

### (a) Implicit Negation



I **pretend** to be happy.



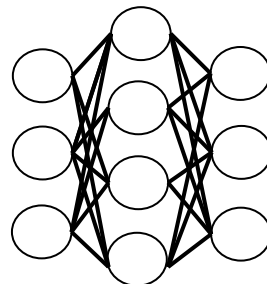
Predicted Label: **Positive**

**Positive** Confidence: 0.93

**Neutral** Confidence: 0.05

**Negative** Confidence: 0.01

Sentiment  
Classification  
Model



---

### (b) Self-conflicting Semantic



I **like** this food that you  
think is **disgusting**!



Predicted Label: **Neutral**

**Positive** Confidence: 0.16

**Neutral** Confidence: 0.61

**Negative** Confidence: 0.22

Sentiment  
Classification  
Model

