

## Scenario

Machine learning and artificial intelligence has gained notoriety with public missteps for bias, racially insensitive outcomes or over-hyped potential. For example, credit algorithms have been found to extend more credit to males or that natural language models associate Muslims with violent terms. In this case you are asked to review the text of all incidents, understanding frequent terms, associations of interest or other relevant findings such as topic modeling. The case is nuanced as it is applying machine learning techniques to descriptions of machine learning missteps. One on hand, the findings may inform practitioners for common “issues” or help regulators spot trends with misbehaving algorithms.

## Contextual Information:

The database is a curated database of 1200+ articles covering a number of specific incidents. It is robust though not likely complete. Multiple industries, algrotihm types, and applications are covered. For example, resume and text processing issues are contained in the database alongside scheduling conflicts causing humanistic hardships for low-wage workers.

## Technical Considerations:

* Provide some insights such as frequent words, emerging topics by date, channel, or the media channel such as [www.businessinsider.com](http://www.businessinsider.com)
* You may use a subset of the data provided based on your analysis and/or compute constraints. If so, you must show the manner the data was reduced in a script and describe why in your presentation.

## Non-Technical:

* Describe the preprocessing steps and why are they are applied to the documents
* Describe the various techniques used to create the visuals and findings in a PowerPoint to a non-technical user

## Project Deliverables include

1. R scripts for data processing & exploration “lastName\_TM\_AIID\_case.R”
   1. Your script(s) must account for all aspects of the material in your presentation to ensure the presentation is data driven (no cheating with Excel or other tools!)
2. PowerPoint of any visualizations, findings and descriptions of non-technical results as if presented to the thinktank’s leader. “lastName\_TM\_AIID\_case\_.pptx”
   1. The PowerPoint must be accompanied by a voice over embedded in the file, or screenshare video uploaded

## Example Data

The incident date table

Table

Description automatically generated

The incident summary text table

Graphical user interface, table

Description automatically generated

## Citation Information:

McGregor, S. (2021) Preventing Repeated Real World AI Failures by Cataloging Incidents: The AI Incident Database. In Proceedings of the Thirty-Third Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-21). Virtual Conference.

Original Data: <https://incidentdatabase.ai/>

## Criteria for Success

The case material will be evaluated according to the following criteria. Each is worth 5pts for a total of 20pts.

## **Organization of content**– Logical ordering of ideas, modeling artifacts, applicable visualizations in slides

## **Organization of code**- R Code is well organized, concise, and free from error

## **Text mining process** – Recognize the type of data mining problem, adherence to established main data and text mining steps.

## **Completeness** – Understood impact, and mined the data for relevant insights/recommendations

PLAN

* OBJECTIVE GOALS
  + DISCUSS DATA
    - WHAT IT IS
    - EXAMPLE ARTICLE
  + HYPOTHESIS / ASSUMPTIONS
* DATA PIPELINE
  + CUT UP DATA BY TIME PERIOD
  + PRUNING?
* ANALYSIS
  + PERFORM PCA/WORD COUNTS BY TIME PERIOD
  + SENTIMENT ANALYSIS CLUSTERING BY TOPIC CLUSTER