**Confirmation Generation**

It's a hard problem to translate from NL to ThingTalk, and different from Q&A, virtual assistant commands make side effects! We need confirmation!

Since there are infinite number of possible compound commands, it's impossible to provide manually created confirmation. We need machine learning!

---

**Almond Virtual Assistant**

Almond lets users issue compound commands in natural language and automatically translates them to programs in a formal language called ThingTalk. Compound commands combine APIs from an open-source, crowdsourced Thingpedia.

- **User input**
  - Get a cat picture and post it on Facebook with caption "Funny cat!"

- **ThingTalk**
  - `(image (from-flavor get)
    (image (facebook-post-pictures (insert-text image "Funny cat!"))
    )`.  

- **Execution result**
  - *Image of a cat*

---

**Model**

A Seq2Seq Model with attention
- **Encoder:** 1 embedding layer; 1 layer, bi-directional GRU
- **Decoder:**
  - 1 embedding layer
  - 1 layer GRU with attention (Luong et al 2015)
- **Loss function:** masked cross entropy
- **Hyperparameters**
  - Learning rate: 0.01
  - Batch size: 128
  - # of hidden units: 64
  - # of epochs: 100–300
  - Early stop after BLEU score stop improving within 10 epochs

---

**Dataset**

Dataset collected via Amazon Turk
- **Primitives:** 14,789 paraphrases (1,772 programs)
- **Total:** 9,777 paraphrases (1,638 programs)
- **Total:** 24,566 paraphrases (3,410 programs)

---

**Experimental results**

- **Different learning rate:**

  - ![Graph 1](image)

- **Different size of neural network:**

  - ![Graph 2](image)

---

**Discussions**

The work is still very raw and incomplete at the current state:
- Training loss is still relatively high despite overfitting
- A couple insights gained from a smaller training set:
  - The model tend to append random tokens after a perfect confirmation, penalty for duplication & long output is added, but does not seem to improve much
  - The loss function is not a very good measurement of the accuracy, BLEU score is used for early stopping