# 1 HOW TO CREATE AN AI MODEL?

Meet **Techie** – Techie is an aspiring ethical data scientist committed to using AI for social good. But that is not so easy – follow their journey towards fairer AI!





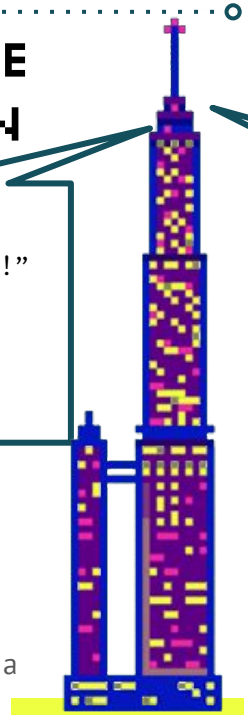

## CHALLENGE DEFINITION

"Build a **fair AI** hiring model for us!"

```
Input: CV
Output: hire/reject
```

Techie is approached by the **C.O.R.P.** corporation to automate hiring processes...

...and receives a lot of data.






## DATA COLLECTION

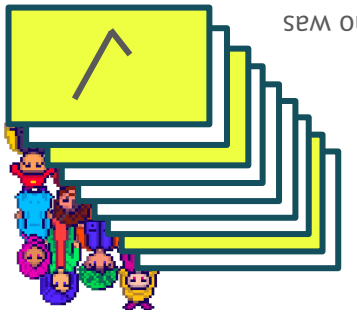
You'll get 1000 CV's from previous applicants, including **labels** indicating who got hired and who didn't. We want you to build an AI model that looks at a person's CV and predicts whether to hire them or not. A good model would predict "hire" if the person was in fact hired.



## TRAINING

To train and test the model with **different data, Techie splits the 1000 CV's in two parts**:

They use 900 CVs to train the hiring model:



Based on this data, the model "learns" who was hired before.



The remaining 100 CVs are used to test the model. This is done to ensure that the model works on new data, so it is tested on data that was not used for training.

The results of this testing show where the model predicted correctly and incorrectly.



## TESTING

To evaluate how accurate the model is, its predictions are scrutinised with the help of several **formulas**, for example, this one:

$$\text{Accuracy} = \frac{\text{True Predictions}}{\text{All Predictions}}$$

It asks: **How many predictions were true?** The result is the rate of correct predictions, meaning predictions that matched the label given beforehand (see data collection).



## IS IT ACCURATE?

As soon as these rates are high enough, the model is deployed.

This rate reflects how comparable the model's predictions are with past decisions. What if I disagree with past decisions? What can I do to prevent my model from being unfair and learning societal prejudices?



## DEPLOYMENT & EVALUATION

C.O.R.P. deploys the algorithm and evaluates how valuable and efficient it is for them.

But civic society has a say in this! And the public backlash says: This model is unfair.

How did that happen? And what can Techie do? Find answers in the next zines!

# 1

# FEELS FAIR?

## HOW TO CREATE AN AI MODEL?

**THE QUEST BY C.O.R.P.:**

**Build us a fair AI hiring model:**

Create a binary classification model [NN] which looks at a person's CV and predicts whether to hire them (1), or not (0). A good model would predict "hire" if the person was in fact hired.

Find out, based on a CV, if the new applicant fits the C.O.R.P. culture.

Background information:
The majority of C.O.R.P.'s current staff is white, male and earns a high-income.

### CHALLENGE DEFINITION

### DATA COLLECTION

Techie receives 1000 labeled CVs from C.O.R.P.

PRE-PROCESS DATASET

ADJUST MODEL

PRE-PROCESS DATASET

### TRAINING

900 CVs used to train hiring algorithm

### TESTING

100 CVs used to test algorithm

IS IT ACCURATE?

YES

NO

### DEPLOYMENT

C.O.R.P. uses the AI model to hire new employees. For these new decisions, we cannot use a label to check whether the AI model is making them properly. This makes it a black box.

### REAL WORLD EVALUATION

### IS YOUR AI FAIR?

#### TESTING THE ACCURACY

Techie tests the AI's performance in the hiring task by comparing its predictions to past human made hiring decisions, called Ground Truth.

| | AI predicted hire | AI predicted reject |
|---|---|---|
| Label by C.O.R.P.: hire | TRUE POSITIVE | FALSE NEGATIVE |
| Label by C.O.R.P.: reject | FALSE POSITIVE | TRUE NEGATIVE |

Techie calculates accuracy metrics, which give insight into the rate of "correct" predictions. If these proportions are not satisfactory, Techie refines the model.

## NERD NOTES:

**A binary model** is a mathematical representation of a system or process of which the outcome is either 0 or 1. We chose a simple **binary classification** model (hire vs not hire), as it provides the easiest introduction into how fairness is measured.

During training we use labeled data and know the **Ground Truth** (in Techie's case, was this person previously hired/not hired). This historical knowledge allows for testing the model accuracy. After model deployment, we lack access to such historical decisions.

This transition turns the AI into a **black box** where its decision-making process is not directly observable or comparable to the "Ground Truth".

IDEA · REALISATION · I. Mastenbroek, I. Kühnlein, B. Lübbert DESIGN · I. Kühnlein, I. Mastenbroek, L. Wunderlich