# Take-Home Assignment

Aurora Articles is a specialist online blogging platform who owns the rights to all of the articles published on their platform. They have published their entire catalogue on Human Native which has been licensed by an AI company called Blog Bot. While using the catalogue to train their most recent blog generative model Blog Bot has noticed some of the articles contain personally identifiable information. They would like to report this to Human Native AI so that it can be resolved but there is not currently a process.

The team at Human Native AI have just found out about the problem from Blog Bot and legal, operations, research and engineering teams have come together to work out a solution.

You have all agreed to build a new service that allows licensees to report data they expect is in violation of local laws or regulations. The service will allow consumers to provide structured information about where, why and how the data is in violation. As a reminder Human Native AI supports all types of multimedia (text, image, audio, video, animation).

Human Native currently has a team of people in operations that review the manually submitted violations and mark them as valid or invalid. They receive nearly 100,000 violations a month and this number is growing rapidly. The engineering team is looking for feedback on their data model about if it makes this machine learning task easy or hard, but have decided to store flagged data as follows:

```
Dataset(org_id, id, name, type)
Data(dataset_id, id, value, flag)
```

As part of the machine learning research team you have been tasked to build a model that can automatically flag data that might be in violation.

**Assumptions**

- You can pick one of text, image, audio, video to deal with
- You can assume you don't need to worry about persistence
- You can assume you will need to generate some fake data to train and test the model
- You can assume your model has access to all previously flagged data items as well as those verified as correct by the operations team

**Requirements**

- You can use any language or framework you're comfortable using. With the exception of proprietary languages/tools such as Matlab and Wolfram
- Your solution should be tested and/or outline how you would assess its performance

**Submission**

- You should submit your solution either by sharing a git repository or archive with the person who sent you this assignment
- You should include a README file describing how to run your solution as well as describing any assumptions you made as part of the solution