Hungry Hornets: Checkpoint 5

We built a natural language processing model that allows users to ask questions directly to it, to retrieve information on officers who have drug and alcohol abuse and medical allegations against them.

We asked the model these questions for every single officer:

1. What is the race of [X]?
2. What is the gender of [X]?
3. What is the birth year of [X]?
4. what is the allegation count of [X]?
5. What is the sustained count of [X]?
6. What is current salary of [X]?

Here is the accuracy of each question:

Gender	Race	Birth Year	Allegation Count	Sustained Count	Current Salary
86%	73.33%	68.89%	71.11%	68.89%	66.67%

We felt this was an important task as it could allow this type of information to be readily accessible to non-technical people in the public.

Due to constraints in computational power with the model itself, we limited our analysis to officers with more than 2 allegations of drug and alcohol abuse, of whom have had at least 1 sustained allegation and have a numerical value listed for their current salary.

For the questions that were answered incorrectly, the model correctly identified the column for which to find the answer, but provided an answer from an incorrect row (i.e. from a different officer). This could be occurring due to clashes when retrieving the same value, for example when two officers have the same birth year or race.

Going forward, an approach such as this would be markedly improved by being able to ask more complicated questions to the model, such as questions that combine different rows and columns or require some form of computation such as an average. At present, this is not possible with an acceptable degree of accuracy.