# Probabilistic Deep Learning

## for Fast Radio Bursts Project Proposal

遠馬印驤

# What is the Problem?

We wish to tackle two issues:

1. There exists a large dataset of classified FRBs done by the a community however there hasn't been any real statistics done to quantify the performance.

2. We want to eventually build an automatic detection process where we can reliably rely on the system to classify potential FRB candidates.

# But Why Deep Learning?

Not all problems are good fits for deep learning. So what does DL bring to the table?

Deep learning has always presented itself as a solution to replace simple image recognition tasks and has a track record in mimicking and at times out performing human performance.

We expect the flow of data to eventually one day overwhelm the manual efforts in verifying FRB candidates.
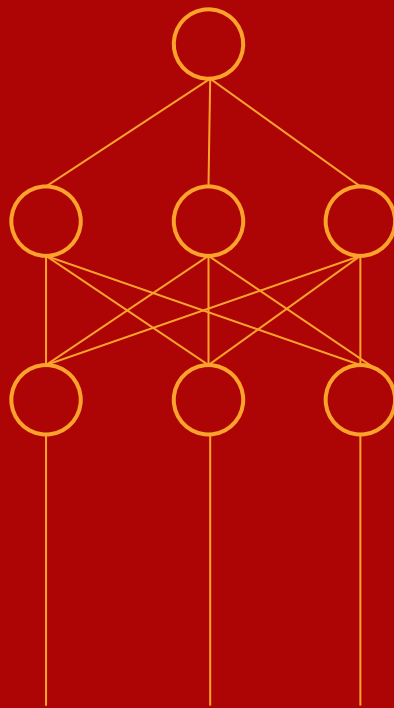
# Why are Probabilistic/Bayesian models powerful?

Bayesian/probabilistic models are designed to learn distributions rather than "point" weights. In a typical neural network, the feedforward is completely deterministic. In a Bayesian neural network the outputs are stochastic. This means that the outputs are samples from distributions rather than discrete values thus the model can output both a classification AND an uncertainty associated to the classification.

Furthermore, Bayesian neural nets are more robust and are designed to prevent overfitting especially with small data samples. This presents itself as an interesting detection method to search for transients where data could be disproportionately filled with negative samples in respect to positive ones, or vice versa.

Paper Reference **https://arxiv.org/pdf/1806.05978.pdf**
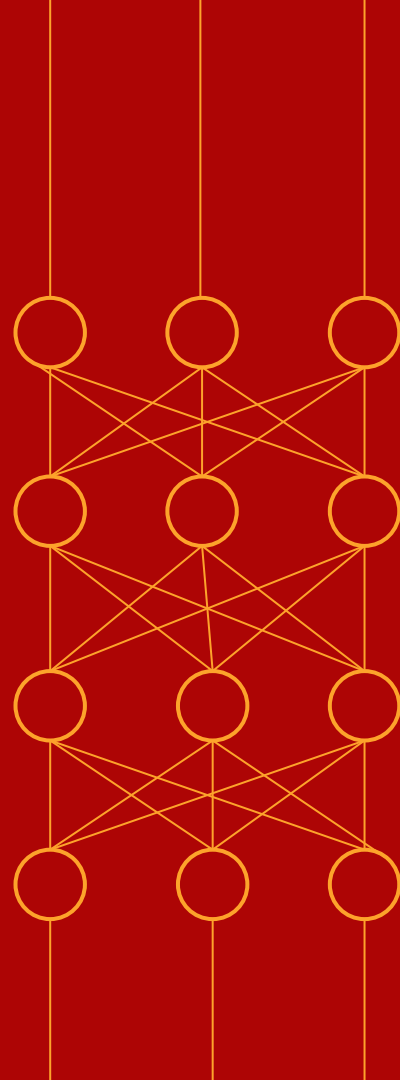
Timeline
# Research Questions

1. Can deep learning models approximate the performance of a community based effort in classifying FRBs?

2. Can this model then suggest anything about human level performance? If so, what does it suggest? If not, why?

3. Can we use this technique in the future for automated detection? If so, how? If not, why?
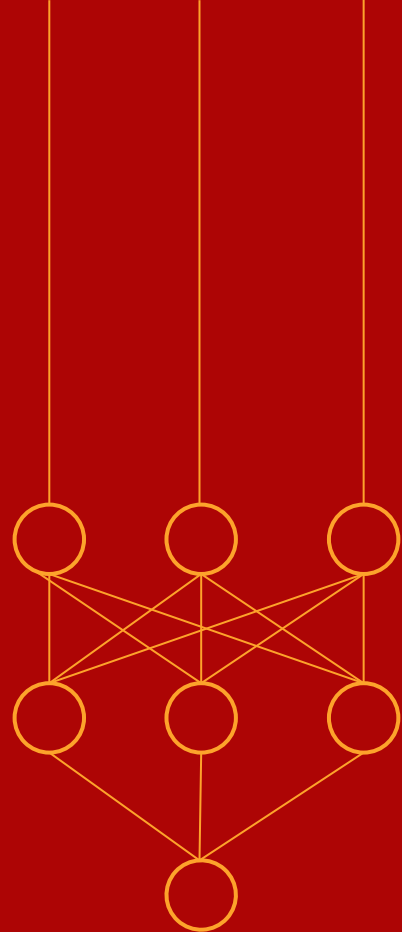
Timeline?
# Scope of Project

Project involves three main components:
**Simulation, Model Development** and
**Performance Assessments**

**Simulation:** We have limited amount of real observation data. We need to figure out how to realistically augment existing data and characterize our events [2 months?]

**Assessment** : we then use the simulations and run diagnostics on the model and conclude statistics. [1 month]

**Model Development:** Dedicated time to engineering the model. [1 month]

Timeline?
# How will this be done?

**Step 1:** Build simulation data — This will take the most amount of time

**Step 2:** Build CBNN on community data

**Step 3:** Evaluate model on community data

**Step 4:** Evaluate model on simulated data — This takes the most amount of time

Results
# Deliverables

A number of results from the project:

1. Python based FRB simulation that can inject bursts into existing filterbank or HDF5 files?

2. Model based on CBNN technique which can be potentially used for automated detection strategies.

3. We can then characterize the false positive or false negative rate of the model and conclude some approximated results about the community based search efforts!

4. Course work deliverables, presentation, paper etc.

# Find me at...



Personal Website: Peterma.ca
Email: Peterxy.ma@gmail.com
Twitter: @peterma02
Github: @Petchma