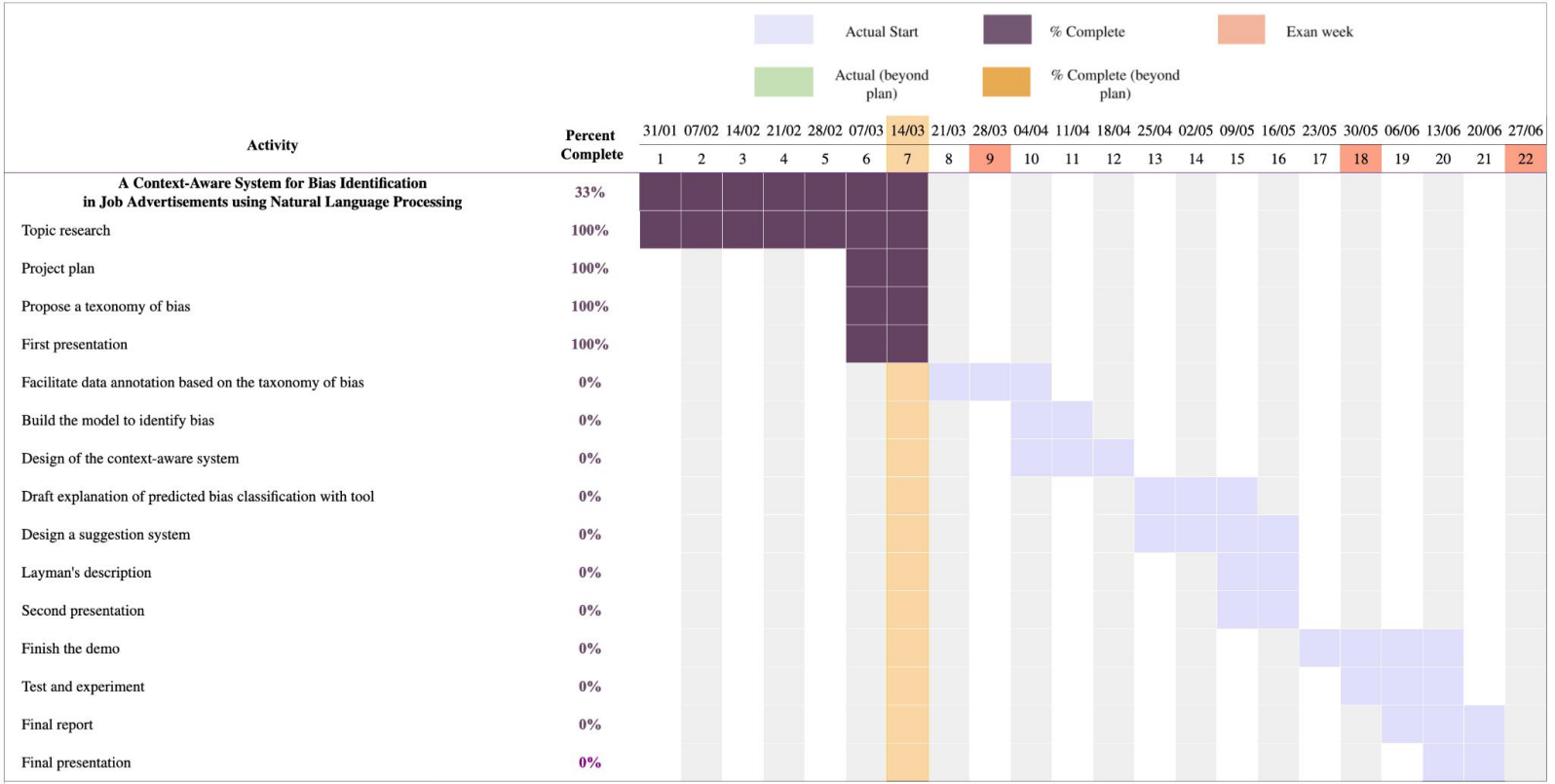
Deliverables

- A demo
  - Presents a bias identification system
  - Suggests unbiased alternative words
- A report
  - Explains the design of the demo
  - Provodes solutions for our research questions

Planning – Milestones



Planning – Gantt Chart



Planning – Phase 1

Activity	Percent Complete	31/01	07/02	14/02	21/02	28/02	07/03	14/03
		1	2	3	4	5	6	7
A Context-Aware System for Bias Identification in Job Advertisements using Natural Language Processing	33%							
Topic research	100%							
Project plan	100%							
Propose a taxonomy of bias	100%							
First presentation	100%							

- Taxonomy of bias
  - Prodives an easy way to understand, organize and show the category of bias

Planning – Phase 2

Activity	Percent Complete	21/03 8	28/03 9	04/04 10	11/04 11	18/04 12	25/04 13	02/05 14	09/05 15	16/05 16	23/05 17	30/05 18
A Context-Aware System for Bias Identification in Job Advertisements using Natural Language Processing	33%											
Facilitate data annotation based on the taxonomy of bias	0%											
Build the model to identify bias	0%											
Design of the context-aware system	0%											
Draft explanation of predicted bias classification with tool	0%											
Design a suggestion system	0%											
Layman's description	0%											
Second presentation	0%											

- Facilitate data annotation based on the taxonomy of bias
  - Improve the existing annotation
- Design the bias identification system
- Design a suggestion system for the biased language

Planning – Phase 3

Activity	Percent Complete	23/05	30/05	06/06	13/06	20/06	27/06
		17	18	19	20	21	22
A Context-Aware System for Bias Identification in Job Advertisements using Natural Language Processing	33%						
Finish the demo	0%						
Test and experiment	0%						
Final report	0%						
Final presentation	0%						

- Finish the demo
- Test and experiment

Risk Analysis and contingency plans

- Unavailable of a team member
  - Keep active communication
  - Assign tasks to each mumber
  - Keep track of individual progress

- Delay caused by disagreement over project direction and methodology
  - Set regular internal meetings
  - Discuss with supervior

- Unfamiliar with the topic
  - Do more research
  - Self-study

References

[1] R.Frissen.A-machine-learning-approach-to-recognize-bias-and-discrimination-in-job-advertisements, 2021
[2] J.Doughman and W.Khreich.Gender bias in text:Labeled datasets and lexicons.CoRR, abs/2201.08675, 2022

Thank you for your attention