



Chatbot Learning Models Multitask Learning and NLP

Kenneth Conroy, VP Data Science, Finn AI

When it comes to programming a banking domain chatbot, there are 2 key areas of focus to understanding as much as possible of what the customer wants to do.

The first is knowing what you want your bot to be able to do - what are you trying to optimize, and importantly, what it isn't trained to answer. At Finn, we call this our **Retail Banking Coverage**.

The second area is optimizing the accuracy at which you can predict what the user wants to do - how accurately can you match user input to the action they want to perform / question they want answered. We call this **Retail Banking Recall**.

The latter has many nuances - can the model understand in isolation what the user wants to do? Can the model make use of prior conversations or previous questions to better understand what the

user wants to do? Can the model be trained at a fine-grained level to do specific tasks, not broad all-encompassing answers which provide less value?

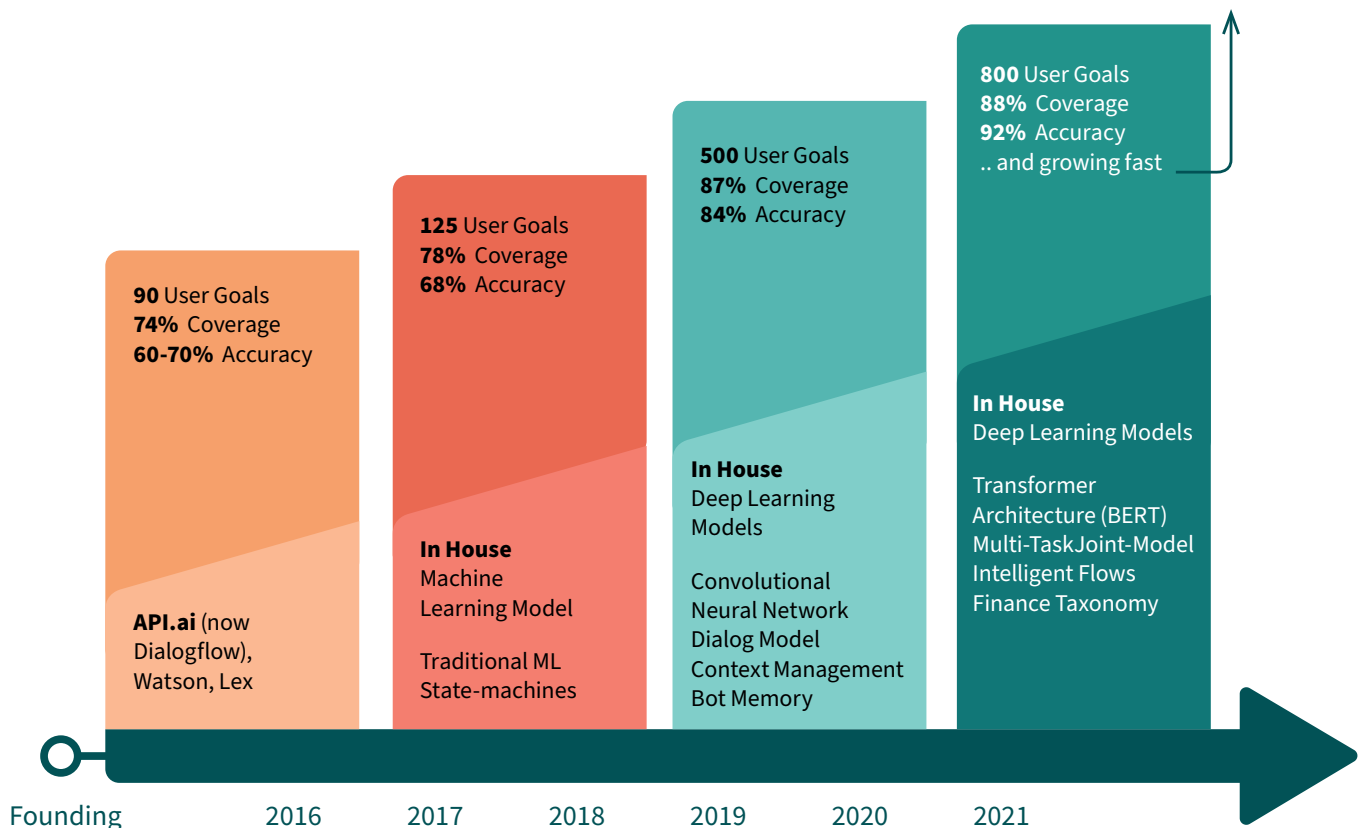
Finn AI creates chatbots for financial services - banks and credit unions, in the retail banking domain. Focusing on a single domain means we can provide understanding at a fine level of detail, using high quality training data consisting of millions of real-world interactions sent to our chatbots deployed on banks and credit unions. Data integrity is important with this level of complexity, as is a robust annotation framework.

Having this great data and a data-centric approach allows us to have many specific user goals and achieve both high retail banking coverage and recall. Recent advancements in Natural Language Processing (NLP) have centered upon pre-trained models and transformer architectures to optimize the accuracy of classification. In a focused vertical domain where thousands of different user objectives need to be classified, you need the flexibility to expand as your requirements grow. As there is a trade-off between retail banking coverage and recall, increasing retail banking recall will allow more flexibility in expanding retail banking coverage.

We also have a convolutional neural network for intent recognition, encompassing hundreds of user goals (what the user wants to do). When a human listens to a conversation, they are understanding contextually what the user is referring to: speaking about a branch while talking about trees is different to speaking about branches referring to banks.

Our latest model iteration takes a multitask learning approach and utilizes techniques taking advantage of pre-trained models to optimize accuracy, thus improving retail banking recall.

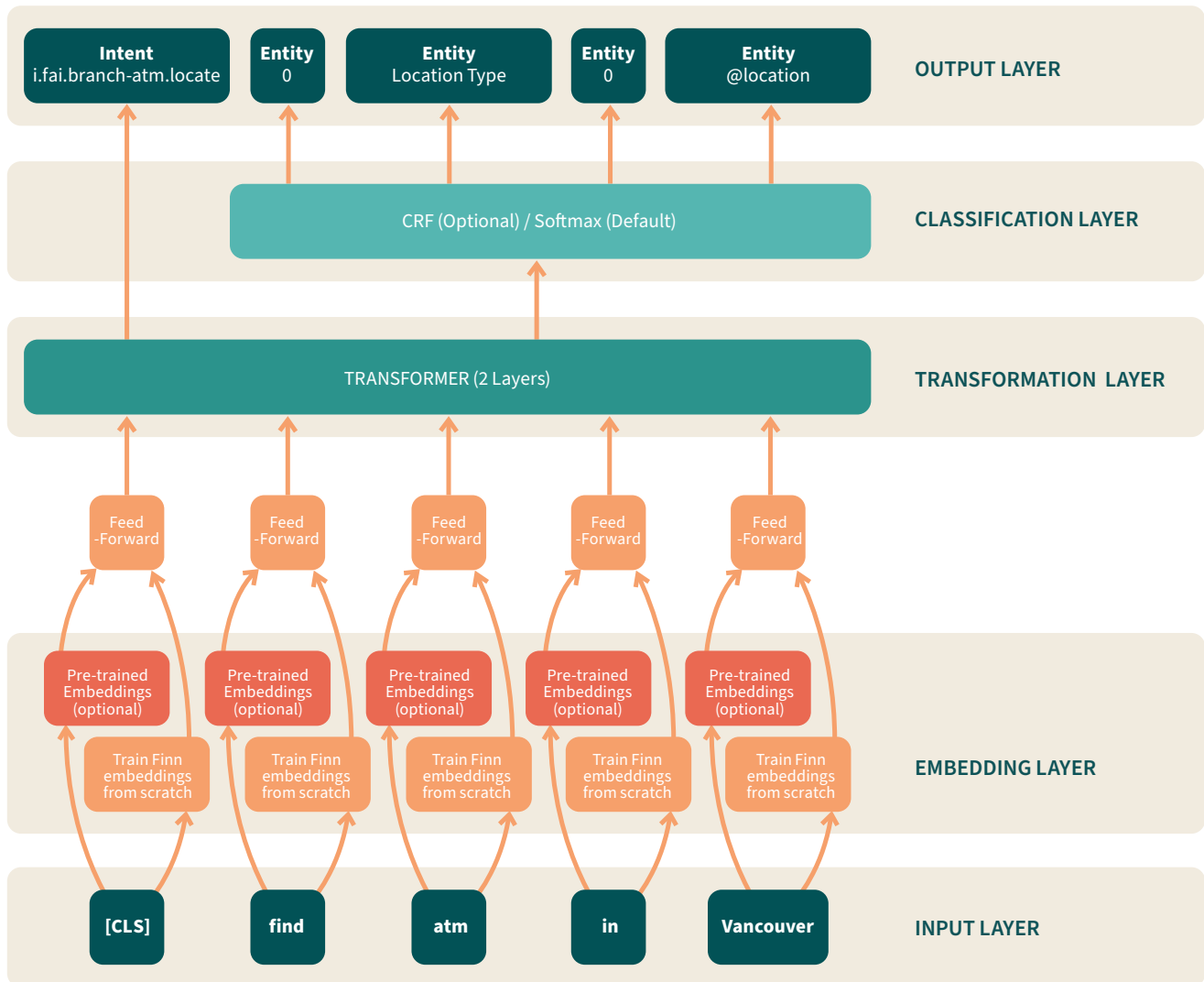
Evolution of Finn AI NLU



With multitask learning, we are looking for the intent and entities in parallel when making predictions on what the user intends to do, utilizing a Transformer architecture as seen below:

Finn NLU Model (Intent Classification + Named Entity Recognition)

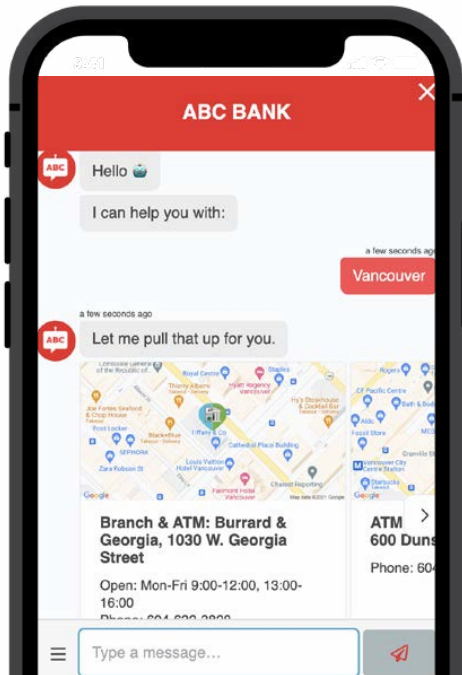
A simplified version of the architecture



We are leveraging pre-trained embeddings to give the model an understanding of English in a generic sense. This works by understanding how words relate to one another and other actions or activities in general use. This is a good foundation for NLP, but needs to be improved upon to truly understand what a user wants to do using data from the domain and the channel in which the model will be deployed. For this, Finn uses millions of unique utterances labeled to specific needs of a user to determine what they want to do in the retail banking space.

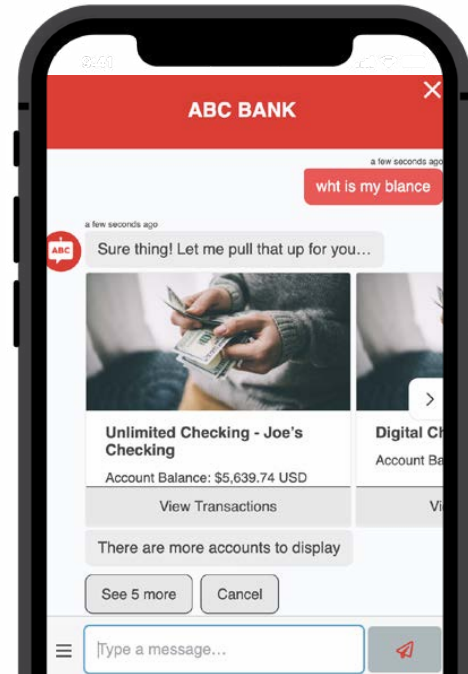
The results have been impressive so far: Retail Banking Recall (Accuracy) with our model was 84% when detecting one of 500 user goals prior to the development of the new architecture. With the new architecture, the accuracy jumps to 92%: A massive 8% improvement.

There are other benefits too: for instance, our bot is trained to work mostly with conversational language - full sentences where the user states what they want to do and with what product or service. When only products or services are mentioned, the model is less certain what to do: what is the context of this request?



With our memory feature, we can rely on prior conversation to help make the decision. For instance, a user may ask for interest rates, and then mention a specific product on the next input - then we can give the interest rates for this product. However, in cases where there is no prior input and the user just enters something short and without context, what should the model do? With our new architecture, when an entity is found, it can help direct the conversation to the most likely intent behind the user's input. For example, the utterance "Vancouver" will show all branch and atm locations closest to Vancouver.

Another noticeable benefit is the behaviour when the bot encounters spelling mistakes. Previous models were based on word tokens, with some "fuzzy matching" capabilities on entities only. With the new model, the bot takes into account sequences of characters, rather than words, and is much better at handling spelling mistakes throughout an utterance. As an example, "wht is my blance" provides the balance response as seen in the **screenshot below**.



Aside from these examples, the increase in overall accuracy means we have more and more users getting what they want on the first time, with a 33% decrease in false positives. We anticipate this will have a knock-on effect when it comes to user goal completion and containment.

The improvement to our model means we now have more bandwidth and potential to expand coverage, by defining and implementing new user goals.

About Finn AI

Finn AI builds conversational AI for banks and credit unions to improve their digital customer experience on mobile, online, and call center channels. Packaged as an all-inclusive Managed Service, Finn uses AI to solve the problem of creating an outstanding digital experience that increases customer success, engagement, and loyalty.

For more information visit finn.ai