

# Automating neuroscience model identification and characterization

Ziqing Ji<sup>\*1</sup>, Siyan Guo<sup>\*1</sup>, Yujie Qiao<sup>1,2</sup>, Robert A McDougal<sup>1</sup> • 1 Yale School of Public Health; 2 Princeton University • \* contributed equally

### Introduction

Computational neuroscience allows insight into biological phenomena that cannot be easily measured directly. To do this effectively requires us to know the literature and to build on that which has been done before. This is challenging because:

- Models are published in a wide range of journals, often jointly with experiment.
- Models are complicated and cannot generally be fully defined in a paper.
- GitHub and other generic sharing sites do not provide standardized annotations or model visualizations.

## ModelDB (modeldb.science)



- Launched in **1996** (Gordon Shepherd)
- Home to over **1800** source codes for published neuroscience models using over **100** simulation tools/languages
- Consistent metadata fields and terminology
- Identification of reuse
- Identification of key literature cited by models lacksquare

### **Defining a ground-truth**

Ambiguity in defining computational neuroscience was considered through a cross-comparison of manual annotations (based on prompt given to GPT) with SPECTER, GPT-4, and GPT-3.5 outputs, which is then evaluated by Cohen's Kappa for annotator agreement.







The number of irrelevant responses increase.

- Boost model reuse and model refinement for