ELEC 3841 Internship Report

Name: Wan kwong Fai

UID: 3035780142

Major: Computer Engineering

**Part 1: History, background, and main business of the Company**

HKT Limited, also known as Hong Kong Telecom, is one of the largest telecommunications companies of Hong Kong. It has a dominant position in fixed-line, mobile, IDD and broadband services in Hong Kong. HKT Group is a subsidiary of PCCW since 2000, after it was acquired from Cable & Wireless plc.

The history of this company has more than 150 years. In 1871, The predecessor of Cable & wireless established its Hong Kong branch, for its submarine communications cables that connect from Hong Kong to Saigon, Fuzhou, Manila, Labuan, and Macau. In 1981, the Hong Kong branch of Cable & Wireless plc was incorporated as Cable and Wireless (Hong Kong) Limited (later known as Hong Kong Telecom International; HKIT).

While Hong Kong Telephone Company, which established in 1906, which operate for the fixed-line telephone as business at first, and in 1983, it started to build their own mobile radiotelephone service, and went public in the next year.

In 1987, Cable & Wireless (Hong Kong) and Hong Kong Telephone Company merged to form a new telecommunications group, with the new holding company called Hong Kong Telecommunications Limited, replacing Hong Kong Telephone Company as a listed company on the stock exchange of Hong Kong and one of the constituents of Hang Seng Index. At 2000, Cable & Wireless HKT was acquired by PCCW, and renamed to PCCW-HKT Limited. And at 2008, HKT Group Holdings Limited (HKT) was formed to hold the telecommunications service, media, and IT business of the PCCW group.

The man business of the HKT include commercial and international business, online network provider, mobile network provider, entertainment, and online transaction.

The commercial and international business is aimed to provided ICT services to small, medium, and large enterprises, like the Smart Identity Card System for the Hong Kong SAR Government.

While for the network, it provided both internet and mobile communications in Hong Kong, like Navigator, CSL Mobile /1010, SUN Mobile etc.

For the entertainment, HKT has now TV, which provided news, sports and infotainment programming and interactive services to Hong Kong public. ViuTV also is running by HKT.

While for online transaction, HKT Payment Limited is a developer of “Tap & Go”, a prepaid mobile payment for Hong Kong users, which provide HKT-eye over-the-top media services and Internet Protocol TV service delivered to firmware-modified tablet computer, and PPS (Payment by Phone Service) – a bill payment service provided by HKT and EPS.

**Part 2: Organizational structure of the department with which you have worked.**

I worked in the department called Technology and strategy and development, it is a relatively new department inside the company, it set up at 2019, with HKT innovation lab, office locate at the PCCW towel, Taikoo Place, 39/F. It also owned a place at Cyberport. This department is relatively few people when compared to others department, including interns, only have 14 people. This department is built to aim to discover and testing different technology trend and try to build the prototype, guidelines and inspire the other department in HKT to embrace or know more about the latest trend of the cutting-edge technology. Within this tiny team, it is divided into two major teams, the tech team, and the commercial team. Team is mainly for research and development about the latest technology and try to use them within our company, while the commercial team is trying to represent the department to communicate with other teams or to the external customer, like writing the business plan and presentation to others. Although the team is small, and is a relatively new department inside the company, we had built a lot of idea prototype and try to adopt them inside the company and inspire the other team to have more knowledge on the latest technology. For example, our department has a lot of the sharing talk, to share the latest technology to the company, also, we have an internal newspaper -tech watch, issue every two weeks, to tell what the news about the technology is, and how they will influence others. Also, our team also represent the HKT in many events, like booths, to demonstrate our latest product to the public.

**Part 3: Schedule of work undertaken by me**

From 5th June to 31st August 2023. Monday to Friday, 9am to 6pm.

**Part 4: Description of work that you have completed, and experience gained &**

**Part 5 My comment on the work done.**

As mentioned as above, I worked from June, the first month I am just doing research on different technology. As the ChatGPT has become a more and more popular topic now, so one of the focuses of research will be on the ChatGPT. I am responsible for looking up the Azure OpenAI, which is owned by Microsoft, and the advantage of using Azure’s one instead of just using the OpenAI one is because Azure is promised not using our data to train the model. At the same time, we try to expand the usability of the large language model (LLM), as the current LLM can only just answer you according to the knowledge it already knows. And it is just a language model, which also limited the ability of it. To solve this question, I and my teammate also look up how to make the answer for it to be better, and how can we allow it to get more information and enhance the ability of it. For the first one, we find out that there are two types of method to enhance the answer quality, by better prompt or by fine-tuning the model.

The first one, having a better prompt. We can address this by different means, one is giving situation and the steps-by-steps instruction to the LLM, and allow the LLM to think more on it, because there are a more steps, so it also need to answer more to get more prompt. The second one, which I think is better, is by a few shots approach.

Like you are giving some examples of the question answer pair, the LLM can have a better understanding on question answer relationship and get you a better approach.

On the first month, not only trying to know more about how to use the LLM, but I also spending a lot of time to understand the principal behind it. As I did not take any AI/ ML course before, so it is a relatively hard time for me to directly go deep into the principal of LLM, so the first month, what I have done, is learn different AI/ML concepts online, to have a better understanding of it. And at the same time, learning how to be a professional “Prompt Engineering”, which aim to make the LLM to answer better according to different prompt. And at the end of June, I use streamlit library, which is a python library provide a quicker way to build the website with simple code. Besides the job inside the company. I also have a chance to represent the HKT to attend the Hong Kong Open Conference on 10th June in HKSTP. At there, I meet a lot of people, and share a lot of the experience with other people, like Google, Microsoft, Amazon etc. Also, presenting the demo to the participates also train my social skills and the English communication skills.

In July, 3 interns including me, keep continue doing research on the LLM. 3 of us doing research on different LLMs, like I look into the AzureOpenAI and OpenAI, while others look into the Google PALM and Vicuna, which is an open-source project on the hugging face. At the same time, we learn about the basic of LLM, like what is transformer, how the data is being train etc. And we also look up a open source project – LangChain, which is a python / JS library to extend the ability of LLM, like connecting to the network, connect to your email account, and more. But unfortunately, when we start using it at the late of June, it doesn’t have a full document on the support for LLM except OpenAI ones, so we need to build the connection between the LangChain and different LLMs. And at the same time, I really experience the world is fast changing. For example, in late June, I try to build a local vector storage for LLMs so that the LLMs can according to our database to answer the question. I use the Chroma as the local vector database storage and try to use different embedding method to build the information. However, when I successfully build the prototype of it, Microsoft announced that they release congestive search for their large language model, which is overlap with my work. So, I stop to improve my one and try to switch to use Microsoft one, but we need to wait for the access right from Microsoft, so until now, the date I write this report – 13th August, we still haven’t gotten the access right from Microsoft, so I just continue to build my one first. Besides the problem with database system, we also meet a lot of the problem. Like we trying to connect the Zapier, one of the famous automate software, which can be using the LangChain to connect the LLM to do a lot of the automation. But we failed to connect it with the models. I have also set up the new server for our department use, but one of the funniest things is, when our new server arrived, it does not contain the power supply, so we need to purchase it and wait one week more. And as no one in my department know how to set up it, I take in charge to do that. I spend almost 1 day to assemble the server, and try to install the OS and the according set up, like the IP cable, SSH connection, the requirement software set up, which spend me around 2 days, but I think it is quite a unique experience, as this is the first time I seeing the real server and try to play on it, no just on textbook or the pictures. After building up the server, we try to run the Vicuna model with fine tuning on it, although we only use the smallest version of Vicuna -7B model, it also take us around a week to train it have some ability to understand the Chinese words, which originally only have the knowledge on English, and we have the experience to experience how to train a model, we need to connect the data, format it, and seeing the learning rate and the overfitting/ underfitting problem, which I learn a lot of the AI/ML knowledge inside, and inspire me how data is important is, as the hardest thing is how to prepare the data set, while the training part is actually easier. While for the website, we originally using streamlit, which has a very simple UI look up, we try to use one of the projects on GitHub to build a new UI design.

电脑萤幕的截图

描述已自动生成

The one using streamlit

电脑软件截图

描述已自动生成

The new one

However, switching to the new UI has a lot of the problem, the first one is the new UI is using react, which Is on top of the JS, HTML and CSS. Which I never learn before. So, we spend around 2 weeks to learn the basic of JS, HTML, CSS and react, and spend around 1 week to try to understand the source code. As others two interns they have more knowledge on this, I passed the front-end part to them, I just focus on the backend of it, I make use of flask, which is a very popular backend server for python application and try to connect this UI to my flask server. At the same time, I need to learn the principal of how they connect, like the rules of using API, what is the header of it, in which format? How to receive it in UI and backend etc. Besides the LLMs, I also explore more Ai application, one of it is about drawing. Some stable diffusion on hugging face already provides a simple version of it. Diffusion principal behind is making plenty of noise on the image, and according to the prompt to erase some of the noise, so that it can create the image according to the prompt. One of the famous companies who did that is called mid-journey, which you can just type the prompt on discord and the mid-journey bot can response the image you want. As it is also a part of the AI, so I decided to connect it into ours one, the hugging face one is simple, just download the diffusion model to the server, and use the GPU to run is fine. But mid-journey doesn’t provide any API for us to use. So, I refer on one of the projects on GitHub, I build a bot program to grab the content in discord, to send the prompt command and receive the mid-journey bot response.

Apart from that, in July, our department also has an event called “共創明teen”, which is aimed to provide some event for some s1- s3 secondary school students, to have a more knowledge on the latest technology. I demonstrate one of the demo sessions and answering their questions. I think this is meaningful as we can share more knowledge to others.

While in the August, so far, I have done two kinds of things, one is trying to use LLMs to make the tech watch for our department. Tech watch is an internal news that issue by our department, which will talk about the last 2 weeks technology trend. As the writer for the tech watch will be retired soon, so my boss wants us to think of how to uses the AI to automatically do this kind of work. I separate the job into different steps, like using online tools to search the hot topics about technology within 2 weeks, and then use some online grabbing technique to grab the content from the website, and cut the news into pieces for LLMs to summarize the content, after that, I put the summary of the news and a role for the LLMs to generate a title and the content for the tech watch. Besides that, I also use LLMs to generate a more detailed one, with the role to share the technology and how they are related to us, with some suitable content, like asking them to subscribe the news etc., all by LLMs and python program, and the detailed one will be posted on the social media of HKT Innovation lab. This thing can greatly reduce the time needed for our department, as before, we usually spend 7-8 hours on this thing but now, with the aim of LLMs, we can just let the AI to finish it, which also is one of the use cases of using LLMs.

Another thing I am trying to do is the code interpreter feature of our project. In July, OpenAI has announced they now have the code interpreter function for their GPT 4 model, which enable the LLMs to execute some of the code, which is very powerful, and it has the capability to do a lot of things. So, we also want to build something like this in our project, so we think of using a Jupiter notebook as a server and allow LLMs passed some code and return the result like error messages back to the LLMs, to do the further analysis. However, we still get stuck at some of the points, and I think it is not possible to finish it before my departure. Nevertheless, I think our prototype has already been inspired how we can use the AI, how to use LangChain to make the LLMs more suitable for our task. So overall, I think it is meaningful for what I have done in these few months.

**Part 6 How did I build human relations with my seniors and co-workers.**

My senior is busy with her jobs, so she is not directly in charge of our job. My work usually co-work with the others interns, as we need to build the website and the project together, we need to communicate to each other very often, every day, we will talk to each other like what we are going to do, and the expectation of the today progress, and at what part we might need others help etc. As we are having a limited time, we usually just focus on our part, and then try to merge them together. And as we all need to use the same server, so we need to take extra careful when we need to delete the items, and if needed, we need to notice them that you want to reboot the server, like you would like to reboot the server at around 5pm, and everyone need to save their work before that time. Besides the job, we usually have the lunch together, during the lunch, we will share a lot of the story among us, not only related the jobs. Also, in our office, there has a room called demo room, inside the demo room, there has a lot of servers and a PS5. Usually, when we get stuck in our work or we just want to have some relax time, we will just go to the demo room and play the game together. Sometimes, our boss will also join us, and during the playing, we build the friendship with each other. Apart from that, in the event like said above, we need to go out and do communication and the set up together, everyone will have their own duty, and if some of us missed that, the other will help them.

**Part 7 General remarks on the Industrial Training**

Overall, I think this internship is valuable, as just sitting in the lecture room and reading the lecture notes doesn’t really the skills you need for the work. For the work, we need to be ready for learning things, as the world is keep changing, especially the Tech related jobs. In this internship, I learned a lot of the latest technology and some of the coding skills.