GDPR Compliance Assistance

BACHELOR OF TECHNOLOGY
in
Electronics And Communication

By
Ritik Sharma(IIITU18246)

SCHOOL OF ELECTRONICS

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY UNA HIMACHAL PRADESH

JUNE 2021

BONAFIDE CERTIFICATE

This is to certify that the project titled GDPR Compliance Assistance is a bonafide record of the work done by

Ritik Sharma(IIITU18246)

in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Electronics and Communication of the INDIAN INSTITUTE OF INFORMATION TECHNOLOGY UNA, HIMACHAL PRADESH, during the year 2018- 2022.

Under the guidance of

1. Mr.Umesh Verma
2. Mr.Rohit Jagati
3. Mr.Suman Kumar
4. Mr.Sudhanshu Surana

Internship viva-voce held on: _____

Company Seal   Examiner (Umesh Verma)

ORIGINALITY / NO PLAGIARISM DECLARATION

I certify that this internship report is my original report and no part of it is copied from any published reports, papers, books, articles, etc. I certify that all the contents in this report are based on my personal findings and research and I have cited all the relevant sources which have been required in the preparation of this project report, whether they be books, articles, reports, lecture notes, and any other kind of document. I also certify that this report has not previously been submitted partially or as whole for the award of degree in any other university in India and/or abroad.

I hereby declare that I am fully aware of what constitutes plagiarism and understand that if it is found at a later stage to contain any instance of plagiarism, my degree may be cancelled.

Ritik Sharma(IIITU18246)

ABSTRACT

GDPR Compliance Assistance is a platform where users can ask a query to Support Platform i.e Chatbot for any context associated with GDPR and the Model will render the best 5 possible answers to the question. It utilizes a Previously trained Question-Answering model for the answer generation part along with Rasa-x customize chatbot that provides predefined dialogue flow for better answer extraction.

ACKNOWLEDGEMENT

I would like to thank the following people for their support and guidance without whom the completion of this internship in fruition would not be possible.

I would like to express my sincere gratitude and heartfelt thanks to Mr.Umesh Verma for the unflinching support and guidance, valuable suggestions and expert advice. His/her words of wisdom and expertise in subject matter were of immense help throughout the duration of this project.

I also take the opportunity to thank our Director and all the faculty of School of Computing/Electronics, IIIT Una for helping me by providing necessary knowledge base and resources.

I would also like to thank my parents and friends for their constant support.

Ritik Sharma(IIITU18246)

TABLE OF CONTENTS

Title	Page No.

School of Electronics/Computing, IIITU   [Internship code] : Roll number

Chapter 1

Introduction

1.1 What is GDPR?

GDPR = General Data Protection Regulation

GDPR is a regulation that requires businesses to protect the personal data and privacy of EU citizens for transactions that occur within EU member states. And non-compliance could cost companies dearly. Here's what every company that does business in Europe needs to know about GDPR.

1.1.a What are the 7 principles of GDPR?

1. Lawfulness, fairness and transparency.

2. Purpose limitation.

3. Data minimisation.

4. Accuracy.

5.Storage limitation.

6. Integrity and confidentiality (security)

7. Accountability.

1.2 Question-Answering Overview and Basics

Question answering is a task in **information retrieval** and **Natural Language Processing (NLP)** that investigates software that can answer questions asked by humans in natural language. In **Extractive Question Answering**, a context is provided so that the model can refer to it and make predictions on where the answer lies within the passage.
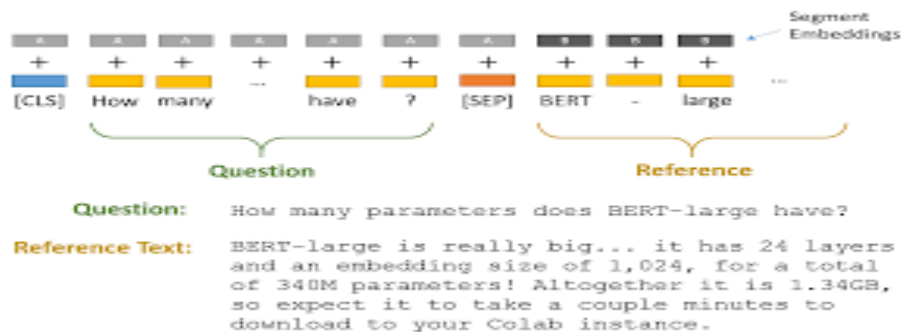
←- This the basic Diagram of how we pass out context and Question to the transformers architecture and get our Answer Start and End Token within the reference context .
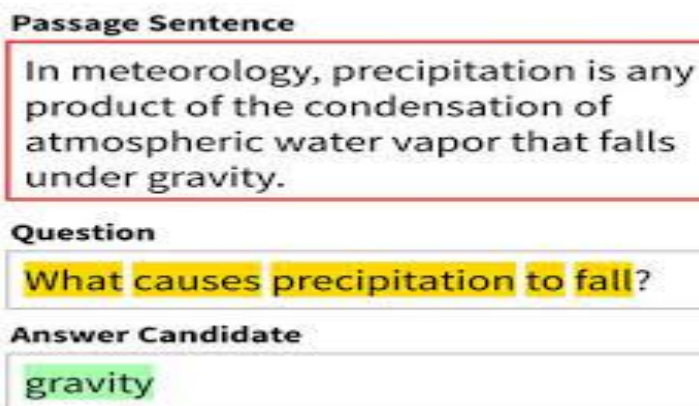
Question-Answer Example where we pass passage as our reference text and question whose answer is in reference text so we get "gravity" as answer

Chapter 2

Models And Architectures Used

2.1 BERT

Bidirectional Encoder Representations from Transformers is a Transformer-based machine learning technique for natural language processing pre-training developed by Google. BERT was created and published in 2018 by Jacob Devlin and his colleagues from Google.

!! To understand Bert we need to understand its parent architecture Transformers

2.1.a Transformers

The Transformer in NLP is a novel architecture that aims to solve sequence-to-sequence tasks while handling long-range dependencies with ease. It relies entirely on self-attention to compute representations of its input and output WITHOUT using sequence-aligned RNNs or convolution.
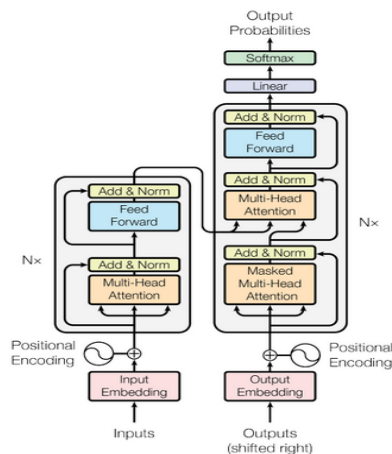
This is the architecture of Transformers where we can see that there are two sections encoder and decoder and it uses multi-headed attention blocks along with positional encodings with feed forward neural networks
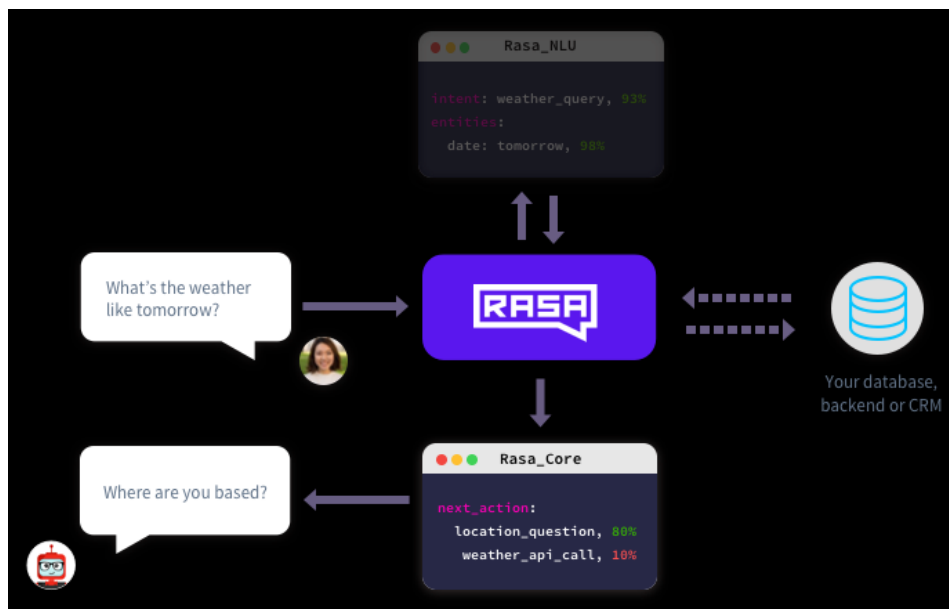
Bert   is a Encoder Part of Transformers
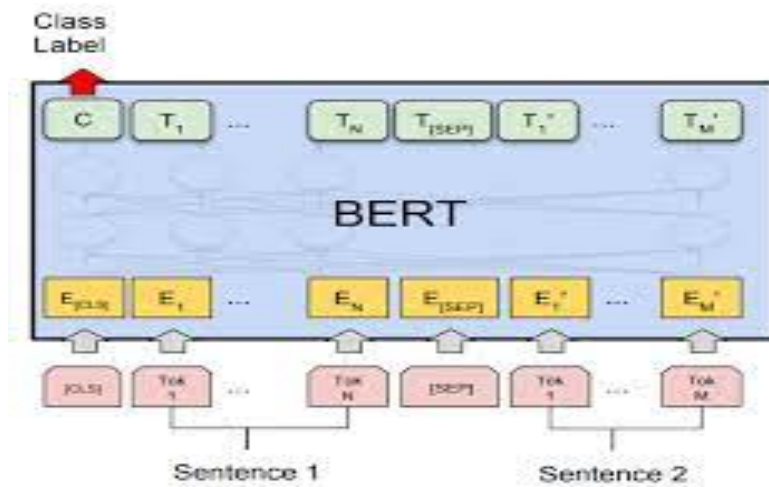
Like for the text processing part bert uses 3 layers (segment embeddings,position embeddings,tokens embeddings)

2.2 Rasa-X

"In the project I have used Rasa-X for making Chatbot with predefined dialogue flow "

Rasa X is a tool for Conversation-Driven Development (CDD), the process of listening to your users and using those insights to improve your AI assistant.

key Features of Rasa-X

1. layers on top of Rasa Open Source and helps you build a better assistant
2. is a free, closed source tool available to all developers
3. can be deployed anywhere, so your training data stays secure and proprietary
4. **View and annotate conversations:** Filter, flag, and fix conversations that didn't go well to continually improve your assistant.
5. **Deploy anywhere:** Ready-to-deploy Docker containers and orchestration to run Rasa on premise or using your favourite cloud provider.

2.2.a   Understanding Rasa-X Files and WorkFlow
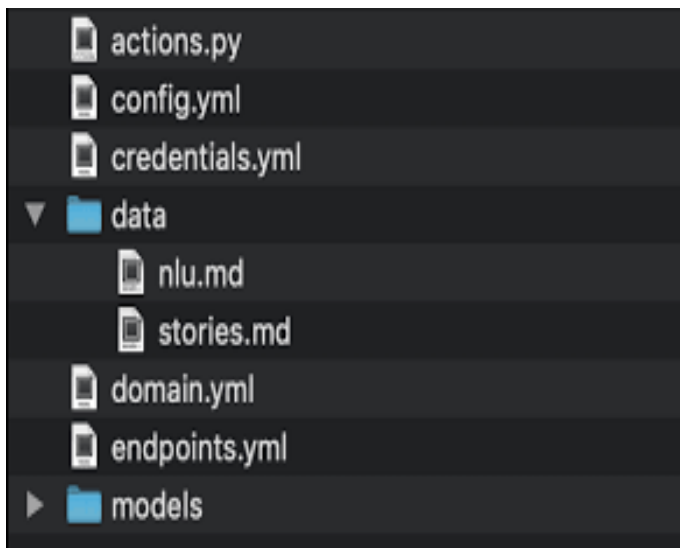
Rasa has Some files associated with it so we need to understand all of them

1. **Config.yml**

   configuration of your NLU and Core models. In-case you are dealing with Tensorflow or Spacy, you need to define such a pipeline here. To handle this file, you show knowledge about Machine Learning and Deep Learning.

2. **credentials.yml**

   details for connecting to other services. In case you want to build a Bot on Facebook Messenger, Microsoft Bot Framework, you can maintain such credentials and token here. So basically you just need to add Facebook, slack and Bot framework related configuration, rasa will automatically do rest for you. Remember that you need to host Rasa over https domain. During development, you can use **ngrok** as a testing tool.

3. **nlu.md**

   your stories. This is required for Rasa Core. There is something called "Dialog Flow in Rasa" where Rasa Core controls the flow of the conversation between you and the chatbot, so for that flow, you need to train the chatbot using these stories. So in case you want your chatbot to be very perfect in different contexts (stories) you can add those stories here.

4. **domain.yml**

   your assistant's domain. This file combines Different Intent which a chatbot can detect and a list of Bot replies. Remember you can define your Custom Action Server Python method name here (in underscore format), so that Rasa will call that python method for you

5. **actions.py**

your assistant's domain. This file combines Different Intent which a chatbot can detect and a list of Bot replies. Remember you can define your Custom Action Server Python method name here (in underscore format), so that Rasa will call that python method for you
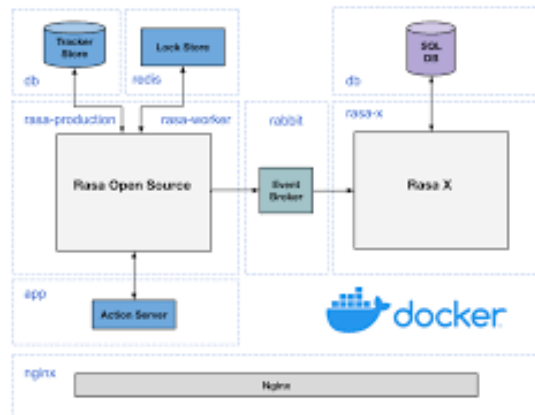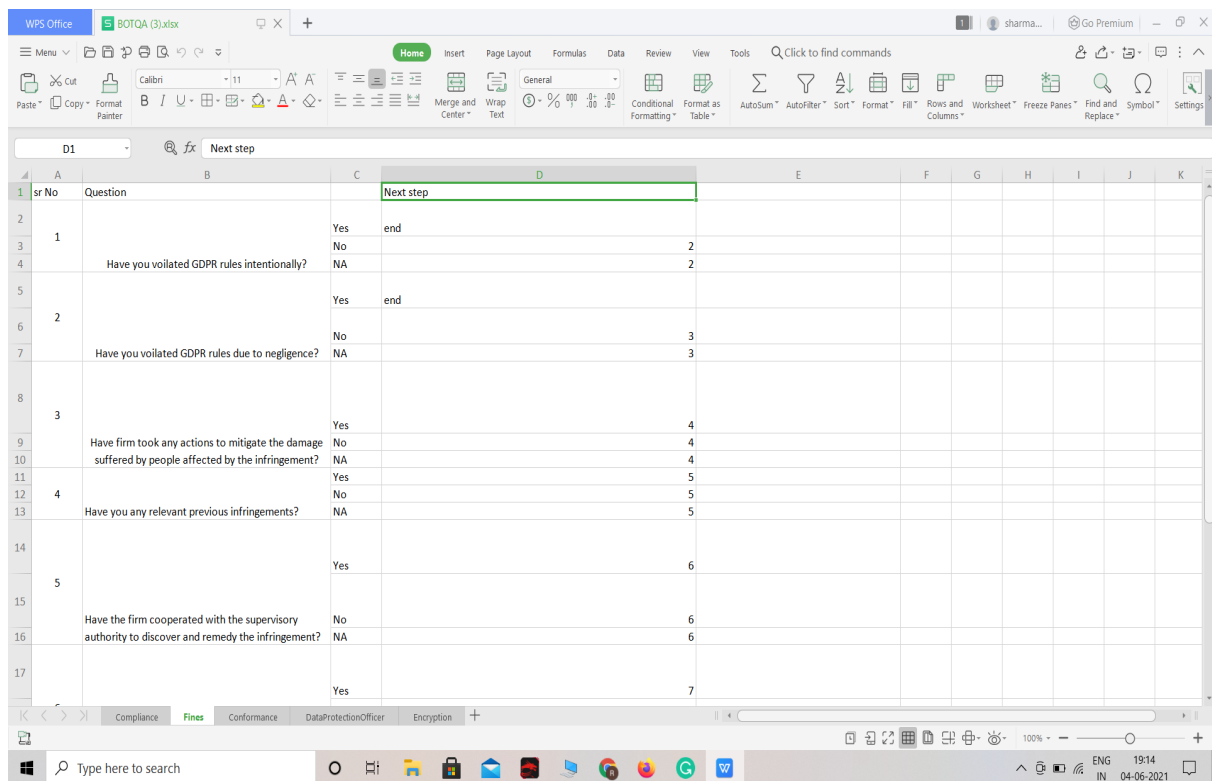
Fig: Rasa-X Action Server Hit Flow

Chapter 3

Project Workflow And Integration Of Rasa-X With BERT QA Model

3.1 Dialogue Flow of Rasa-X for GDPR Compliance

In this we have defined Predefined set of question that will going to display one after another and we have different flow at every level i.e if user select answer option 1 for the level first(first question) then it will go to 1.1 flow and if he selects option 2 then he will go to 1.2 flow like this manner

we can see that we have different steps at every level of question

3.2 BERT Question Model

If user gets his answer in early stage in the conversation then there will no hit in the action server and if user gets to the last question then Rasa-X will allow user to enter his query and then this query will be passed over to pretrained BERT model that was fine tuned on the "question - answering" task for particularly GDPR document(Domain Adaptation). and we have written a customize python script for that in actions.py file of Rasa-X which takes question i.e user entered query as input and returns generated answer from the BERT model

Basic Workflow

References
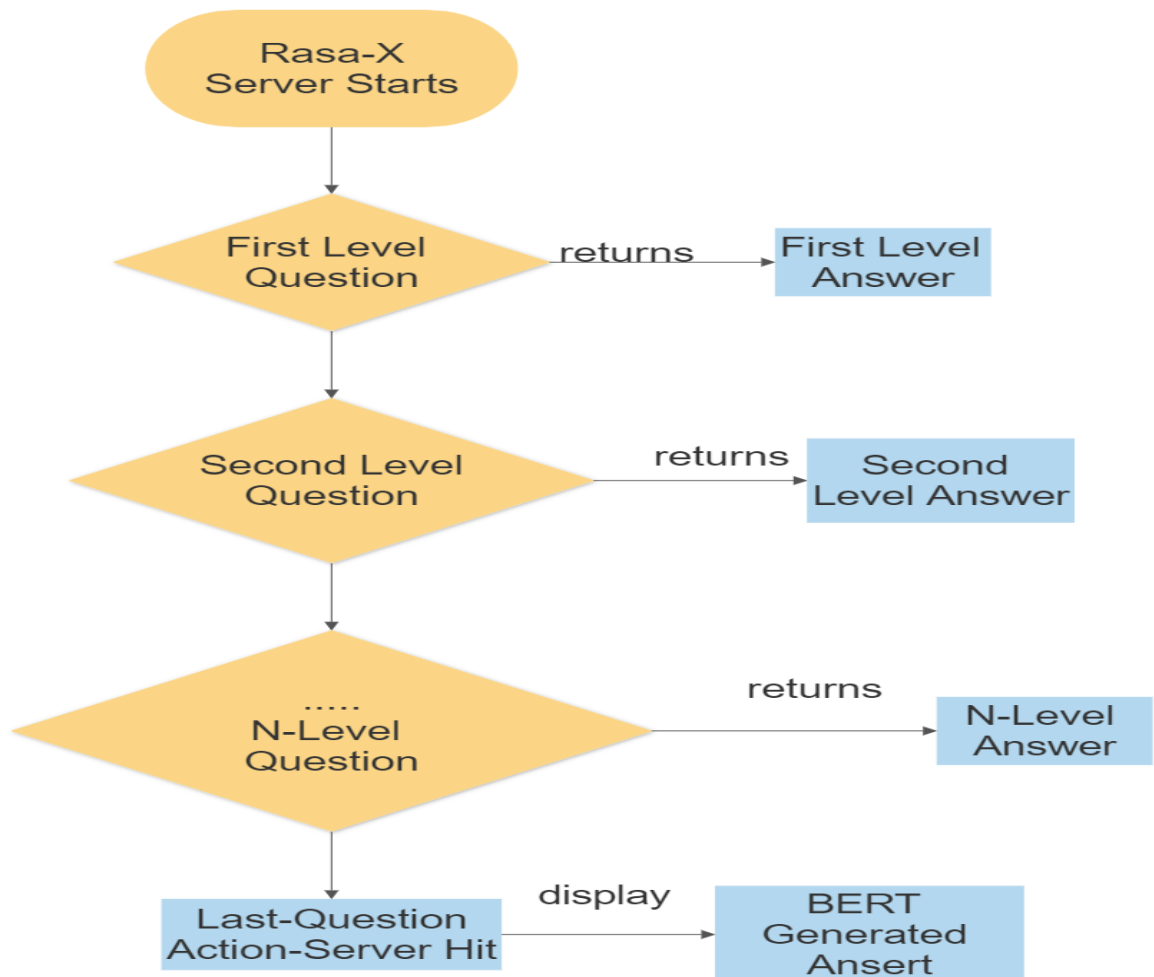
[1]   https://towardsdatascience.com/pre-training-bert-from-scratch-with-cloud-tpu-6e2f71028379 (Pretraining of BERT)

[2]   https://qa.fastforwardlabs.com/pytorch/hugging%20face/wikipedia/bert/transformers/2020/05/19/Getting_Started_with_QA.html (Fine Tuning of BERT-Domain Adaptation for QA Task)

[3]   https://www.pragnakalp.com/nlp-tutorial-setup-question-answering-system-bert-squad-colab-tpu/

[4]   https://simpletransformers.ai/docs/qa-model/   +
        https://github.com/pytorch/fairseq/blob/master/examples/roberta/README.md
        (Training of Roberta model)

[5]   https://jalammar.github.io/visualizing-neural-machine-translation-mechanics-of-seq2seq-models-with-attention/ (Recurrent Neural Network and Language Modelling)

[6]   https://www.youtube.com/watch?v=TQQlZhbC5ps(Transformers Architecture)

[7]   https://huggingface.co/models (hugging face pre-trained models hub)

[8]   https://rasa.com/docs/rasa-x/ (Rasa-X Basics Introduction)

[9]   https://www.pragnakalp.com/chatbot-development-tutorial-rasa-x/(Chatbot Development)

[10]   https://www.wired.co.uk/article/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2018#:~:text=GDPR's%20seven%20principles%20are%3A%20lawfulness,new%20to%20data%20protection%20rules. (GDPR Basics and Introduction)

Appendices

Appendix A

Code Attachments

!!! Code and exact details of this project Cannot be displayed publicly under Non-Disclosure Agreement(NDA).

<u>**NON- DISCLOSURE AGREEMENT**</u>

THIS NON- DISCLOSURE AGREEMENT ("Agreement") is entered into on this ¸ 2021 (effective date) by and between:

Sphinx Worldbiz Limited a company duly incorporated and existing under the Companies Act, 1956 having its registered office at Sphinx Worldbiz Limited, A-27B, Sector 16, Noida (hereinafter referred to as "Disclosure/ Company" which expression unless repugnant to the context hereof, shall mean and include its subsidiaries, representatives, sister concerns, affiliates, successors and assigns, of the FIRST PART.

AND

Mr./Ms. , **Ritik Sharma** s/o **Rajinder Prasad Working** as Software Trainee for the project

"GDPR Compliance Assistance" Under MR.Umesh Verma at

Sphinx Worldbiz Limited & are hereinafter collectively referred to as the PARTIES.

The term "Disclosure" shall mean the Party to the Agreement disclosing Confidential Information and the term "Recipient" shall mean the Party to the Agreement receiving the Confidential Information.

WHEREAS:

A. It is anticipated that, during the course of discussing and considering possible transactions or relationships between the parties, it may be necessary for either of the parties to disclose certain confidential and proprietary information in written, oral, visual and/or physical/sample form to Recipient or the Recipient may otherwise come into control or possession of certain Confidential Information relating to the Disclosure, including those of the customers of the Disclosure which is considered as confidential information by the Disclosure as defined hereunder. The proprietary and confidential information to be disclosed by the Disclosure to the Recipient shall be subject to the terms and condition set out herein below.

NOW THEREFORE, to facilitate the free exchange of information in consideration of the premises set forth herein, the parties hereto have agreed to enter into a Mutual Non Disclosure Agreement as follows:

1. '**Confidential Information'** shall mean and include any information disclosed by Disclosure to the specifically authorized personnel of the Recipient, either directly or indirectly, either orally or in writing, by inspection of tangible objects relating to Disclosure (including, without limitation, documents, prototypes, samples, media, documentation, discs and code) . Confidential Information shall include, without

Signature (Employee)_____

limitation, any building plans, site maps, security features, materials, trade secrets, know-how, formulae, processes, ideas, strategies, data, network configurations, system architecture, designs, flow charts, drawings, hardware, media and the contents thereof, proprietary information, business and marketing plans, financial and operational information, information about customers (either present, past or prospective) of the Disclosure, all business, financial, operational and any other information or data, in any form, supplied to the Recipient by, or at the direction of the Disclosure relating to the Transaction, including any copies, excerpts, notes, memoranda summaries, analyses or compilations thereof generated by the Recipient (collectively, the "Information"), the information regarded as confidential by the customers of the Disclosure, all non-public information, material or data relating to the current and /or future business and operations of the Disclosure and analysis, compilations, studies, summaries, extracts or other documentation prepared by the Recipient based on information disclosed by the Disclosure and any other information considered as confidential information by the Disclosure and specifically notified to the Recipient.

For purposes of this Agreement, Confidential Information shall be limited to information of a business nature. Such information shall **not** include confidential processes, formulations, designs, source code or other information that would otherwise be considered technical in nature or that could be used in the creation of a product, solution or business process.

2. For purposes hereof, "Confidential Information" **shall not** include:

    (a) information in the public domain at the time that it was provided by the Disclosure or which subsequently came into the public domain other than as a result of breach of this Agreement; or

    (b) information obtained from a third party unaffiliated with the Disclosure (provided such party was not bound by confidentiality agreements with the Disclosure); or

    (c) information independently developed by the Recipient without reference to the Information; or

    (d) information in the possession of the Recipient prior to its disclosure by the Disclosure to the Recipient; or

    (e) information which the Recipient is required to disclose by law or under any legal

School of Electronics/Computing, IIITU   [Internship code] : Roll number

process, provided that the Recipient gives the disclosing party a reasonable opportunity before such disclosure to contest the legal process or seek a protective order; or

(f) information outside the scope of the Definitive Agreement.

3. **Ownership**- The Disclosure retains the sole and exclusive ownership and intellectual property rights in the confidential information and no license or any other interest is granted to the Recipient by virtue of this Agreement or any other agreement between the parties, unless expressly agreed to the contrary by the Disclosure in writing prior to such disclosure.

4. All Confidential Information:

(a) shall be treated as confidential by the Recipient, taking such action as shall be

Signature (Employee)_____

necessary or desirable to preserve and protect the confidentiality of the Information and in any event using means not less than those used to protect its own confidential information;

(b) shall be used solely for purposes of evaluating the Transaction as provided in this Agreement, and for no other purpose; and

(c) shall not be disclosed to any third party and limit access of the same only to those of Employee's personnel, agents and representatives on a "need to know" basis (each an "Authorized Person") without the Discloser's prior written consent. Provided however that, prior to disclosing the Confidential Information, the Recipient shall apprise each of such person who receives the Confidential Information to be cognizant of and to comply with the terms and conditions of this Agreement.

5. **Covenant Not to Compete** - The Employee also agrees that at no time during the term of their employment with the Company will they engage in any business activity which is competitive with the Company nor work for any company which competes with the Company. For a period of two (2) years immediately following the termination of their employment, they will not, for themselves or on behalf of any other person or business enterprise, engage in any business activity which competes with the Company.

6. **Non-solicitation-** During the term of employment, and for a period of two (2) years immediately thereafter, the Employee agrees not to solicit any employee or independent contractor of the Company on behalf of any other business enterprise, nor shall they induce any employee or independent contractor associated with the Company to terminate or breach an employment, contractual or other relationship with the Company.

7. **Soliciting Customers After Termination of Agreement** - For a period of two (2) years following the termination of employment and relationship with the Company, the Employee shall not, directly or indirectly, disclose to any person, firm or corporation the names or addresses of any of the customers or clients of the Company or any other information pertaining to them. Neither shall they call on, solicit, take away, or attempt to call on, solicit, or take away any customer of the Company on whom they have called or with whom they became acquainted during the term of their employment, as the direct or indirect result of their employment with the Company.

8. **Injunctive Relief** - The Employee hereby acknowledges (1) that the Company will suffer irreparable harm if they breach their obligations under this Agreement; and (2) that monetary damages will be inadequate to compensate the Company for such a breach. Therefore, if they breach any of such provisions, then the Company shall be entitled to injunctive relief, in addition to any other remedies at law or equity, to enforce such provisions.

9. In the event that the Recipient or any of its Authorized Persons becomes legally compelled (including, without limitation, by law, rule, regulation or governmental regulatory or administrative or similar agency as part of a judicial or administrative proceeding (including, without limitation, by deposition, interrogatory, request for information or documents, subpoena, civil or criminal investigative demand or otherwise)) to disclose any of the Confidential Information, the Recipient shall provide the Disclosure with immediate written notice of such requirement so that the Disclosure may seek a protective order or other appropriate remedy or waive

Signature (Employee)_____

compliance with the confidentiality obligation in relation to such information. In the event notice prior to disclosure is not possible, the Recipient will give written notice as soon as reasonably possible following such disclosure.

10. **Duty to protect**- The recipient undertakes to protect the Confidential Information with a reasonable degree of care as it uses to protect its own Confidential Information, to protect the Confidential Information received by it under this Agreement. In addition to such degree of care, the Recipient agrees not to, in any way disclose, copy, reproduce, modify, use, or otherwise transfer (including temporary transfer) the Confidential Information or any media containing the Confidential Information to any other person or entity without obtaining prior express written consent from the Disclosure. The Recipient shall at no time create any charge, lien or encumbrance of whatsoever nature on the Confidential Information. These restrictions shall also apply to any media containing the Confidential Information. The Recipient shall take all steps necessary/ appropriate to protect the confidential information against unauthorized disclosure or use. The Recipient shall immediately notify the Disclosure of any unauthorized disclosure or use of any confidential information that comes to the Recipient's attention, and shall take all action that the Disclosure may reasonably request to prevent any further unauthorized disclosure or use of such confidential information.

9. **Term**- This Agreement is effective on the Effective Date, and shall remain in effect throughout the term of your employment with the Company and for a period of two years thereafter.

10. **Return of Materials**- When Employee's employment with Company ends, for whatever reason, Employee will promptly deliver to Company all originals and copies of all documents, records, software programs, media and other materials containing any Confidential Information. The Employee will also return to Company all equipment, files, software programs and other personal property belonging to Company.

11. **Indemnity**- The Recipient shall be responsible and liable to the Disclosure for any unauthorized disclosure or any breach of the terms and conditions herein contained.

15

The Recipient hereby indemnifies and agrees to keeps the Disclosure saved, defended, harmless and indemnified from and against all such direct costs, charges, expenses and damages whatsoever arising therefrom.

12. **Relationships**- Nothing contained in this Agreement shall be deemed to make an Employee a partner or joint venturer of the Company for any purpose.

13. **Severability**- If a court finds any provision of this Agreement invalid or unenforceable, the remainder of this Agreement shall be interpreted so as to best to effect the intent of Company and Employee.

14. **Integration**- This Agreement expresses the complete understanding of the parties with respect to the subject matter and supersedes all prior proposals, agreements, representations, and understandings. This Agreement may not be amended except in a writing signed by both Company and Employee.

Signature (Employee)_____

15. **Waiver**- The failure to exercise any right provided in this Agreement shall not be a waiver of prior or subsequent rights.

16. **Attorney Fees and Expenses-** In a dispute arising out of or related to this Agreement, the prevailing party shall have the right to collect from the other party its reasonable attorney fees and costs and necessary expenditures.

17. **Governing Law**- This Agreement shall be governed in accordance with the laws of India.

18. **Jurisdiction**- The Employee consents to the exclusive jurisdiction and venue of the Courts located in Noida in any action arising out of or relating to this Agreement. The Employee waives any other venue to which the Employee might be entitled by domicile or otherwise.

IN WITNESS WHEREOF, THE PARTIES HERETO HAVE EXECUTED THIS AGREEMENT AS OF THE DAY AND YEAR FIRST ABOVE WRITTEN.

Signature: Ritik Sharma

SPHINX WORLDBIZ LIMITED

Signature: Preeti Sharma

By: Nitish Kumar

Title : Software Trainee

Date: 07-May-2021

Employee Name: Ritik Sharma

THANK YOU

School of Electronics/Computing, IIITU   [Internship code] : Roll number