The **ABCs** of **NLP**

Instabot

Instabot is one of the best artificial intelligence (AI) chatbot platforms available. The first step in creating an AI chatbot is to better understand a branch of artificial intelligence called natural language processing (NLP). Learn more about NLP in our quick study guide.

What is NLP?

Natural language processing (NLP) is a branch of artificial intelligence that helps computers understand, interpret, derive meaning, manipulate human language, and then respond appropriately.

In practice, NLP is accomplished through algorithms that compute data to derive meaning from words and provide appropriate responses.

How does Instabot use NLP?

Instabot allows you to build an AI chatbot that uses natural language processing (NLP). You can easily get started building, launching and training your bot.

The Basics

This is a practical, high-level lesson to cover some of the basics (regardless of your technical skills or ability) to prepare readers for the process of training and using different NLP platforms.

Our first task is to understand the lingo.

Disclaimer: NLP is a field that is under construction and constantly evolving. The use of vocabulary words may vary across different NLP platforms. So you may want to stay close to new, emerging trends and not to get too closely focused on semantics.

In order to better understand these terms, let's use an analogy to visualize these terms in a more practical context.

With NLP, it's very similar to training a new gardening store staff member on the job. You want to train the new associate to do certain things: interact with customers, figure out what they're saying, and interpret the customer's intent from that interaction.

Terms to Know

Intent:
Understanding of what a user wants. There are numerous phrases, sentences and expressions that have the same meaning. Intents are way of categorizing meanings for a string of words into a similar grouping.

Entities:
Entities are building blocks of intent-based systems that allow you to extract and categorize information from a string of text or a phrase.

Fallback Intent:
A Fallback Intent is really just a catch-all phrase that NLP systems can use when they aren't able to understand the actual intent of the user.

Training Data:
Data used to train an algorithm to better understand entities, intent and context.

Intent: An Example

The purpose of establishing an “Intent” is to understand what your user wants so that you can provide an appropriate response.

In the example above, the user is interested in understanding the cost of a plant.

In the gardening store analogy, the new hire won’t know anything about the way the store operates or how to best help customers, so you’ll have to give them a lot of historical requests that are similar to what your customers might ask them.

During training you might tell the new gardening store hire that “these types of questions relate to pricing requests,” or “these questions are relating to the soil types we have.” A vast majority of these requests will fall into different buckets, or “intents.” Each bucket/intent have a general response that will handle it appropriately.

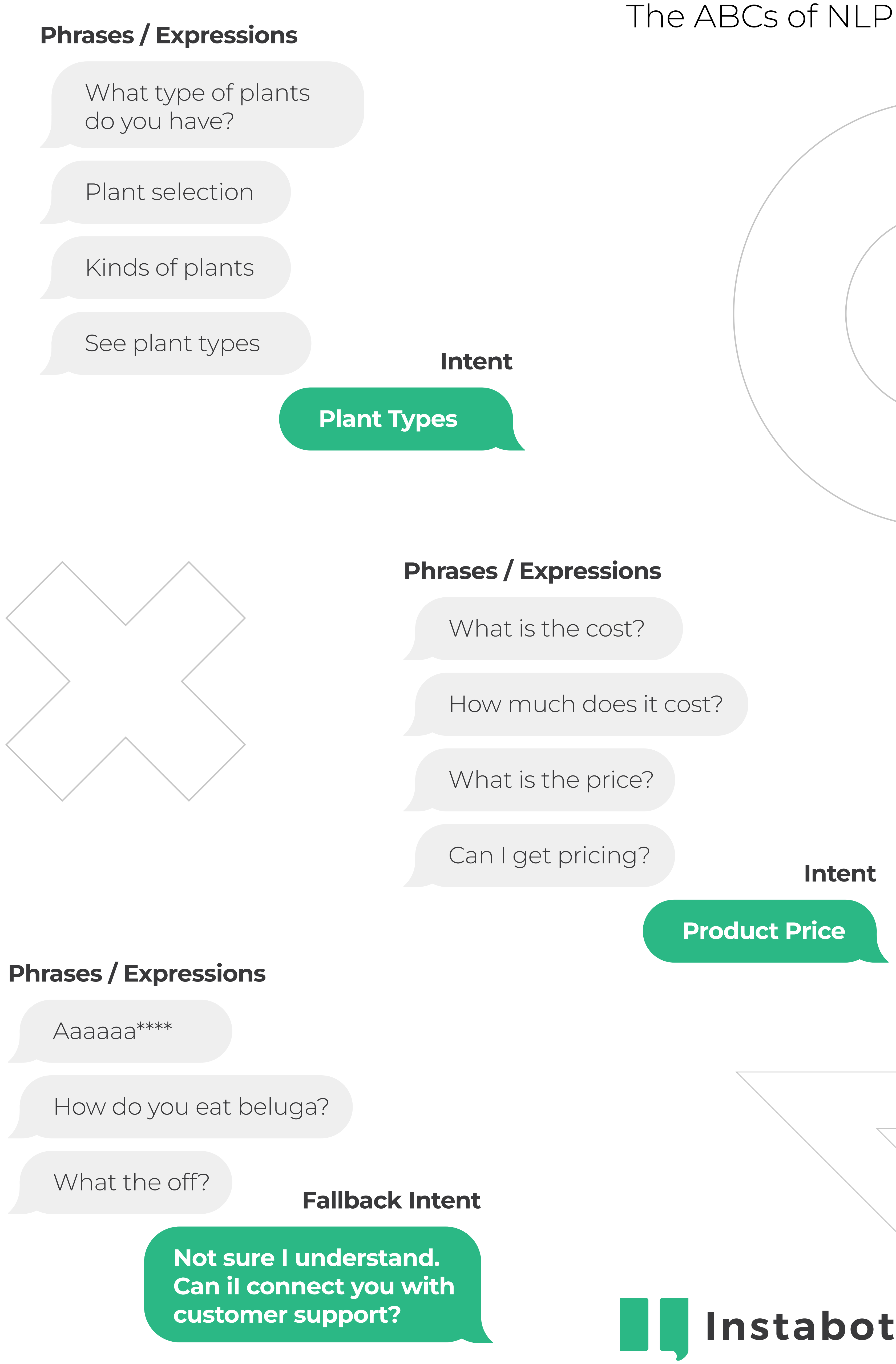
Although it comes easily to humans, in practice, having an AI system derive intent is a challenge, and due to the newness of this technology, it can be prone to errors. Having a “Fallback Intent” serves as a bit of a safety net in the case that your bot is not yet trained to respond to certain phrases or if the user enters some unintelligible or non-intuitive input.

In the diagram above, when a user responds to the bot with content the bot has yet to be trained on, i.e. “How do you eat beluga” or something unintelligible like “Aaaaaa****”, its fallback response would say “Not sure I understand. Can I connect you with customer support?”

The gardening store associate can be trained to learn certain types of intents and respond appropriately, but new hires have limits to their knowledge. If they’re not trained to handle certain types of things, they have to recognize that they don’t know how to handle that request, and they will then hand it off to someone who can help—that’s a fallback intent.

If a gardening associate were to be asked about free shipping for an online purchase, in real life they likely won’t know how to answer that request; it’s not in their repertoire of knowledge. Their fallback will be to direct the customer to someone who can help out with that question, probably saying something like, “Someone at our customer service desk can help you with that.”

Next are “Entities,” which help you to extract and categorize information from a string of text or phrase that you are given.



Phrases / Expressions

What type of plants do you have?

Plant selection

Kinds of plants

See plant types

Intent

Plant Types

Phrases / Expressions

What is the cost?

How much does it cost?

What is the price?

Can I get pricing?

Intent

Product Price

Phrases / Expressions

Aaaaaa****

How do you eat beluga?

What the off?

Fallback Intent

Not sure I understand. Can il connect you with customer support?

Training: An Example

In the example above, these are examples of ways in which NLP programs can be trained, from data libraries, to messages/comments and transcripts.

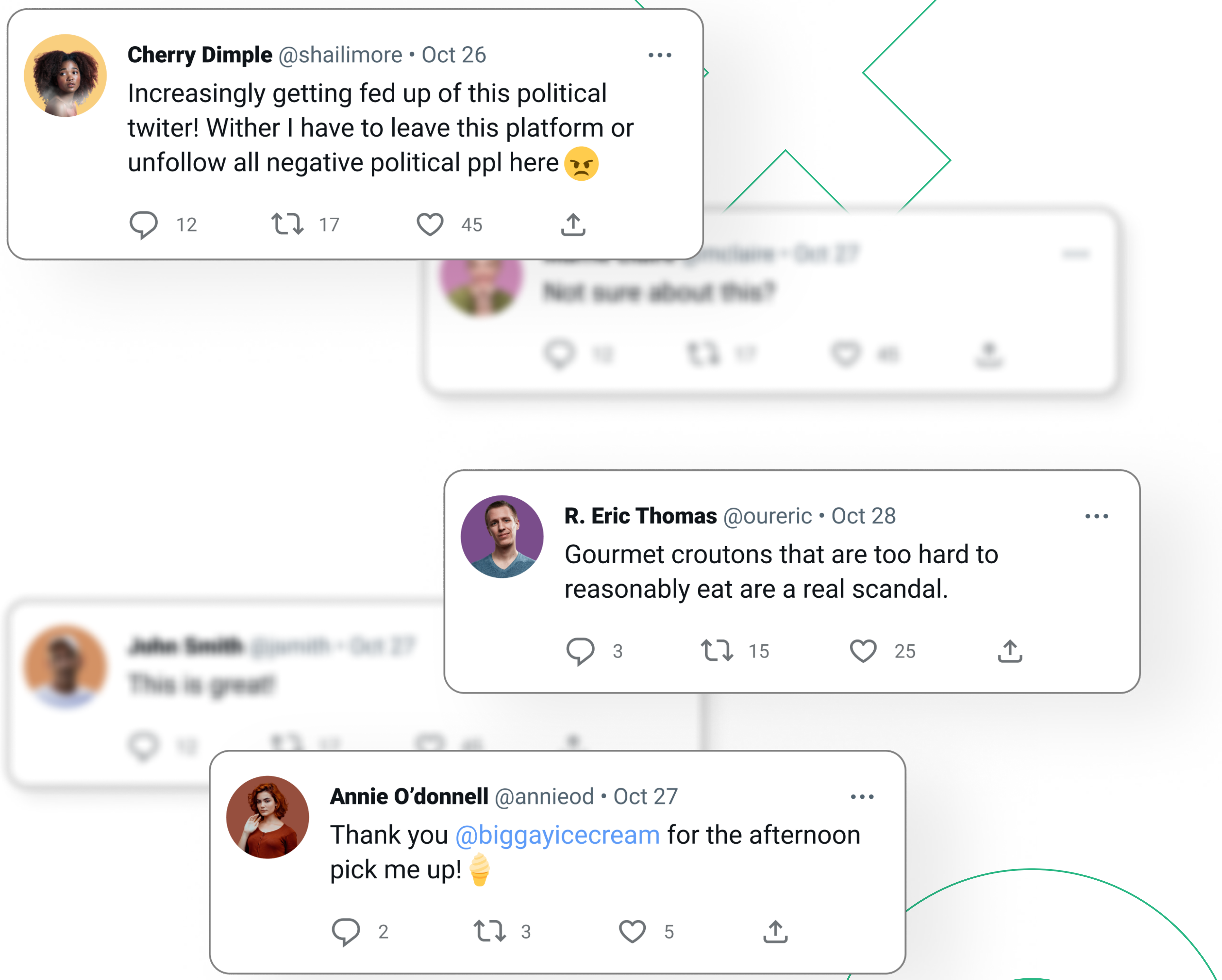
Providing expressions that feed into algorithms allow you to derive intent and extract entities. The better the training data, the better the NLP engine will be at figuring out what the user wants to do (intent), and what the user is referring to (entity).

Training data is often labeled and organized. Test data is a separate set of data that was not previously used as a training phrase, which is helpful to evaluate the accuracy of your NLP engine.

In the gardening store analogy, you start training the new hire from a level of zero knowledge, then you provide them with a large amount of historical data, tell them what it corresponds to, and put them out on the floor to see how they perform. At this point, the associate (or NLP system) has about 60-70% competency.

Training starts at a certain level of accuracy, based on how good training data is, and over time you improve accuracy based on reinforcement.

In practice, training material can come from a variety of sources to really build a robust pool of knowledge for the NLP to pull from. If over time you recognize a lot of people are asking a lot of the same thing, but you haven't yet trained the bot to do it, you can set up a new intent related to that question or request.



Cherry Dimple @shailimore • Oct 26
Increasingly getting fed up of this political twiter! Wither I have to leave this platform or unfollow all negative political ppl here 😠
12 17 45

[illegible] @mclaire • Oct 27
Not sure about this?

R. Eric Thomas @oureric • Oct 28
Gourmet croutons that are too hard to reasonably eat are a real scandal.
3 15 25

John Smith @jsmith • Oct 27
This is great!

Annie O'donnell @annieod • Oct 27
Thank you @biggayicecream for the afternoon pick me up! 🍦
2 3 5

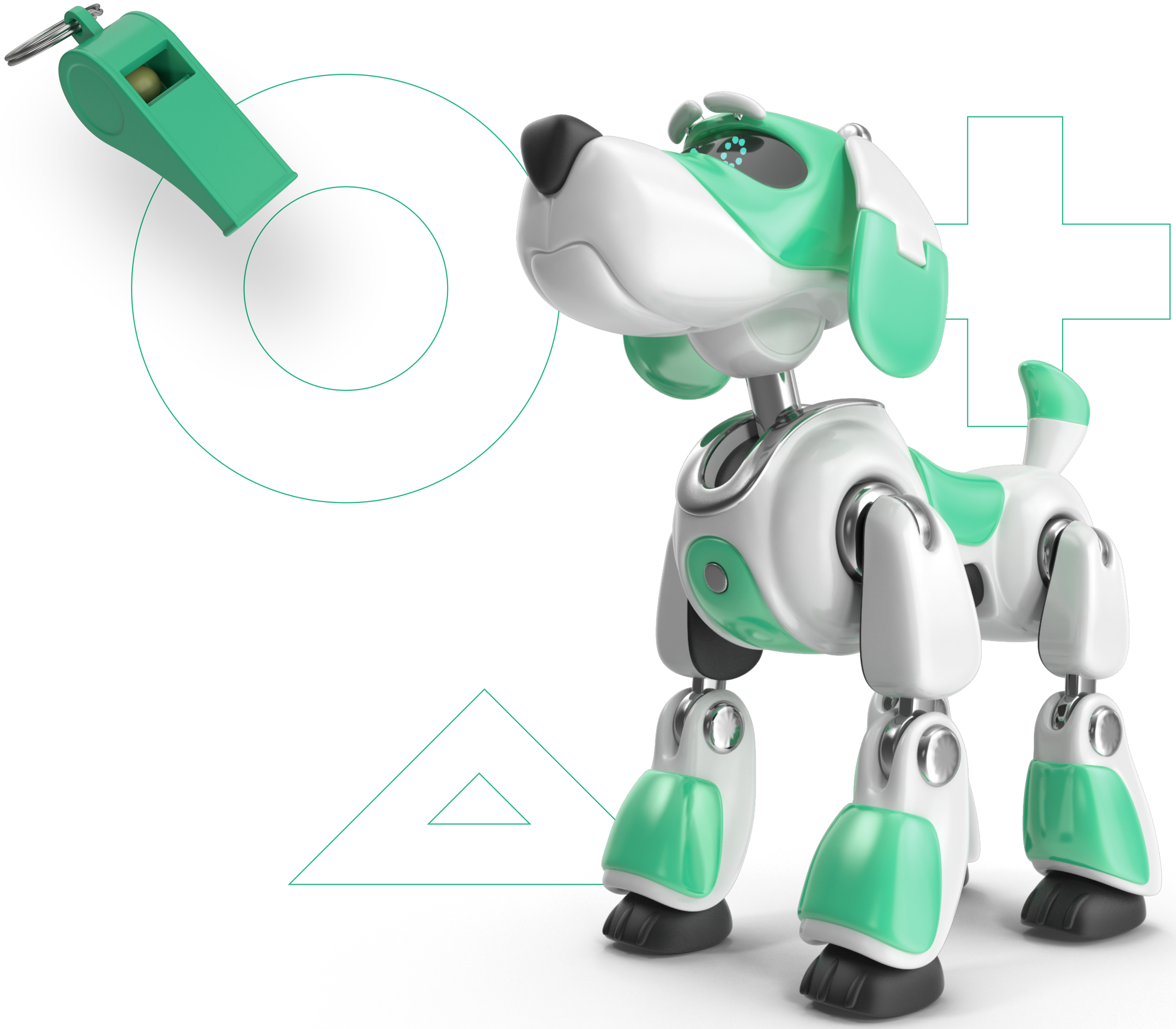
Negative	Positive	Neutral	Irrelevant
572	519	2333	1689

Total
5113

NLP Machine Learning

The next step, to fortify the NLP program's efficacy, is to incorporate NLP machine learning, which is essentially a feedback loop system where you take a look at the behavior of how the NLP platform has operated, to then provide validation of responses. Validating the responses is "Pavlovian" in a sense, with positive reinforcement, the machine or NLP program begins to notice patterns and better understand how to interpret the language and inputs it is fed, getting smarter over time.

Over time, the gardening store manager will observe how well the associate is doing, and either affirm the associate's performance, or in the case of questions they do not know how to answer, give them the right response to handle those requests, therefore strengthening their knowledge and increasing their level of competency.



This diagram demonstrates an outline of the basic process to gather, input,   and test data to train an NLP program.

GATHER TRAINING DATA

1

Gather training data from messages, emails, etc.

SCRUB/CLEAN/LABEL

2

Scrub your data, clean it, and label data to make training data

SELECT YOUR NLP PLATFORM

3

You can choose from a variety of platforms including: Watson, Dialogflow, or wit.ai

INSERT TRAINING DATA + ENTITIES

4

Input your training data and entities into their platform

LAUNCH + TEST

5

Input your training data and entities into their platform

PROVIDE FEEDBACK

6

Provide validation so that you can leverage machine learning algorithms.

NLP Training in Practice

One of our successes with NLP and machine learning leveraged Google's Dialogflow platform to train our bot to answer pricing questions. The process included:

1) Building out a bot to collect data- (Gathering user intent phrases)

Using open-ended, free-response question prompts in our bot conversations with users allowed us to collect information about questions users have, that a basic decision tree bot was not always equipped to respond to accurately. This bot always allowed users to "Ask me anything," in order to collect that data and parse through it for patterns or FAQs not included in the bot's initial structure.

2) Downloading all data- (Understanding user intent)

The next step was to review all the data collected from the bot's free-text questions to better understand what users are looking for from the bot, and common questions that could be added to the bot's repertoire of knowledge. Part of bot building and NLP training requires consistent review in order to optimize your bot/program's performance and efficacy.

3) Organize & Label Data- (Establish intents, training data, and entities)

While going through the responses, it's important to categorize them based on user's intent, especially since the same question or request can be worded and phrased in so many different ways. After categorizing the data, it's much easier to come up with groups of entities that correspond to the different user intents, and therefore will contain the most pertinent information with which to train the NLP program. The most popular and more relevant intents would be prioritized to be used in the next step.

4) Input into NLP Platform- (NLP Training)

Once intents and entities have been determined and categorized, the next step is to input all this data into the NLP platform accordingly. It's easy to group responses by category as well.

We used Google Dialogflow, and recommend using this API because they have access to larger data sets and that can be leveraged for machine learning.

Some Challenges When Employing NLP

t’s important to note that this technology is not perfect just yet. Since the technology is still relatively young and consistently developing at a rapid pace, there are some challenges when it comes to NLP, including:

Audio Speech Recognition:
There are many words or phrases that sounds similar depending on pauses and cadence.e.g. Youth in Asia v. euthanasia

Context:
Unless you have context, it is sometimes challenging to see the meaning in language.e.g. “We saw her duck.” Is it a bird that belonged to a person? An observed female person ducking? A person named ‘We’ who saw her pet?

Parsing Language:
Instances in which someone speaks out of grammar and being understood.e.g. “Tank working” is the same as “The tank is working.”

Polysemy:
Situations in which one word meaning several different things.e.g. “I need some paper.” “I wrote a paper.” “I read the paper

Language is a bit complex (especially when you’re talking about English), so it’s not clear whether we’ll ever be able train or teach machines all the nuances of human speech and communication.

That’s it!

Instabot