INTERNATIONAL JOURNAL OF ADVANCE RESEARCH AND DEVELOPMENT

(Volume 4, Issue 4)

Available online at: www.ijarnd.com

Training an AI chatbot on AWS

Kartik Jaitly

Student, Maharaja Agrasen Institute of Technology, New Delhi, Delhi

ABSTRACT

This paper aims to simulate a fully functioning Artificially Intelligent chatbot on Amazon Web Services (AWS). The chatbot learns from the dataset feed to it. This project consists of an Artificially Intelligent System i.e. an agent acts by learning from the data set fed to it. A model is trained on the Cornell University Movie corpus. After training on the movie dataset, the chatbot can hold on to fun conversations. This uses Amazon Web Services for training the model in the cloud and establishing the required infrastructure in the cloud to make it scalable and secure.

Keywords— *Artificial intelligence, Machine learning, Amazon web services, Elastic computing*

1. INTRODUCTION

The project focuses on the implementation of cloud computing for the purposes of AI chatbot applications and also the security of the cloud system.

The platforms used for the project include; Amazon Web Services, Anaconda, Tensorflow, Pandas, Numpy
The project leads to a better understanding of the cloud computing platform Amazon Web Services (AWS). Also, the techniques used to create the AI chatbot such as the Tensorflow API and python modules like Pandas, Numpy, Tensorflow and setting up the virtual environment through anaconda.

1.1 Conversational Agents

Chatbots, likewise called Conversational Agents or Dialog Systems, are an intriguing issue. Microsoft is making enormous wagers on chatbots, as are organizations like Facebook (M), Apple (Siri), Google, WeChat, and Slack. There is another flood of new companies endeavouring to change how buyers connect with administrations by structure purchaser applications like Operator or x.ai, bot stages like Chatfuel, and bot libraries like Howdy's Botkit. Microsoft as of late discharged their very own bot engineer structure.

Numerous organizations are planning to create bots to have regular discussions indistinct from human ones, and many are professing to utilize NLP and Deep Learning systems to make this conceivable. Be that as it may, with all the publicity around AI it's occasionally hard to tell reality from fiction.

In this arrangement I need to go over a portion of the Deep Learning strategies that are utilized to fabricate conversational operators, beginning off by clarifying where we are at this moment, what's conceivable, and what will remain almost incomprehensible for somewhere around a short time. This post will fill in as a presentation, and we'll dive into the execution subtleties in up and coming posts

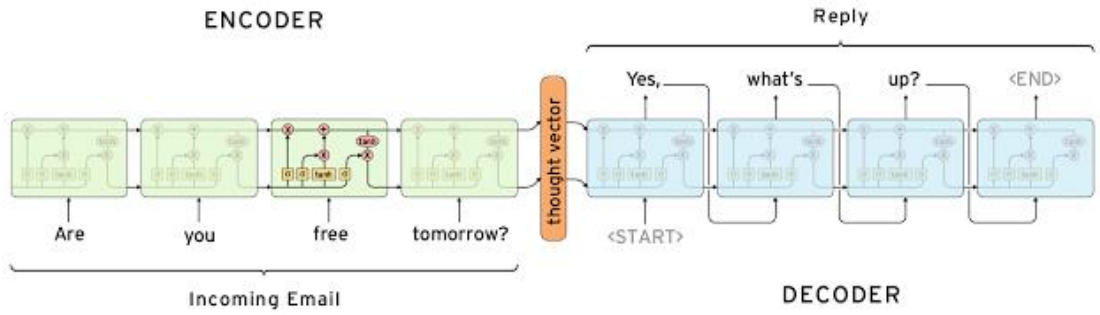
1.2 Retrieval-Based vs. Generative Models

Recovery Based versus Generative Models: Recovery based models (simpler) utilize an archive of predefined reactions and some sort of heuristic to pick a fitting reaction dependent on the info and setting. The heuristic could be as straightforward when in doubt based articulation coordinate, or as unpredictable as a gathering of Machine Learning Classifiers. These frameworks don't produce any new content, they simply pick a reaction from a fixed set.

Generative models (harder) don't depend on pre-characterized reactions. They produce new reactions starting with no outside help. Generative models are ordinarily founded on Machine Translation procedures, however as opposed to making an interpretation of starting with one language then onto the next, we "make an interpretation of" from a contribution to a yield (reaction).

The two methodologies have some undeniable upsides and downsides. Because of the vault of high-quality reactions, recovery based techniques don't commit syntactic errors. Be that as it may, they might be unfit to deal with concealed cases for which no fitting predefined reaction exists. For similar reasons, these models can't allude back to logical element data like names referenced

before in the discussion. Generative models are "more brilliant". They can allude back to substances in the information and give the feeling that you're conversing with a human. Notwithstanding, these models are difficult to prepare, are very liable to commit syntactic errors (particularly on longer sentences), and normally require colossal measures of preparing information.

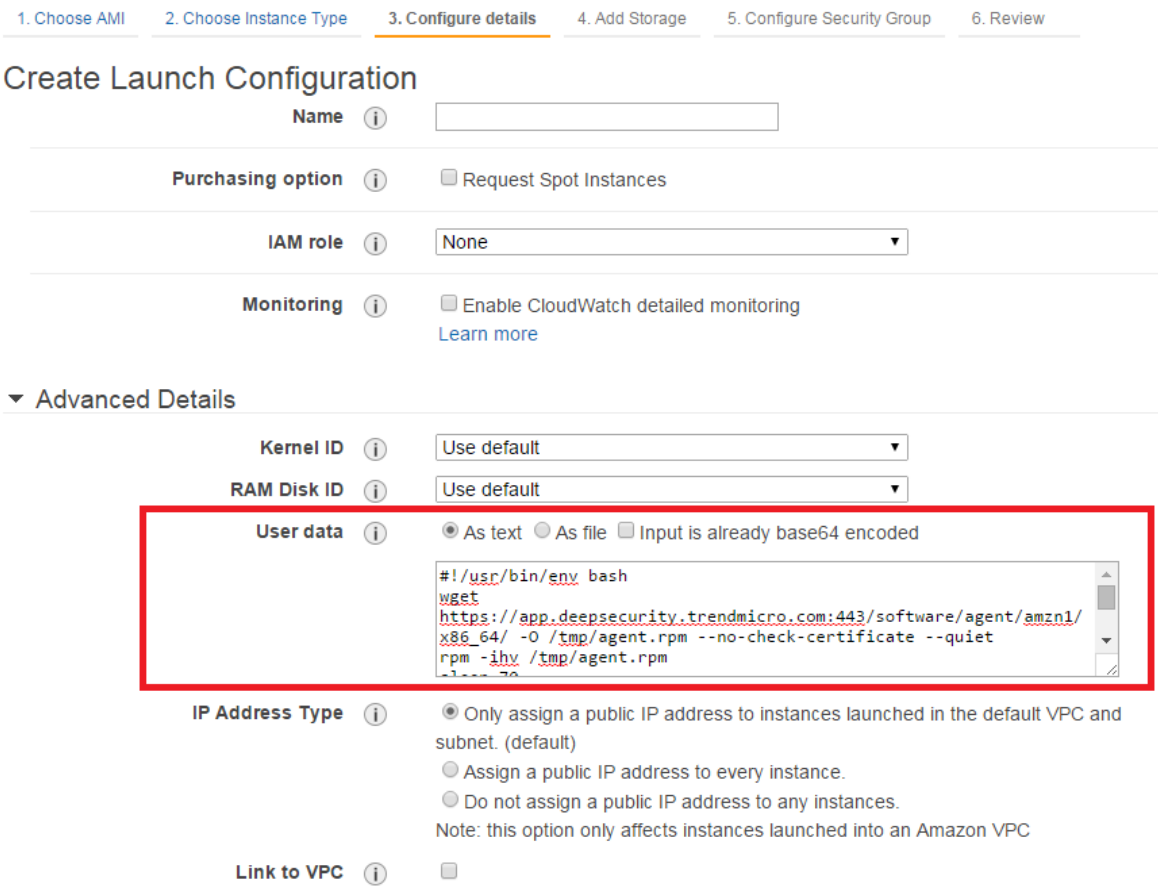


Conversational demonstrating is an imperative assignment in common language comprehension and machine knowledge. Albeit past methodologies exist, they are regularly confined to explicit areas (e.g., booking an aircraft ticket) and require hand-created rules. In this paper, we present a straightforward methodology for this errand which utilizes the as of late proposed grouping to succession system. Our model speaks by foreseeing the following sentence given the past sentence or sentences in a discussion. The quality of our model is that it very well may be prepared to start to finish and therefore requires many less hand-made guidelines. We find that this clear model can produce basic discussions given a huge conversational preparing dataset. Our fundamental outcomes propose that, regardless of upgrading the wrong target work, the model can chat well. It is capable of concentrate information from both space explicit dataset, and from a substantial, uproarious, and general area dataset of motion picture captions. On an area explicit IT helpdesk dataset, the model can discover an answer for a specialized issue by means of discussions. On a boisterous open-space film transcript dataset, the model can perform straightforward types of sound judgment thinking. Not surprisingly, we additionally find that the absence of consistency is a typical disappointment method of our model.

Profound Learning procedures can be utilized for both recovery based and generative models, yet examine is by all accounts moving into the generative bearing. Profound Learning models like Sequence to Sequence are remarkably appropriate for creating content and specialists are wanting to gain fast ground around there. In any case, we're still at the beginning periods of structure generative models that work sensibly well. Generation frameworks are bound to be recovery based on the present.

2. IMPLEMENTATION

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides secure, resizable compute capacity in the cloud. It is designed to make web-scale cloud computing easier for developers.



Amazon EC2's simple web service interface allows you to obtain and configure capacity with minimal friction. It provides you with complete control of your computing resources and lets you run on Amazon's proven computing environment. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change. Amazon EC2 changes the economics of computing by allowing you to pay only for capacity that you actually use. Amazon EC2 provides developers with the tools to build failure resilient applications and isolate them from common failure scenarios.

The implementation includes the hosting of a chatbot on cloud platform AWS using Elastic Load Balancer in front of the infrastructure provided by the elastic compute instance (EC-2 Instance).

This project also provides solutions for security using AWS Guard Duty and TrendMicro's Deep Security API.

3. CONCLUSION

This project concludes on the above shown experimental results of both the AI Chatbot and Cloud Security implementations done.

This AI conversational Chabot has a lot of applications ranging from B2C deliverables as well as mental health 24*7 assistance, customer service, among many others. It replaces and bogs down the need for multiple apps and different interfaces to perform some basic user actions/activities and requests. It is flexible to train using different data sets and will adapt accordingly, which expands the level of applications and uses.

For cloud security, the few options that can be considered are AWS guard duty, security groups, and trend micro deep security APIs. These security features are scalable, customizable and come with an ease of implementation. They have been widely used for enterprise-level security models.

The importance and relevance of this project stem from the fact that in today's day and age with data surge exponential the need for security is equally dire.

4. REFERENCES

[1] Wikipedia, "Deep Learning", Website URL: https://en.wikipedia.org/wiki/Deep_learning
[2] Amazon Whitepapers, Website URL: https://aws.amazon.com/ec2/
[3] Trend Micro "Deep Security API",
[4] Thomas N. T., Amrita Vishwa," An E-business Chatbot using AIML and LSA,"2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India
[5] Oriol Vinyals, Quoc Le, "A Neural Conversational Model" Submitted on 19 Jun 2015 (v1), last revised 22 Jul 2015 (this version, v3)

