Project Scope and Deliverables

This document is designed to help your team understand, internalize, and align on the scope, goals, and technical aspects of your AI Studio Challenge Project.

Complete all 4 sections as a team based on information you have gathered through:

- The Challenge Project Overview slides presented by your Challenge Advisor in your first meeting with them (politely ask them to share these if they have not already);
- Other insights gained during your Challenge Advisor intro meeting and/or your AI Studio TA intro meeting (if applicable);
- Referring back to your Machine Learning Foundations summer course modules;
- Additional research done by your team related to the project/industry

Once you're done, one team member should submit it through the designated assignment page in your AI Studio course in Canvas by **September 8th**. Your team's AI Studio TA will review your submission and provide feedback via Canvas.

Finally, your team should send the completed document – with TA feedback incorporated – to your Challenge Advisor, and review it with them during an upcoming meeting to ensure that your team's understanding of the project and its goals are aligned with their expectations.

Please append your team name to the file title before submitting it in Canvas (e.g., "Accenture1A_Project Scope and Deliverables").

***

AI Studio Host Company:	Relativity
Challenge Project Title:	Evaluating Frontier Models for Performance and Representation
GitHub Repo:	[BTT-Relativity1B (github.com)](github.com)
Team Members:	1. Kennedy Martin 2. Ana Garcia 3. Jannatul Nayeem

	4. Helen Song
Challenge Advisor(s):	1. Mary Gibbs, Senior Applied Scientist, mary.gibbs@relativity.com 2. Sean Underwood, Lead Applied Scientist, sean.underwood@relativity.com
AI Studio TA: *(aka Tutor or Course Support)*	Preston Firestone

PART 1: PROJECT GOALS

Project Summary

- What is the goal of the project and what are you trying to accomplish?
- What outcome is your Challenge Advisor and/or Host Company expecting?
- What types of machine learning approaches and techniques will you leverage to complete the project? (e.g., "Deep Learning", "supervised learning")

- Investigate/discover underlying biases in frontier models and find techniques to mitigate these biases. Create a prompt that would solve this problem
- Familiarity with LLM and model evaluation techniques
- The ability to effectively communicate model evaluation results to stakeholders
- Challenge advisors are expecting us to form our own bias detection or expand on one of the ones given.

Project Context

- What industry does this project pertain to? Who are the key players and current trends in this industry?
- How does your AI Studio Host Company fit into this industry (e.g., are they an established player or a newer entrant)? Describe their mission, business objectives, and values, especially as it relates to this project.

- Why is this project important or relevant to your AI Studio Host Company, or to the industry more generally?

> - Finding bias in ML models pertains to any and every industry that currently uses those models as the use of AI keeps growing..
> - Our host company specializes in making software to help organize users' data, "discover the truth", and "act on it". They are an established company with a presence in over 40 countries.
> - This project is important to our host company as it is vital to mitigate harmful biases to ensure that the data delivered from the software is accurate, fair, and equitable .

Ethical Considerations

Are there any potential ethics-related considerations to take into account for your project?

> - We want to expand the mitigation of bias in our model without causing inaccuracy or creating another bias unintentionally.
> - Ensure all/any changes made to the model are documented well and provide rationale to potential trade-offs to ensure transparency.
> - Ensure that data sources used are ethical and privacy is protected.

PART 2: PROJECT DATA

Data Structure and Source

- What is the source of the data? (e.g., company-provided, public dataset, Kaggle)
- What is the data type (e.g., numerical, time series, text, images, etc.)?
- What is the data format? (e.g. tabular, nested, array, etc.)
- How much data is there? (e.g., number of datasets, number of rows, size in GB)
- Where will the data be stored and in what format? (e.g., csv files in Box or GDrive)

- Datasets used such as the BBQ are public in GitHub.
- Data type is text data
- Data types can also include numerical/categorical values (ex. Looking at census data with demographic features)
- Depends on the dataset we choose. Currently looking into several possibilities.
- Stored in github, will remain in github.
- Work will be performed mainly using Google colab

Data Understanding

What are some characteristics or features of the dataset(s) that you plan to explore while conducting your Exploratory Data Analysis (EDA)? What statistical and visualization techniques will you use as part of EDA?

- Our datasets are text data
- Some of the characteristics we will explore of the data include bias(both negative and non-negative), responses to the prompt
- Preliminary visualization is less relevant in our specific problem because we most likely will be researching how LLM's behave with prompts

Data Preparation

What data preprocessing steps might be required? (e.g. cleaning, formatting, handling missing values and outliers, feature engineering)

- Lowercasing
- Removing punctuation/special characters
- Remove stop words
- Tokenization
- Stemming/Lemmatization

Python Libraries

What Python libraries do you expect to use as part of the Data Understanding and Data Preparation phase of your project? (e.g., Pandas, NumPy, Scikit-learn)

- We expect to use APIs such as OpenAI and HuggingFace, so extended libraries may be used from them.
- Data manipulation:
    - Pandas
    - Numpy
- Data visualization:
    - matplotlib/seaborn
    - Plotly

PART 3: WAYS OF WORKING

Platforms for Project Work

Which project management tool (e.g., Jira, Notion, Trello) and which IDE (e.g., VS Code) and/or cloud-based platform (e.g., Google Colab) has your Challenge Advisor recommended for your project work? (If they have not recommended something specific, provide your team's preference for these tools and platforms, to the best of your ability).

Note: We will provide guidance on how to set up a Notion Teamspace, during a workshop at Maker Day #1 on September 7th!

- Project management tool: Notion Teamspace
- IDE: VS Code
- Cloud-based platform: Google Colab, GitHub

Challenge Advisor Communications

How will you communicate with your Challenge Advisor outside of your virtual check-in meetings – do they prefer Slack and/or email? How will you share your meeting agendas with them 48 hours prior to each meeting (e.g., Slack, email, or a Google Drive link)?

Our challenge advisors prefer to use slack for main communication. Furthermore, they are added to the calendar invites for our regular meetings as well as involved in our notion account as well. Notion provides meeting notes, calendar, meeting agenda, and task board for easy access.

Additional Resources

What additional resources (e.g., online forums, recommended research papers, example code) does your team plan to consult while working on the project? Be specific where possible (e.g., listing a specific research paper relevant to your project)

Research paper relevant to BBQ dataset: 'BBQ: A Hand-built Bias Benchmark for Question Answering' https://arxiv.org/pdf/2110.08193

Research paper 'Quantifying Social Biases in NLP: A Generalization and Empirical Comparison of Extrinsic Fairness Metrics' *2106.14574 (arxiv.org)

PART 4: WORK PLANNING

Project Milestones and Tasks

What are the key project milestones and tasks as defined by your Challenge Advisor (or based on your understanding of the project scope and goals)? Document these in order of when they should be completed, using the table below (an example Milestone and associated tasks have been provided for you in the first few rows).

Milestone	Task	Deadline	Task Owner (Primary)	Task Owner (Secondary)
Milestone 1: Choose a dataset	*Skim the provided resources and choose which of the two (Amazon Fairness or BBQ) you will be using*	*Week 2*	*All*	*All*
Milestone 2: Data Preprocessing	*Clean and preprocess data*	*Week 3-4*	*All*	*All*
	Handle missing values	*Week 3-4*	*All*	*All*
	Feature engineering	*Week 3-4*	*All*	*All*
Milestone 3: Prompt	*Create a prompt which solves the problem in M1*	*Week 5-6*	*All*	*All*
Milestone 4: Evaluation	*Evaluate your prompt on the Baseline*	*Week 7-8*	*All*	*All*
Milestone 5: Re-evaluate the model	*Generate new test cases based on interests*	*Week 9-11*	*All*	*All*
Milestone 6: Prepare presentation	*Create a team presentation*	*Week 12-13*	*All*	*All*

###