FACULTY OF COMPUTING
UTM Johor Bahru

SEMESTER 1 2023/2024

Subject : Technology Information System (SECP1513)
Section : 03
Task : DESIGN THINKING REPORT
Group Name : PENANG CENDOL GROUP
Group member :

NAME	MATRIX NO.
ADRIANA ZULAIKHA BINTI ZULKARMAN	A23CS0035
ALI REDA ALI ABOUSEADA	A23CS4003
KNAN FADI DAWARH	A23CS4013
KWAN JIN PING	A23CS0232
NG DE SHENG	A23CS0254
TANG JASMINE	A23CS0277

Table of Contents

1.0 INTRODUCTION

Artificial Intelligence (AI) represents a groundbreaking technological advancement, revolutionising various industries with its ability to mimic human intelligence. AI systems, often driven by machine learning algorithms, can analyse vast amounts of data, recognize patterns, and make intelligent decisions. The impact of AI is profound, transforming the way we live, work and interact. Its ability to process and interpret data at an unprecedented scale has paved the way for numerous innovations. In the context of social media, AI has ushered in a new era of personalised experience and content moderation.

In recent years, AI has become an integral part of many applications, with significant implications for social media platforms. AI in social media operates through algorithms that analyse user behaviour, preferences, and engagement patterns. These algorithms continuously learn and adapt, ensuring a dynamic and personalised user experience. The core mechanisms include natural language processing (NLP), computer vision, and deep learning, enabling platforms to understand and respond to user interactions effectively.

However, the insidious spreading of deepfake content around social media presents challenges for AI analysis, thus has the potential harm to society. Deepfake contents are AI-generated synthetic media that often involves the modification of data and manipulation of visuals and audios, which appears to be realistic, but fake content. As AI-generated media is currently new to the market, it's hard for audiences to identify its validity and authenticity.

This is why we decided to introduce our product - Beagle Content Filter. It is a machine learning product that targets AI-generated content, blocking it from being recommended to users and entering user's interface. Also, not to mention that the product also promotes a more sophisticated, comprehensive, in-depth personalised content recommendation to enhance users experience. Our product acts as a software plugin, adding advanced functions to the host program without altering the host program itself.

2.0 DETAILED STEPS AND DESCRIPTIONS

1. EMPATHY

Empathy is the stage where we should be aware of the issues that our respondent is facing. Young adults who could experience issues with social media were our target audience. We asked twenty participants to complete a survey and interviewed 1 student from UTM. We questioned the respondent about their personal information on the survey since it may help us figure out what issues they were having. We also inquired about the social media platforms they use and the kinds of inappropriate content they have seen.

2. DEFINE

Define is the step when we determine the types of inappropriate social media content and issues that the responder is facing. We determined their issues by comprehending and analysing their responses after reading the survey responses.

3. IDEATE

The ideate stage is the brainstorming process when we can come up with every potential solution for the issue statement. We've come up with several solutions for the problems. We have specified the goals and parameters of our program and have chosen the best platform for it.

4. PROTOTYPE

The prototype stage is when we pick the best concept out of all the ones that could work and turn it into a product. We began working on the prototype after deciding on a solution. We designed the interface, functionality, and interaction of the AI content detector. The prototype was made with A3 paper, marker pen and pencils.

5. TEST

Test is the stage when our product should be put to the test by our users. We value their comments since it may help us find the program's shortcomings and make additional improvements. Once the prototype was finished, we gave the respondent a tour of it and went over the capabilities of the AI content detector. In addition, we requested their opinion on the program.

3.0 DETAILED DESCRIPTION

3.1 Problem Statement

The Covid-19 pandemic has exacerbated the issue of misinformation in social media content, with algorithms often recommending content that violates guidelines. Other than mis-informational content, there are unethical content creators that profit with plagiarised content or AI-generated content without any endeavour. Incautious practices of recommending contents featuring controversial content are likely to harden people's views and steer them toward political radicalization. As for content creators, their benefits are compromised when content is recommended to an unsuitable target audience. The current system's lack of transparency and reliance on human reviewers has led to increased human errors and a decline in the credibility and client base of social media platforms.

3.2 Solution

The incapability of the current system to flag inappropriate content has led us to think of our solution : a product that functions similar to existing services such as AdBlock, AdGuard or even the spam box of Gmail. It will hide the content and comments that are inappropriate, i.e. violence, pornography, plagiarism etc. Other than that, it shall also be able to accept keywords from the user to customly filter out their disliked content. It shall be an extension that is installable on the Extension Store of browsers.

Upon installation, the extension will direct the user to our main page, where they will be prompted by several questions so that the AI can learn about the user's preferences. After the first setup, the main page shall act as an interface to allow for more advanced controls such as the switches for enabling the hiding categories of content and comments, or to add new categories.

For daily use, the user can open a small window via the extensions button on their browser toolbar to access simple and limited controls. Other than that, the user shall only interact with the extension when it detects content/comments that are suspected to be inappropriate. It will do so by highlighting the suspected content/comment for the user to decide whether to see similar content/comments in the future.

3.3 Team Working

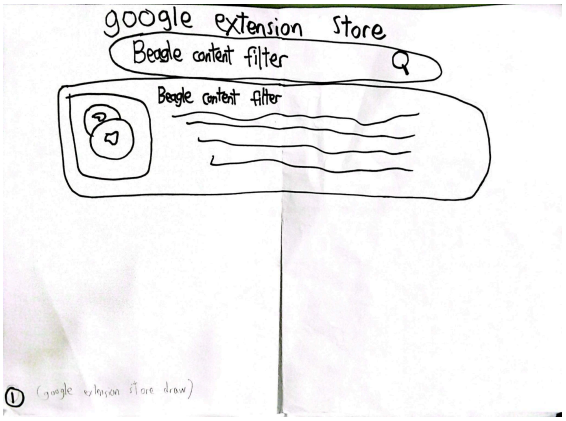
Ng De Sheng and Kwan Jin Ping were responsible for brainstorming and planning while Tang Jasmine and Adriana Zulaikha binti Zulkarman were the analysts of the survey forms and provided their opinions on the project. Ali Reda Ali Abouseada and Knan Fadi Dawarh were responsible for the survey. Overall, our team had some communication issues but eventually we smoothed everything out.

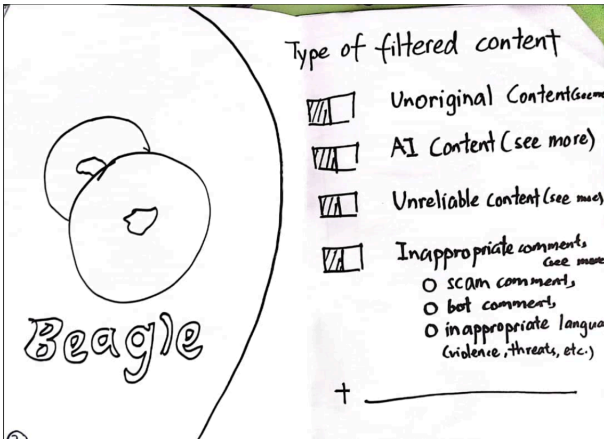
4.0 DESIGN THINKING ASSESSMENT

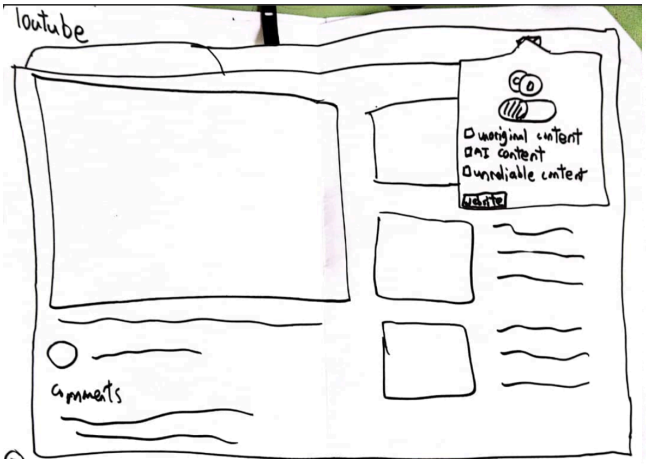
We are making our AI detector by using a design thinking method. First, we learn about the people who will use it through study. We've found the exact issue we want to fix, and are concentrating on what users need. Thinking up new ideas means coming up with many thoughts, not making quick choices, and picking the best ones. We have made our ideas real by creating simple models and improving them with helpful comments from users. Putting it into action means figuring out how to build, picking the best tech, and making a fair schedule. We have put our AI finder to use. We made a plan for spreading it, making sure the instructions are easy and we monitor how well it works all the time. Looking back at the whole trip, our review after putting things in place shows what we learned. This gives us ideas for future times and ways to keep getting better.

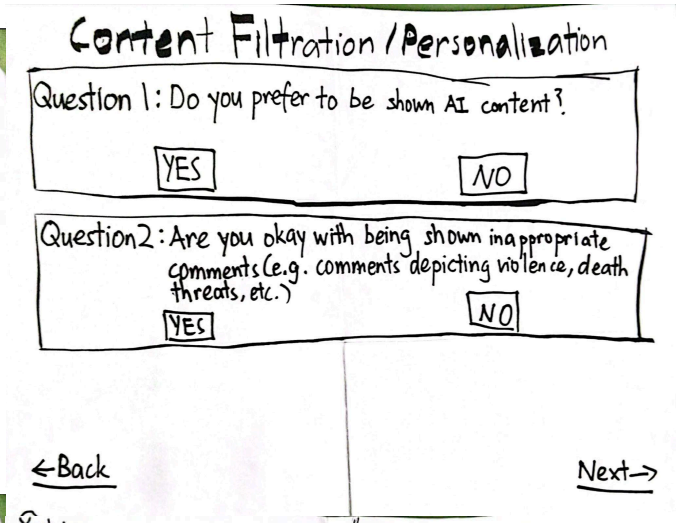
5.0 DESIGN THINKING EVIDENCE

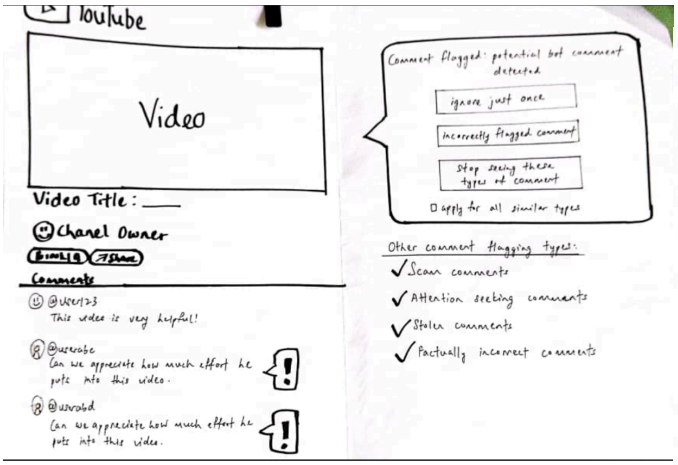
5.1 SAMPLE WORK

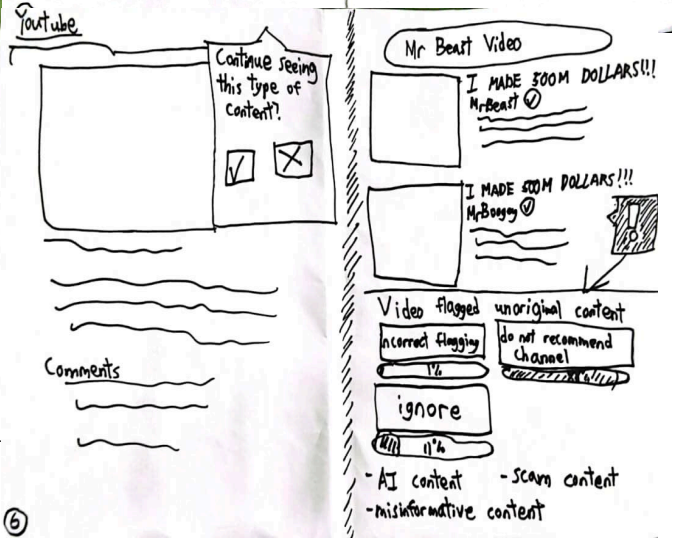








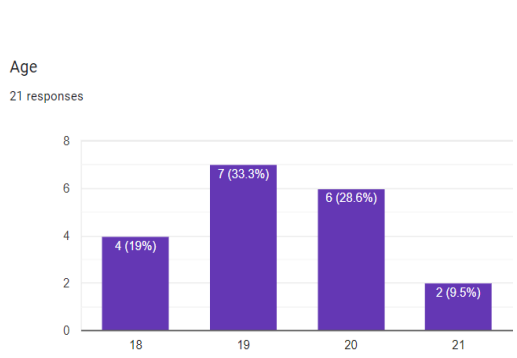


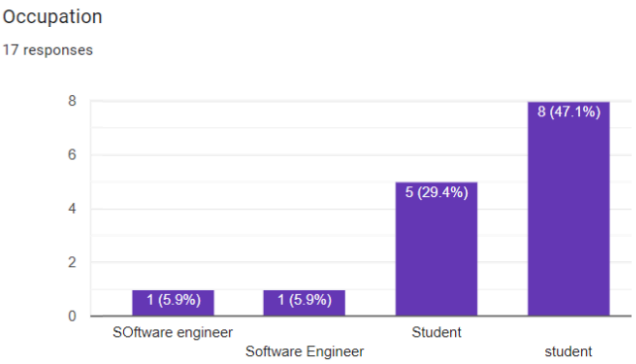


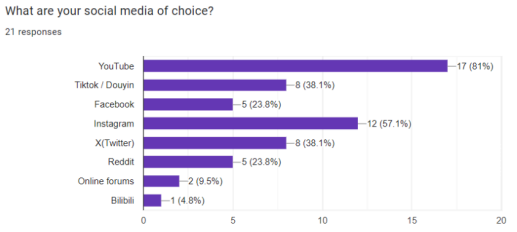
5.2 RECORD FOR EACH PHASE

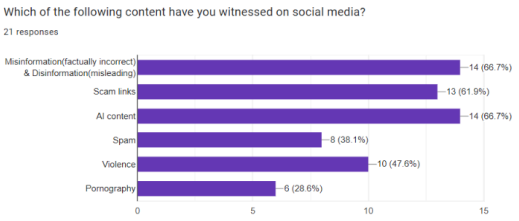
5.2.1 Empathy

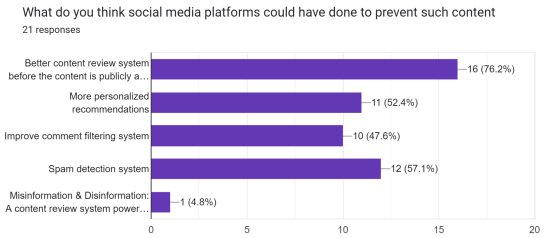
We have conducted some surveys on people who are frequent users of social media to ask about their opinion on the appearance of inappropriate content surfacing on social media. The questions for the survey and their choices of answers are shown in the images below:

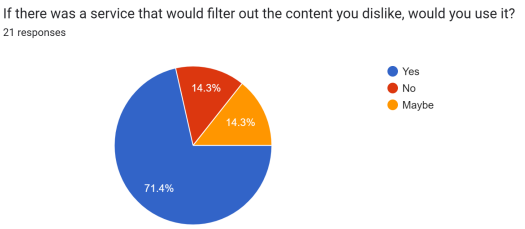












Furthermore, some interviews were conducted on a certain amount of people, and one of the interviewee's answers and questions asked are listed in the table below:

1.	Please introduce yourself (name, age, occupation)	My name is Havesh A/L Ravi, I'm 19 years old and I'm a UTM degree student.
2.	Do you often go online? What are the social media platforms that you usually browse?	Yes, I do go online and I usually use Instagram most of the time and also YouTube.
3.	Have you encountered any inappropriate content that is unsuitable to be publicly displayed and openly discussed? (plagiarised content/misinformation/ acts of violence or acts of racism homophobic/etc)	Actually I have seen many of the inappropriate content that you mentioned and if you ask me it's pretty disturbing and uncomfortable.
4.	Have you seen improper comments that appear in the comment section? (Ads promotion, comment bot, etc)	Very often, especially on Instagram because most other kids our age of our generation don't have any restrictions and no one supervises what they are doing. about the comment bots: yes I do see a lot of them especially in celebrities' Instagram pages, if you go to the comment section there's a lot of bots and it's most likely not under the supervision of Instagram's security system, and most of which are always scams.
5.	As for your concerns, how familiar are you with the role of AI in shaping online experiences?	I would say AI can be used as a standard for some cybersecurity systems like creating an anticheat security software that blocks unnecessary bots. I do often use AI like most other students, I use softwares such as chat GPT or Bing AI and if you ask me those AI softwares do help me in certain cases but not all the time, especially when I'm doing my maths homework I would use chat GPT to check whether the answers are correct or not, or when i don't particularly understand some assignment questions i would use chat GPT to check.
6.	(Video shown to interviewee) https://www.youtube.com/watch?v=ghDcOhx_TsI Have you ever noticed AI generated content online? If so, how do you feel about it? Has that affected your user's experience?	I recognize this YouTuber(in the video) because he's pretty famous, his name is Kwebbelkop and he used to record his videos in real life and not use AI and what I can see from the video is he uses AI instead of recording the video on his own. if you ask me, it's pretty bland and not entertaining because there's not much emotion in it. AI doesn't have emotions so when you use AI,

		you can't see expressions on his face like laughing, smiling or any emotion at all so it's not that entertaining.
7.	In your opinion, how does the integration of AI in content creation impact the authenticity and reliability of information online?	In my opinion, AI and AI based content creation are unreliable and dangerous at the same time, since a lot of people uses social media such as TikTok or Instagram as a source of information, and these information could be false or inauthentic, and people are quick to make decisions based on these information given instead of doing their own proper research through google or other means.
8.	In your opinion, are there any measures to ensure the responsible utilisation of AI to foster a healthy online environment?	I would say no, because AI is continuously evolving every year, and maybe in the future, AI can be used to do even more outrageous things than what we can do right now, and that is inevitable.
9.	What improvement/changes are you expecting? (Any specific concern? / change of algorithm / filtration system)	I would say it's going to be based on what the government wants to try to do with the algorithm. because instead of putting all that negative content, perhaps the social media developers can start making algorithms that recommend the younger generation about education or discipline. I also think a filtration system is very important; as I've said, there's a lot of inappropriate content online and the social media developers should take more initiative in strengthening their security. If this is accomplished, maybe there's still hope for a healthy environment for the future generations.

5.2.2 Define

From the results of the survey and the interviews above, a common problem could be identified:

The weaknesses and oversights of the current content filter in use by social media sites.

Majority of social media users, mostly who use youtube, have claimed to have witnessed the appearance of inappropriate content on the social media sites, such as misinformation(content that provides incorrect info), violent content, and/or pornographic content. Furthermore, scam advertisements or scam comments could be seen every now and again on the site, which could lead to many unsuspecting users being caught in the trap. The prevalent appearance of AI content seems to also be flooding their feeds, which hasn't been an issue until recent times. Even though social media sites already have their own built-in filter in their systems, it is evident from our survey that the system could potentially be outdated and require some remodelling to fit in with current times.

5.2.3 Ideate

In order to solve the problem mentioned above, we have come up with a solution to create a **more personalised and smarter content filter that checks the authenticity of the content as well as block inappropriate content**.

At first, we had the idea to make this a website where the content would be recommended to users on the website itself, however a website is much more harder to run and requires dedicated servers, not to mention our website would just be displaying content on other social media sites, which the user might find it hard or tiring and just prefer to use the social media site itself.

So we opted to make it a google plugin instead, where users could still browse their social media sites at their respective websites, but still have the functionality of our filter system.

5.3 Prototype

5.3.1 Functionalities of our prototype

a. Detect content that contains misinformation

One of the downsides of the current content filter that is being used by social media sites in our current times is they are unable to discern truthful content to content containing misinformation, and still requires the users to manually report those types of videos to initiate a video takedown process. Overall, this process is long and unreliable, as multiple reports need to be made for a human moderator to take potential action, otherwise it goes through their AI content checker, which could still lead to either no action being taken or even abused to false strike a creator's content. With our content filter, it automatically runs through some simple google research using AI to determine if the information in the content is reliable or not.

b. Detect scam content

With scam contents, people fall victim to the scams before knowing it's a scam. With our filter, the AI can do its research to determine if the contents of the video is a scam or not before the user indulges deeper into the scam itself. Even though social media sites such as youtube do implement a comment system to let others share what they think of the video, there are multiple ways to put down the comments, such as bot comments praising how reliable the video is, not to mention the creator themselves have the option to enable or unable comments on their videos, so the comment system is evidently not effective enough.

c. Detect unoriginal content

The filter also has the option to filter out unoriginal content, like videos which are 1 for 1 replications of another popular video. The filter, using its AI, would compare the videos with their thumbnails, duration, content, description, upload date and more to determine which is the unoriginal video and filter it off the user's recommendation if they want. As it stands, Youtube does not have a proper originality filter system, so clone creators could exist and steal views from unsuspecting viewers. With our filter, the views that could have gone to the clone creators would be funnelled to the original creator instead.

d. Detect AI content

As of the 21st century, AI content has been slowly emerging with the rapid advance of technology in this day and age. Therefore, social media sites are slowly seeing a plethora of AI content being pumped out by thousands of creators. These videos take almost zero effort to make, yet the creators of those videos earn just as much as an actual hardworking creator does. So, our filter is able to detect if the content is made by AI or not and give the user the choice to continue seeing this type of content or not. With less AI videos being recommended, more views can be funnelled towards the actual creators that put actual heart and soul into their videos.

e. Better content filtration system

Due to having a large user base and multiple system implementations, social media sites such as youtube are not able to properly filter out all the inappropriate content out of each user's feed with 100% efficiency. With our filter which is more personalised and less system implementations, we can filter out the content with much more ease and efficiency to boost the user's entertainment value on their favourite social media sites. The filter is also able to filter out inappropriate comments such as comments that depict violence or contain scam links.

f. Detect comment bots

One of the more prevalent happenings in youtube's comment section are the comment bots, which often reply under a youtube video to boost the engagement of the video and have the algorithm promote the video to more people. Sometimes these comment bots also comment on popular youtuber's comment section to get people to click on their profile to check out their content by spamming the same comment on every single comment available. Although doing so is not illegal, it is extremely immoral, so our filters are able to detect and hide these types of comments from the users.

5.3.2 Comparison with other similar existing filters

Below is a comparison between our filter and some already existing filters that have similar functionality with ours.

Functionalities	**Our filter**	**F.B. Purity(google extension)**	**Youtube's own content filter**	**Ad blockers**
Hide content based on user's preference	Yes	Yes	Yes	No
Can be used on multiple social media sites	Yes	No	No	Yes
Blocks ads	Yes	Yes	No	Yes
Flag inappropriate content on social media sites	Yes	No	Yes	No
Hide comments based on user preference	Yes	No	Yes	No
Detect and hide content containing misinformation or unoriginal content	Yes	No	No	No
Detect and hide scam links	Yes	No	No	No
Detect and hide comment Bots	Yes	No	No	No
Detect and hide AI content	Yes	No	No	No

TEST

After the prototype was finished, it was presented to the class and we described the functionalities of the prototype. Some hard hitting questions were thrown at our prototype, but the questions were answered with ease. Overall, not many people find the prototype problematic and would be considered a successful first testing.

6.0 REFLECTIONS

NG DE SHENG

This group project has led me to understand the process of design thinking in a very deep sense, which made me think and create complex ideas to help fix the issues related to the project. It has taught me to consider each and every possibility to ensure the best results and to find the perfect solution for the problems to save time and energy.

KWAN JIN PING

The project has given me a better understanding of the design thinking phase of a product or service, including the brainstorming, data collection and producing a prototype. It also highlights the significance of having a good team for problem solving.

KNAN FADI DAWARH

This collaborative project has provided me with a profound insight into the intricacies of design thinking. It encouraged me to generate intricate ideas, carefully analysing each aspect to address project-related issues effectively. The experience taught me to explore all possibilities systematically, striving for optimal solutions to save time and energy.

ALI REDA ALI ABOUSEADA

The project revolutionised my teamwork, critical thinking as well as public speaking skills. The fact that different points of view strengthened our work gave me confidence. It further developed my understanding of design thinking by emphasising the need for a strong team in effective problem-solving. Overall, a change experience for my future projects.

TANG JASMINE

Through this project, I have gained a lot. From forming an international group, starting off with a vague project idea, conducting market analysis, visualising the prototype, presenting it to everyone, and finally ending with a final report. The main core of design thinking is to solve human problems. Besides, I also acknowledged the importance of teamwork.

ADRIANA ZULAIKHA BINTI ZULKARMAN

I learned a lot from this long-term design thinking project and now have a better knowledge of design thinking. My understanding of tolerance grew, and I realised how crucial it was to be in constant contact with the lecturer and other members of my group. To sum up, I reinforced and accomplished these important aspects.

7.0 TASK FOR EACH MEMBER

Members	Task
TANG JASMINE	Intro & design thinking assessment
ADRIANA ZULAIKHA BINTI ZULKARMAN	Detail step and descriptions
KWAN JIN PING	Detailed description
ALI REDA ALI ABOUSEADA	Design thinking evidence (5a)
NG DE SHENG	Design thinking evidence (5b) & reflection
KNAN FADI DAWARH	Task for each member