Neural Network Semantics
Thesis Proposal

Question 1. How can we better understand and control the behavior of neural networks as they learn over time?

Thesis Statement. Neural networks can be treated as a class of models in formal logic, simply by adding an interpretation function. By doing so, foundational questions about neural network inference and learning that were once elusive become natural and answerable questions in logic:

Soundness	answers	"How can we formally verify that a class of neural networks and its learning policies obey certain properties?"
Completeness	answers	"How can we build a neural network that aligns with constraints, even as the net learns and changes over time?"
Expressivity	answers	"What kinds of functions and learning policies are neural networks capable of representing?"

Outline:

1. Introduction
    - Make a helpful & practical example that will undercut the rest of the proposal
    - Motivation & Intro stuff
    - Thesis statement, said explicitly
    - Related work & context (there is a *lot* here! I might have to move it later??)
2. Background & Definitions
    - **Note:** This section will blend from known stuff into this newer idea, but I want to avoid addressing any of the above three questions, or referencing my work in a meaningful way, until the next section.
    - Modal Logic and its Models (incl formal definitions of soundness, completeness, and what I mean by satisfiability/modeling power)
    - Dynamic Epistemic Logic
    - Neural Network Models
    - Common Neural Network Learning Policies
3. Progress So Far & Goals
    - Explain which results (soundness, completeness, satisfiability/model power) over (static, dynamic) were (1) already known/done by others, (2) done by me during my PhD, and (3) are what I plan to do for the remainder of my thesis work. Show it on a picture
    - Divide this section up into Soundness, Completeness, and Modeling Power.
4. Plan
    - A concrete TODO-list with expected dates for finishing up the work.