Motivation
CVS’s Refill Reminders program texts patients to refill their prescriptions, but the current model only understands simple keywords such as Yes/No/Help. Messages outside of this predefined dictionary are invalid and are sent a default response.

These invalid messages make up ~10% of all incoming texts, and results in unfilled refills and unhappy customers. Our objective is to build a chatbot with strong intent detection to vastly improve customer experience.

Data
We manually labeled over 3,000 invalid texts using non-ambiguous guidelines, and ultimately determined that there were 22 granular intents. These 22 intents are then combined into 7 types of messages that CVS will respond with.

Why 3,000 messages? Refill Reminders is just one of many SMS programs at CVS. We wanted to label just enough data to develop a strong model, but also ensure that future models for other programs can be developed quickly with minimal manual work.

Challenges
Though we have hundreds of thousands of texts, they are completely unlabeled.

The texts can have odd and unfavorable structures (e.g., very short texts, people names, medicine names).

The texts can contain plenty of irrelevant samples (e.g., mis-sent messages, gibberish strings).

10,000,000 Ignored Messages, Annually

Modeling
When designing the chatbot model, we explored several models in recent NLP literature. The best approach in both speed and performance was a custom pre-trained DistilBERT. Additionally, this model is flexible and can be painlessly applied to other CVS programs by only adapting the last step of fine-tuning.

Accuracy 84% on 22 Intents Speed 100+ Texts/Second

Unsupervised Techniques:
Zero-Shot Learning, SBERT Clustering

Semi-Supervised Techniques:
Siamese Neural Network, GAN-BERT

Supervised Techniques:
Bag-of-Words Random Forest, SVM

~ $3,000,000 in Annual Revenue
At ~$10 per Refill