RMIT UNIVERSITY

Introduction to IT

COSC2083_ASSIGNMENT 2

Vu Quoc Gia Quan

Dang Hoang Anh Khoa

Mai Duc Minh

Nguyen Thu Thuy

	Member	ID	Name
Team Members	1	s3870273	Nguyen Thu Thuy
	2	s3681447	Mai Duc Minh
	3	s3927120	Vu Quoc Gia Quan
	4	s3836606	Dang Hoang Anh Khoa
Word Count	? TNR 12pt		

Table of Contents

1. Team Profile

Team Name

The team name that our team has chosen is ***Memeiacs***. The origin of our name came from the prefix “meme” that formally meant “an element of culture or system of behavior that is considered to be passed down from one individual to another”. Since the beginning of the 21st century, the word has adopted a new meaning to be “a humorous image, video, piece of text, etc. that is copied and spread rapidly by internet users” (Oxford Dictionary n.d.). Team members at Memeiacs are very up-to-date with newest memes and are always eager to lighten up stressful working hours for others. We are aware it will come to the attention of potential employers and we hope they can see our goofy, bright personality through the name.

Personal information

One paragraph per person, including name, student number, background, hobbies, IT interest and IT experience. This should also include your team’s chosen name.

- ➔ **Project Leader:** My name is Mai Duc Minh and my student number is s3681447 and I am honored to be elected as the project leader. I have studied Software Engineering for 6 semesters in RMIT. I usually play video games and listen to music in my free time. Sports are not counted as hobbies for me because I play sports for physical and mental training. I have some IoT experience with Arduino and household networking. I am interested in Python and have a long term plan which aims at machine learning in the medical aspect.
- ➔ **Secretary:** My name is Vu Quoc Gia Quan, and I have the student ID s3927120. I am currently our team's secretary. I am an IT major, and this is my first semester at RMIT. My hobbies include creating small DIY projects, reading manga, researching random topics, and so on. I do not have much IT knowledge because this is my first time learning programming and IT.
- ➔ **Project Designer:** They call me Ethan and my full name is Dang Hoang Anh Khoa. My student ID is s3836606. I got the designer position for my team project and I'm happy with that although I'm the oldest member. In my leisure time, I like listening to music, playing video games and learning new things from Youtube or external sources are my hobbies. Especially coding something fun such as Microbit and websites. This is my second semester in this school so my academic knowledge is limited. I've experienced some interesting languages such as Python, Css, HTML and a bit about JavaScript.
- ➔ **Project Manager:** My name is Nguyen Thu Thuy, my student number is s3870273 and I am the project manager in the team. I am a Digital Marketing major and this is my fifth semester at RMIT. My hobbies include reading, painting, and writing stories. I like coding although I am still in the learning stage but I am familiar with Python and a little bit of other coding languages.

Team Profile

Mai Duc Minh

Both my personal and career life are indicated accurately in the Myers-Briggs test that I am very goal-oriented. "Practice makes perfect" - My strategy is to set a high goal at first, make a concrete plan and push myself hard enough to follow little by little every day. However, the test also discovered a long-standing difficulty of mine, in which my work style would cause a lot of problems to people who are sensitive to criticism since I accidentally react strongly to the work to obtain best results. I am quite a curious person to every phenomenal thing around me so it is unsurprising that the Learning Style Test determines me as a visionary learner. To come up with a clear and detailed teamwork plan, I usually sketch my ideas on paper and make arrangements so that all the thoughts are logged to make sure I do not forget anything. The Big Five Personality test serves as a confirmation of the accuracy of both tests above and comprehensively reflects my personal character, working style

so that other teammates can understand and coop better. I am willing to learn from others (decent Openness score), I am less affected to external factors (highest Conscientiousness score), I am very disciplined (lowest Extraversion score) but I am not a good communicator and easily prone to overwhelmed when the plan is off-tracked (high Neuroticism score).

Vu Quoc Gia Quan

The Myers-Briggs and a personality test said I am outgoing, analytical, and quick to assess people. But I am strict and do not like unusual answers. The Learning Style test shows that I am a visual learner. Keeping this in mind will help me collaborate with others. Some of my team's ideas may be rejected because they do not meet the textbook's criteria. I need to be more open-minded, but I also need to be more selective with my consumption. Despite my flaws, I enjoy working in a team environment. I can still help others build their own ideas and answers. My ability to elevate ideas and solutions can benefit the entire team.

Dang Hoang Anh Khoa

Although all online tests did not show exactly what I am, I must recognize that I am an introverted person who, once in the blue moon, shares about myself with others in every community whenever I trust them. I am interested in learning knowledge throughout what I witnessed around my life and from external sources. I always accept practicing sections that can help me improve my academic performance such as programming and dealing with code. I think the learning style test convinced me exactly when saying I am a visual person. Beside those tests, I got a high score on the creativity test. To sum up, I would say I can take responsibility in my team as the designer is the exact choice and I will finish my work efficiently.

Nguyen Thu Thuy

The famous Myers-Briggs test convinced me that I am a campaigner, specifically ENFP-A. That means that I am more extroverted and intuitive, and my nature involves more feelings and my type leans towards prospecting. While being quite indecisive, I am assertive with my identity which makes me more confident in my abilities. The second test about the learning style shows that I am a visual learner. The last test was one I chose, the Big Five Personality test, which granted me four factors with an average of 78% (Openness, Conscientiousness, Extraversion and Agreeableness) and one - Neuroticism - with 17% meaning that I do not often feel negative emotions such as anger, anxiety, emotional instability, etc.

Ideal Jobs

Category	Minh - Embedded Software Dev	Khoa - Front end developer	Thuy - IT - Software Architect/ Technical Lead	Quan - AI Engineer (Python, C++, C#)
Skills	-Any approved staff must have background in Electronic, Mechatronics or equivalent -Practical embedded software development is necessary -required to do requirement analysis and play the role of quality assurance/ quality control tester: design and implement the test cases to find bugs and analyse the failures	-Must have at least one years of experience in the same position as front-end developer -Very good knowledge in JavaScript, and CSS3 HTML/HTML5 - Experienced working with more than one JavaScript frameworks such as: React.JS, Angular.js, VueJS, jQuery etc	-BS, MS, and/or Ph.D. degree in computer science, software engineering, or a related technical field -3+ years of experience in the role of Software Architect/Technical Team Leader -Strong knowledge of OOP, design patterns, multithreading/concurrency programming, data structure, and algorithms -Experience defining the architecture of cloud-deployed	-Bachelor's degree or higher in Computer Science, Information Technology or equivalent -Working experience in programming in Python (preferred) or C++ or C# -Working experience in AI systems, like recommendation engine, data preparation, feature selection, target metrics, model evaluation, validation. -Good at study/search skill, communication skill

	-have good English communication skills		applications (AWS/Azure preferred) -Good English communication	-Can read/write English document
Earnings	-13th-month salary bonus + attractive performance bonus+ annual performance appraisal **-**100% monthly basic salary and mandatory social insurances in 2-month probation	-Competitive salary -Monthly bonus	-Average 15 Monthly Base salary per year	- Review for promotion and annual salary and bonus

Benefits of working environment	-15++ days of annual leave + 1 day of birthday leave -Flexible working time -Lunch and parking allowance -Free in-house entertainment facilities and snack	-Opportunities to work onsite -Good working conditions and environment -Company party every month -Company trip 2-4 times per year	-Meal Allowance -Yearly Medical Checkup & Team building activities -Professional and Transparent Working Environment	- Dynamic working environment, rich cultural activities, modern and civilized working conditions - Participating in professional and skill enhancement training courses organized by the Company

Based on what I see in the comparison chart, everyone of us will most likely receive a different job in the future . Each job has its own set of requirements for applicants. Some will appear vague, yet others, such as Thuy's ideal job, will have an incredibly detailed description of what it will require. Most other jobs require at least a few years of experience in the same role or the experience of someone who has worked in that position for many years. Every career requires a profound awareness of its own branches as well as specialization in what the work requires. All four of our jobs pay a decent wage, come with adequate insurance, and, in a few cases, include a monthly stipend for employees. All of the jobs advertise a variety of recreational activities as well as a professional working environment.

Minh's profession provides him with flexible working hours as well as free in-house entertainment and snacks. Khoa's profession provides him with the opportunity to work on-site abroad, allowing him to gain significant experience that will serve him well in the future. Thuy's job is most likely the most challenging of all of our teams' ideal jobs. Her work requires a lot of prerequisites just to

get an interview and a chance at the position. Quan's job appears to be quite simple based on the job description and requirements, but whether this is true or not is debatable.

All three of Quan's, Minh's, and Khoa's positions are likely to be junior engineers; however, Thuy's position is likely to be higher, such as senior engineer or technical lead. Because we all work in different areas of information technology, our career paths are most certainly going to diverge. We shall have several self-development plans to meet our particular demands and to obtain the necessary certification to get the job. We will have to develop on our own, but with the assistance of RMIT. Despite this, I believe we will build an excellent team due to our diverse knowledge of several IT fields. It will assist us in achieving pretty good results in various areas, and we will be able to learn from one another. If everyone studies diverse things, the combination of our expertise will be of tremendous use and will substantially assist us in not only our current but also future undertakings.

2. Tools

Our group's published website can be found in this link.

Our team's website is a shortened version but still displays the full information of the report. As you can see, when we access the website's url, what appears is the homepage including a navigation bar containing the group's self-designed Mememiacs logo; a Drop down box to choose to display information of each group member; the center shows the Project section to navigate to the most dedicated part of the team - details about the idea of building an image comparison tool; Other sections contain the content of IT Work and IT Technologies sections; On the left side of the navigation bar shows the references used in the report.

The content of the homepage shows the meaning of the group's name as well as the motto. When you click on the image to the left of the motto, you will navigate to the Git repository containing the website's content.

When navigating to another page of the website, the navigation bar is always on the top to make it comfortable for the user to navigate different pages and completely return to the homepage just by clicking on the Mememiacs logo which is located on the left side of the navigation bar.

Regarding the content of the pages in the Profile tab of each member, it displays the member's role in the group as well as the personal information that is verbatim from Assignment 1 about interests, test results of online tests as well as statements about its accuracy.

Regarding the content of the Project tab, the idea of building an image comparison tool that was approved by the group as a whole and described in Assignment 2 is more specific than its version in Assignment 1. In particular, not only has the ability to compare images, but also the ability to compare images. The group also describes how to design user interfaces for software, design mock designs for the interface so that readers can specifically visualize the group's ideas as well as shaping the idea of how to interact with the software functions.

The Git repository of the Assignment 2's contents can be found in this link.

This repository contains other content while the team was working on Assignment 2. Specifically, there is a Meeting Minutes folder containing all the details of what the team has discussed or reported progress on the assigned tasks on every meeting. There is also an Interview Content folder containing two documents: one is the questions and the other is the original answer from the interviewee. Moreover, there is an Asm2 Mock Views folder that contains all the mock design of the user interface of the project idea under both PNG file and PDF file.

The Git repository of our group's website can be found in this link.

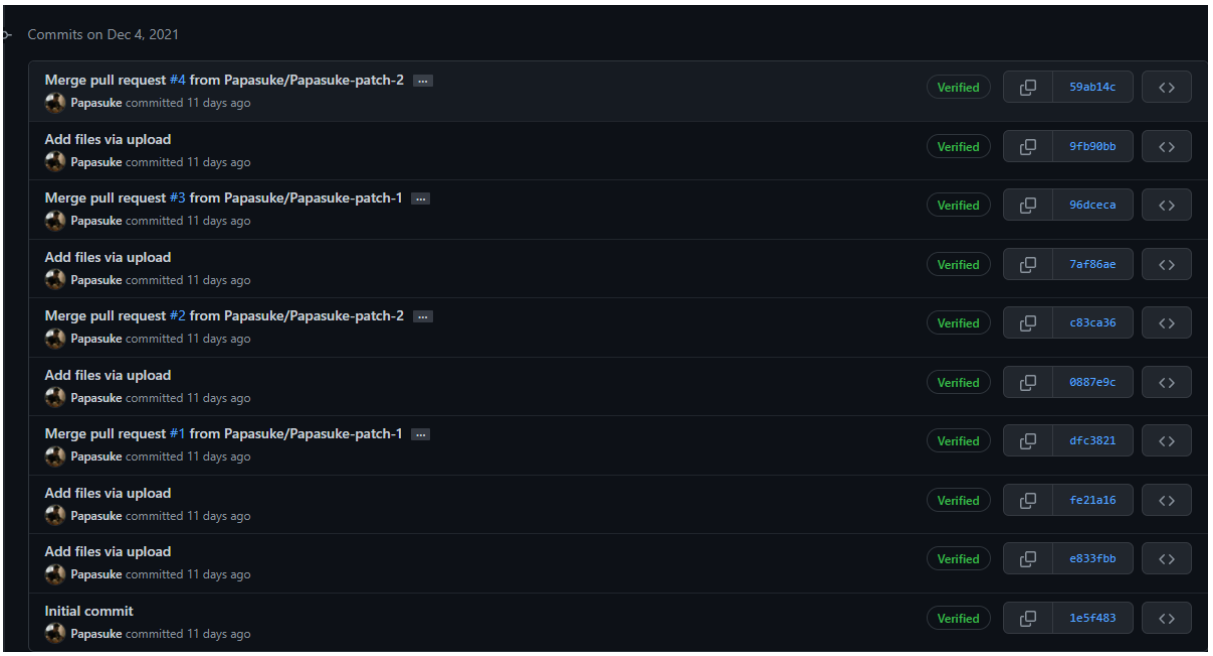

Figure 1. Commit history (1)

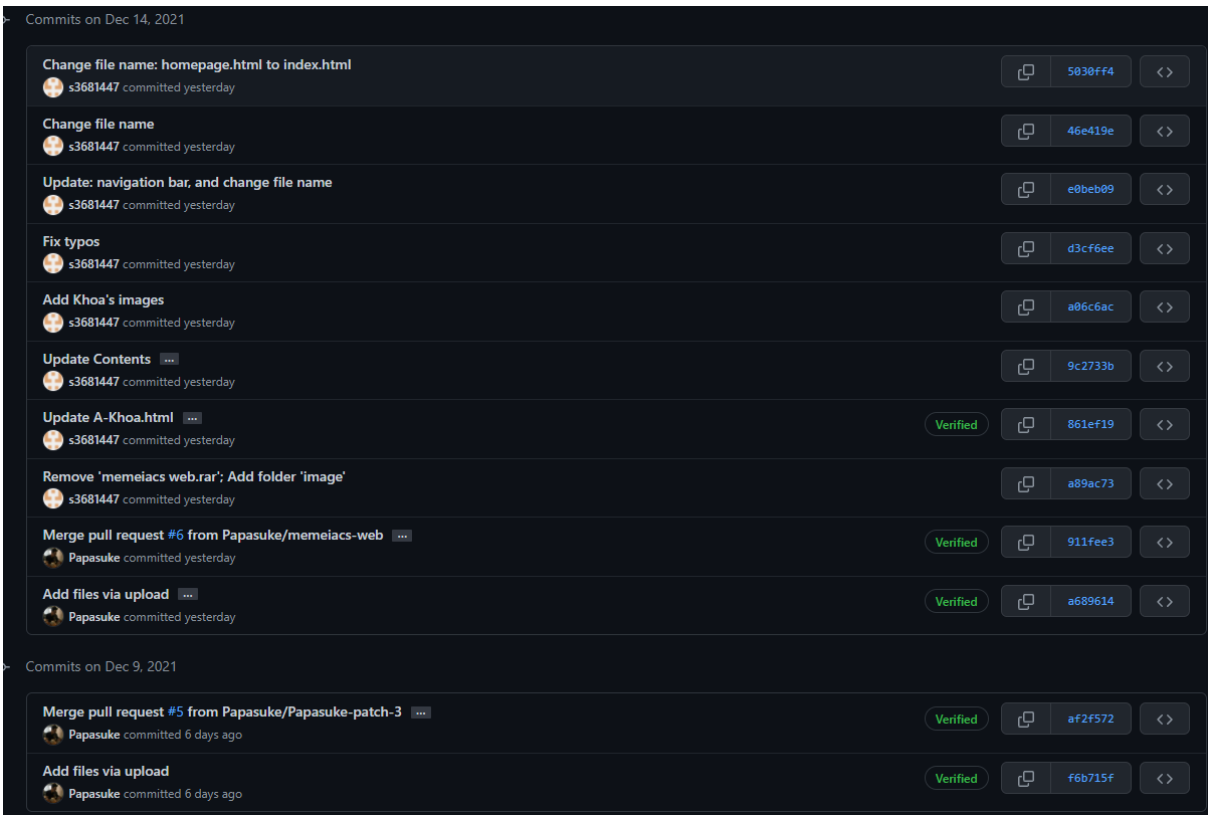

Figure 2. Commit history (2)

The GitHub audit trail images above shows how the team have cooperated to build the website together. As you can see that there are only 2 people who committed all the work but actually every member has actively contributed to the website design. First of all, the way our team works needs to be clarified to avoid unfortunate

misunderstandings. Before starting Assignment 2, the whole group had a meeting and agreed that each member would be responsible for each part of the assignment. That is, each member acts as both a follower and a leader to lead each other. Specifically, each person must prepare their own personal information as soon as possible to complete section 1. Project designer will be responsible for website design, Project manager and secretary will be responsible for sections 3 and 4. Therefore, the group leader will be responsible for section 5. The role of the person in charge is to learn and lead the remaining members together to complete the section they are assigned. Going back to everyone's contribution to the website, the project manager and the secretary are the contributors to the majority of the website's content, then the project designer is the one who builds the website according to the team's agreed upon design and also who has directly committed to the Git repository. Moreover, in the last 2 weeks before submitting the assignment, the project leader participated in supporting and correcting the shortcomings of the website and also finalised and published the website. Therefore, the cooperation was done harmonically.

3. IT Work

The IT professional that has agreed to be interviewed by our team, his name is Nguyen Nhat Trung, currently working as a data analyst for the banking industry. As the world is becoming more digitized and people are quickly switching to cards instead of cash payments, the banking industry is growing to a new height each year. Trung builds and maintains reporting systems with Microsoft SQL Server Reporting Services (SSRS) and Microsoft Power BI Report Server and provides support to the end users.

The department that he works in interacts with the sales center to understand customer needs and help them improve their sales performance as well as with the internal infrastructure department who supports them by keeping the servers running the systems. The most time spent on his work is in analyzing and building data for the reporting system, and the most challenging part of the job is designing and implementing a new system as it requires all technical aspects and timeline. Trung believes that the essence of the IT industry is to provide the end users services with the best quality and improve people’s lives.

Regarding his personal experience in the IT industry, Trung shares that he didn’t choose his career but instead it chose him. He has always been interested in programming since high school. For someone coming into the industry, he suggests reading newspapers and catching up on IT trends such as the future with AI. Trung has been a part of his company for over 5 years and is seeing himself retiring with it.

Trung works from 8AM to 5PM everyday and finds 40 hours a week enough. His favorite tradition/culture in the company is the rewards when one utilises creativity and creates something new, whether it’s a design or a system or a new solution to

improve efficiency. Our IT professional is an ESTP who is known as the doers and enjoys listening to music, playing MMO strategy games as well as tennis. We asked him what advice he would give his past self and Trung said to “choose a certain career and do it best”. Trung has an interesting work-life balance and he’s doing great in the IT section of the banking industry. He was an interesting person to interview and get to know.

4. IT Technologies

Cybersecurity

By looking at Cyber Security Degrees (2021), we can see that there have been numerous new and innovative technologies developed. We have been successful in developing a behavioral analytics system. This will allow us to better predict and respond to threats much more quickly than we were previously able to. We can even prevent an attack before it has a chance to start. This can be used in a variety of different applications. We have improved the efficiency with which we create behavior profiles compared to previous years. A subject's daily routine, their activities, and other factors can be used to create a behavioral profile of them. We will be able to predict when someone is in danger or engaging in unusual behavior, and we will be able to intervene and assist them before something disastrous occurs. This can be used to predict when a cyber attack is going to occur, which is extremely useful. We can gather hacker information and habits, and then predict where and when they will carry out their attack based on this information and habits. What we are looking for is the IP addresses of the hackers as well as their objectives and methods of hacking. We can prevent the attack from taking place if we keep this information in mind. And make algorithms that will take advantage of this information and help us predict the hacker’s behaviours. This is another application of a behaviour analytics system. Another thing we've been able to accomplish is Virtual Dispersive Networking, which is a type of distributed computing (also known as VDN). This allows us to divide our information into multiple parts, encrypt them, and route them through a variety of different protocols at the same time. The attack on information will be significantly more difficult as a result of this. In addition, because the data is dispersed and encrypted, even if the hacker manages to get past the security wall, they will be unable to obtain much data using traditional hacking methods. We can store information in a much more secure manner if we use this method. Another thing that we have been able to accomplish is to use smart grid technology as a security measure. With Padlock, we can establish a more secure communication channel between the central station and field devices, making it less likely that we will be targeted by an attacker. I believe that in the future, we will be able to combine all of these new technologies to create a more secure system, one that will be more difficult to attack in the traditional manner. We will be able to incorporate artificial intelligence into cyber security to assist us in the development of

a more secure and strict cyber security system. The process of developing a security system will take less time and may even be more effective with the assistance of artificial intelligence. Hardware development and personal skill development have both advanced in recent years, which has made all of this a possibility. Our ability to obtain more powerful hardware than ever before will allow us to pursue a much wider range of routes and development paths that would have previously seemed impossible. We learn from our mistakes and strive to improve in areas where we fall short. Along with this, we have improved our personal programming skills and created a security system.

As a result of all of these developments, we will live in a more secure world. Our information can be safely stored on the Internet without causing us any problems. In the future, we will have a much stronger justification for storing our information online rather than in hard copy forms. The people, as well as the hackers, will most likely be the ones who are affected the most as a result of this. People will be more willing to have the information available online if they do not have to worry about their belongings being lost or being stolen. A personal security system that performs as well as an industrial security system will be available to the general public. People's lives will be made significantly easier as a result of this. Hackers, on the other hand, are the ones who will bear the brunt of the consequences. They will be less successful in their hacking and information-gathering efforts for malicious purposes. As a result, there will be fewer hackers, and those who do remain will most likely be the best of the bunch. Hackers, on the other hand, can adapt their methods and approaches to the new security system. This is the essence of hacking; they will adapt to the situation and look for a weakness in the system to exploit. Additionally, we will see a decrease in the number of white hats, or those who hack into a system and then sell their method back to the company they hacked into. This could be a negative development because we will not receive as much information about system flaws and errors as we did in the past.

In all honesty, I do not believe that this would have a significant impact on my life because it is a problem that other important people must deal with. I might be more willing to give my information to companies and corporations since they are less likely to be hacked, and so my information will not be gathered illegally. This will allow me and my acquaintances to have a more positive experience when working with a large corporation. When it comes to how our information is stored and processed, we will have a more secure feeling about it. However, in terms of daily habits or activities, I do not believe that there will be any significant changes. The latest technologies or security systems that have been developed personally do not matter to me. I only choose to use them after they have proven to be more stable and reliable in their performance.

Artificial Intelligence (AI)

As can be seen in Analytics Insight (2021), significant progress has been made in the field of AI by 2021. The use of AI in the creation of behavior profiles to predict the possibility of what might happen next is the most notable. It is used to predict and respond to threats as soon as they appear, rather than waiting until they have occurred. This process will become more automated as a result of AI, which will save us a significant amount of manpower and time. We can also see from Appen (2021) that many companies and businesses have begun to incorporate AI into their applications. It can be found in both external and internal applications of various kinds. Despite the fact that the COVID-19 pandemic was a truly devastating event, it provided an opportunity for AI and machine learning to advance much more quickly than they had previously. We have been able to incorporate AI into a wide range of daily and professional activities, allowing them to become more automated. This year has demonstrated both the potential applications of AI and the rapidity with which the technology is evolving. We've never used them as much as we do now. Our shortcomings remain, despite all of our advantages. Finding and developing a machine learning model continues to be a difficult task to complete successfully. Finding precise, high-quality data to train the model is still a significant barrier to the advancement of AI. If we want to really push the capabilities of AI, those are the problems that we must face head on and find solutions for all of the current ones. One of the most straightforward things that AI has been able to accomplish is using our personal data to provide us with suggestions and advertisements based on what we like or are interested in learning more about. Another application is the automatic driving system in an electric vehicle. Although I do not foresee this happening in the near future, I do believe that many aspects of our lives will be automated, making life a little bit easier. We could likely have AI products that will assist us in studying by going over our previous study history and assisting us in choosing the right path that will be the most beneficial to us. The development of a more refined automatic driving system for electric vehicles is yet another possibility to be considered. With enough time and data training, we can create a system that can even navigate the streets and traffic of Vietnam – very complicated and, at times, chaotic streets with a lot of traffic – and get lost in it. For this type of development to occur, there are two primary reasons for it. The first of these is the advancement in technology. We have been getting a lot more powerful computing hardware in recent years. This has enabled us to obtain more accurate information and simulations in order to collect data to train the model. The second reason is that we have been able to improve our AI programming methods and approaches as a result of this experience. The programmers have improved their abilities over the previous years. We can learn from our mistakes and apply what we have learned to future tasks. In order for AI to make such a quantum leap in 2021, an opportunity had to present itself; in this case, COVID-19 represents that opportunity. A strong desire for AI usage has arisen, which has resulted in the development of AI becoming increasingly successful.

As a result of all of these breakthroughs and advancements in AI technologies, we have been able to do our jobs more efficiently and make our jobs a lot easier. We've also been able to put them to use in our everyday lives. It makes our lives more convenient and allows things to happen much more quickly than they did previously. With the assistance of AI technologies, we can engage in more productive activities or complete more work. It can assist us in automating tasks that we previously had to complete manually. Unless people learn how to adapt to the new technology, it will not result in their losing their jobs. Although it can be extremely beneficial when used correctly, it can also cause people to become lazier if they become overly reliant on it. This process of automation using AI has the potential to render former technologies obsolete, such as jobs that required manual labor in the past. Still, new jobs will be created to fix and manage this new AI technology, which will continue to be a constant source of growth.

I don't believe it would have a significant impact on me because I always wait for a technology to become stable and dependable before putting it to use. However, because I intend to work as an AI engineer in the future, I will need to become accustomed to testing and implementing new AI technology, methods, and models. If I want to get the most out of AI development, I'll have to learn to go with the flow. I also have to think about and learn a new coding method, as well as find new training data for my AI. With regard to the members of my family and friends, it is unlikely that they will experience any significant changes in the near future. They will hold off on utilizing this type of technology until the appropriate time has come along.

Blockchain and Cryptocurrencies

Blockchain - also known as Distributed Ledger Technology - is a shared database that makes any digital asset unalterable and transparent through the use of decentralization and cryptographic hashing (George Strawn 2019). A simple analogy would compare blockchain to a Google Docs, where a shared document is distributed instead of copied or transferred, which creates a decentralized distribution system that allows every party access at the same time. Blockchain is one of the most important technologies of the century as it makes everything safer and prevents fraud by being transparent. Blockchain is known for being a secure technology, where people who are skeptical of one another can share valuable data because it uses complex math and innovative software rules that are extremely difficult to hack into (Mike Orcutt 2018). Blockchain, to be simply explained, is a chain of blocks, each consisting of data, hashcode and hashcode of the previous block. A hash is a 256-bit number created from a 32-bit nonce - also known as a number that used once, extremely unique like a fingerprint. Since each block contains the information of the previous hash, the first block ever does not have that information and is called the genesis block. If a hacker changes the information in one of the blocks then the hash changes, and the validation of the next block will fail, so will the next block and next blocks. But it can be saved with “proof-of-work” which

slows down the creation of a new block. If a hacker tampers with one block, they have to recalculate the proof-of-work for all the following blocks. Not only that, blockchain uses a peer-to-peer network that allows people access to the whole chain on something called a node. A node is any kind of electronic device that maintains the network and verifies each batch of network transactions called blocks. When a new block is created, the node verifies the block and adds it to the chain (George Strawn 2019). Thus, it is tedious and complicated, making blockchain very hard to hack into. One other concept relating to blockchain are miners, those creating new blocks in a process called mining. One has to go through around 4 billion nonce-hash combinations to be able to successfully mine a block. When that happens, the miner is rewarded financially and the block is accepted by nodes on all networks. Blockchains make cryptocurrencies available. Cryptocurrencies are digital currencies that use encryption techniques to control the monetary units and verify the transfer of funds which makes it nearly impossible to counterfeit. One of the most famous cryptocurrency is Bitcoin. Currently, cryptocurrencies can be traded using the blockchain technology worldwide, and investors make profit from the difference between when they buy and when they sell. Aside from that, NFT - also known as non-fungible tokens - are quickly becoming the future of blockchain technology. Such as collectibles that exist in real life now can be digitized and profited from. Soon, our digital identity will also use the blockchain system to make it unique to the user and a lot more secure, replacing username and passwords right now. Blockchain essentially acts like a cybersecurity system and is expected to be implemented in companies in the future (Gayvoronskaya, et al. 2020).

The possibilities using blockchain are endless. It can make data storing and transferring secure, helping companies save millions of dollars and CEOs fear of losing valuable intellectual property. Blockchain usually goes alongside cryptocurrencies, which is available to the public and mainly traded by investors. Following the cryptos, there are NFTs that are on the rise and are making the future of blockchain, all of which are safe and secure in this network. The people mostly associated with these technologies can be anyone. In the future, blockchain will replace identity recognition systems like username and passwords that are used today.

As someone who trades cryptocurrencies and stocks in my free time, blockchain is a very fascinating concept for me. It is currently used to make networks safe and eases my worries about being hacked. A lot of my friends and family also trade on the side and are very familiar with blockchain and cryptocurrencies and even NFTs, the impact of this technology on us is very big and if it were to fail either because a hacker successfully infiltrates and tampers with the blocks, my investments could be gone within seconds. It matters a lot to me that blockchain works correctly and I feel safe about its security.

Natural Language Processing and Chatterbots

Natural Language Processing or NLP for short used to be a subfield of linguistics and transformed to a branch of computer science, significantly using artificial intelligence that allows the computer to comprehend text and spoken language much like a human can. Another way to describe it is the automatic manipulation of language by software. Combining statistical machine learning, deep learning models, NLPs can fully understand speech and text with the speaker or writer's full intent or sentiment. NLPs can get computers to translate text, respond to commands or summarize lange text rapidly in real time (IBM Cloud Education 2020). In real life, NLPs can be seen everywhere such as GPS, Alexa, Siri or other digital assistants, speech-to-text recognition software, and so on. One of the growing uses for NLP also comes from chatbots, which are becoming increasingly present in our everyday lives. Businesses use chatbots to facilitate the interaction with customers and provide answers and services quicker, leading to a more profitable business. Chatbots are computer programs that are designed to communicate with humans through the internet via chat interface or by voice, much like how humans interact with each other. There are three main types of chatbots, the first one is the most simple one: ***rule-based chatbots***. These bots have pre-written code about the questions and answers so users can click a button and the bots retrieve the answer. However, it usually cannot respond in detail and specifically to the user, making it the simplest but slowest guide. The second one is ***intellectually independent chatbots***. These bots use machine learning to communicate, which means that they are fed large enough amounts of data to learn the connection between problems and solutions. These bots also learn from their interactions with users and become better at recognizing keywords and faster at retrieving answers. The third one is ***AI-powered chatbots***, these are the combination of the other two, making it more advanced. The technology that is put into these chatbots are machine learning, NLPs and AI, all used to understand and analyze human speech, find the right response and answer in human language (Botscrew 2017). Currently, chatbots are implemented by many companies to work on messaging apps, which is becoming more popular than the social media apps. Companies can even target the right demographics depending on the user, making their chatbots more efficient. Facebook has done a market research regarding the value of bots, and has revealed that 2 billion messages are sent between companies and customers monthly, 56% of those prefer receiving help from companies through messages and 53% are more likely to make a purchase from companies that can message. I believe that there are many potentials for chatbots and NLPs in the future. Soon, chatbots may even gain "personality" and speak in a more relevant way such as making use of teen language for companies that target the younger generation. A famous case of company personality is Wendy's with their sassy attitude on the internet, and this energy can be brought into their chatbots as well. Our human language is unstructured text that is ever growing with new "made-up" words every once in a while. NLPs will continue to improve themselves to be able to connect humans and technology better.

Language is the core of human's social life, helping people communicate and understand each other. Before, language was used to signal danger or new locations. Overtime, human language becomes a lot more complicated, where each civilization speaks different languages, where they can express feelings and sentiments, etc. Right now, technology is learning our language, but is still held back by the limitations of AI and machine learning. In the future, this will change, and NLPs will become smarter, chatbots will become more human, and the connection between technology and humans will be stronger.

For me, NLPs and chatbots are very important as many of my friends and family run small businesses that can get too busy to answer and satisfy every customer looking for support. Chatbots come in as digital helpers, giving quick answers at any time of day, even in the middle of the night when customers are panicking about their products. Of course, right now, chatbots and especially intelligent chatbots are still not too common for business owners, but in the near future hopefully AI chatbots will become the norm and facilitate the interactions both for the company and the customers. As someone who is planning on opening a business in the future, I am eager to see how NLPS and chatbots evolve and improve.

5. Project Ideas

1. Motivation

There is a saying that goes "everyone needs three hobbies: one that lets you be creative, one that keeps you in shape and one that makes you money", photography can be all three of these. Unlike other art forms like painting, photography does not require the same technical skills to create the art. Even though it is not easy to take a good photo, all you need is your camera or phone to be just fine together with your creative mind to visualize the scenes that you want to capture. Photography forces you to observe the world around you in ways you probably never have before. Moreover, photography can keep you in shape by forcing you to explore cities and nature as walking around is better than sitting on the couch and watching TV. Finally, photography can earn you money by doing production business or just running freelance photoshoots in a few weddings or conferences. These examples are sufficient enough to prove that photography is a very popular hobby and it is not easy to take satisfactory shots at once most of the time. As technology improves, your camera lets you take multiple shots back-to-back and review the photo immediately so that you can decide whether to take more or not. Usually we do not review the photos and are able to decide which one should be deleted or not. It is not until you return from home and realize that your camera packs a bunch of photos and you would have difficulties reviewing them manually. Understanding the need of better reviewing photo capability, my team has decided to build an image comparison software with Python that not just supports basic functionalities of

reviewing images but also designing friendly UI so that everyone can use the software without reading too much instructions.

2. Overview

An image comparison software is basically known as a tool that can take out similar images for you to decide which photos are the best of the bunch. However, in this course, our team intends to build a software that not just returns the “duplicates” of the referenced photo but also a list of closely similar photos in your collection to be reviewed. Moreover, taking out the best of the bunch of photos does not mean you delete the others. The software should be designed to support basic functionalities such as: create/delete a gallery, choose between the existing galleries to work with drag and drop images, display images in list view and more of such features should be designed on user interface to make interactions become more friendly to the users.

3. Description

Image Comparison Process:

The software is expected to take in 2 directories: one is from the “selected” image (the referenced image that we want to find other similar ones), the other is from the image gallery (each image from the gallery is called “candidate”). The comparison time depends on the performance of our computers but it will not take too much time (1 or 2 minutes) for the software to return a list of perceptually similar images in the gallery so that we can start the manual review.

The heart of our image comparison software is the Difference Hashing algorithm and the principle of the algorithm can be found here (Adrian Rosebrock 2017). Correspondingly, the two input images (the “selected” and “candidate”) are hashed into a string of values based on the visual appearance using the Difference Hashing algorithm. Images that are identified as duplicated or also perceptually similar when they have “similar” hash values. Hash values that are “similar” means that they are exactly the same or “Similar” hashes means that they are exactly the same or within a predefined range of Hamming Distance (Baris Coskun, et al. 2007).

The Difference Hashing Procedure is Broken Down into 4 Steps:

Step 1: Blurring the Image

The image is firstly blurred with Average Blurring technique from (Baris Coskun, et al. 2006) in purpose of minimising the difference between the corresponding pixels of the “selected” and “candidate” due to the influence of neighbourhood pixels. The neighbouring pixels are taken into the filter to calculate the mean values with the center pixel. After the mean value is calculated, this average value replaces the

current value of the center pixel. Browsing this procedure through all of the pixels which means the image has been blurred.

Step 2: Converting the colored image into Grayscale

Implementing the OpenCV Library that Python supports from (Muhammad Waiz Khan 2021), the input colored image is converted into grayscale using the cv2.imread() method. Convert to grayscale means the 3-channel colors of every pixel are converted to a single gray channel. Thus, the hashing process can be faster and benefits the comparison process as the slight difference in color shades are omitted, the images are compared only based on the similar patterns.

Step 3: Resizing the image

After the image is converted into single grayscale color, no matter what the original resolution of the image was, it must be rescaled into exactly 8*9 pixels. Firstly, resizing the images helps images of different sizes now have the same size so that the patterns are easier to compare. Secondly, our purpose is to generate a 64-bit hashcode for each image for comparison. As 8 * 8 = 64 bits but the reason behind why we have to scale to a 8*9-dimension is if we compare each of the 8 pixels in one column with the adjacent pixels in that row (total of 9 pixels in a row), we are going to have 8 differences (the comparison is executed with 2 different pixels, a pixel cannot compare to itself). Therefore, 8 column * 8 differences = 64-bit hashcode.

Step 4: Hashcode difference computation

In this process, each of the pixels in an image is compared with the adjacent pixels, The output will return True if the concurrent pixel has greater value than its following pixel. On the other hand, the output will return False if the concurrent pixel has a smaller value than its following pixel. When the pixel comparison is completed, a boolean matrix is generated, which must be flattened into a 1-D array and sum of elements is calculated with the following algorithm written in Python:

```
diff_hash = 0
arr = diff.flatten()
for i, val in range(len(arr)):
    if val != 0:
        diff_hash += 2**i
```

According to (Hacker Factor 2013), the image comparison process is executed by comparing the hash values of the images which are just calculated above. The 2 images are considered duplicated when their hash values have 0 Hamming distance. Even the Hamming distance between the two images between 1 to 10 is

recommended as perceptually similar, it is recommended that we personally tune this threshold to decide the similarity with our personal preference.

Layout of the Software's Graphical User Interface (GUI):

Figure 3. Homepage view

The layout of the software is planned to be designed as we can see in the mock view above to provide the user-friendly interface for these following features:

➔ Display list of images in scrolling list view

After we select a folder, all the images in that folder will be displayed in the list view at the bottom of the software and you can scroll through the images using the backward and forward arrows or scroll your mouse.

➔ Create/delete a collection:

To create a new collection, navigate to the lower left side, at the Collection frame, click the "+" icon to create a new collection and rename it by left-clicking on the "New Collection" default name.

Figure 4. Create a new collection

To delete a collection, navigate to the lower left side, at the Collection frame, right click to the collection we want to delete and choose “Delete Collection”.

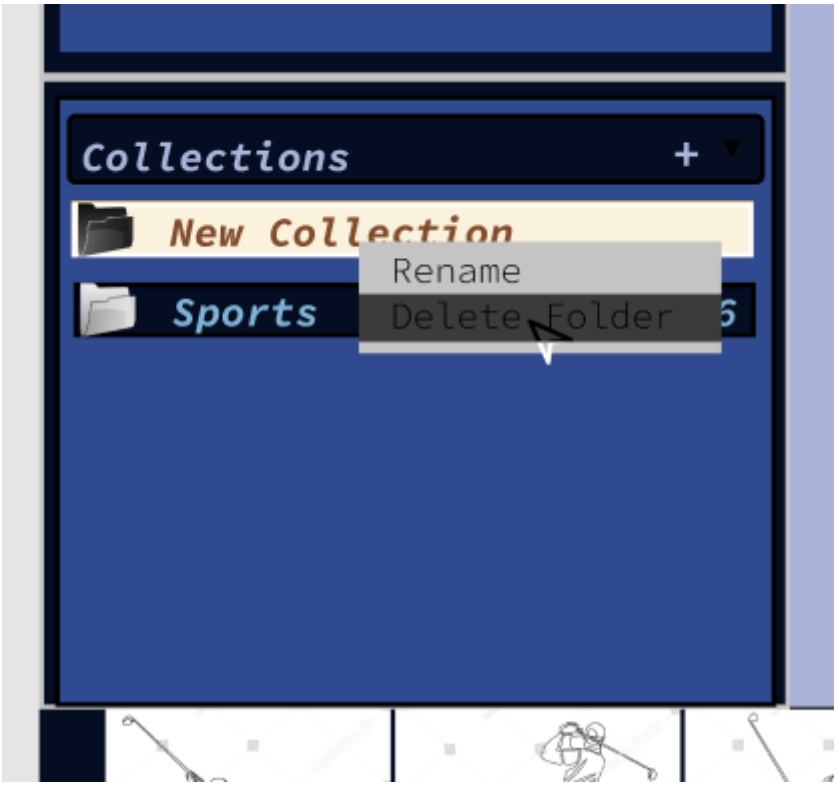

Figure 5. Delete a collection

➔ Choose an existing collection

To choose an existing collection, we also navigate to the Collection frame and use your cursor to left-click on the collection we want to choose.

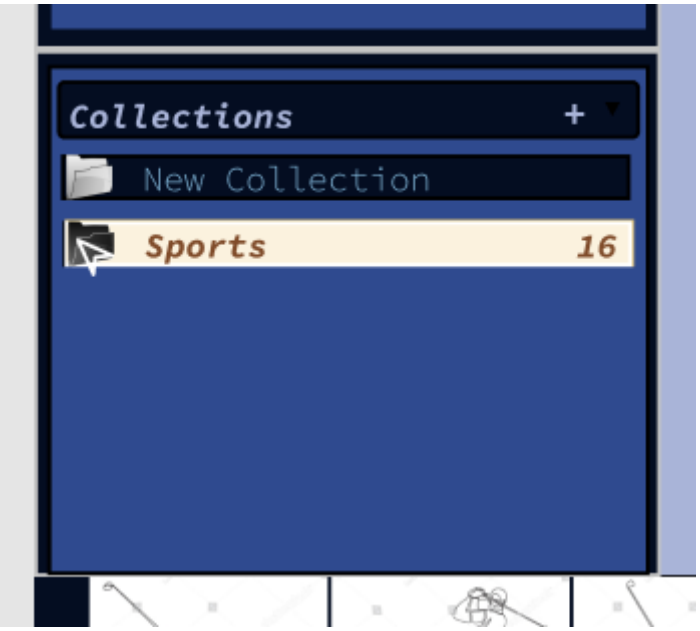

Figure 6. Choose an existing collection

➔ Select multiple images, right-click and choose option (duplicate, delete, rename, view resolution, compare with current “selected” image)

In order to select multiple images, there are 2 options:

+ To select continuous images, we click on the starting image, then hold the “Shift” key on the keyboard and click on the destination image on the image scroll list.

Figure 7. Multiple and continuous images selected in List view

+ To select multiple but not continuous images, we hold the “Ctrl” key on the keyboard and click on the images that you want to select on the image scroll list.

Right-click on the selected images, we can choose between the options: duplicate, delete, rename, view resolution, compare with current “selected” image

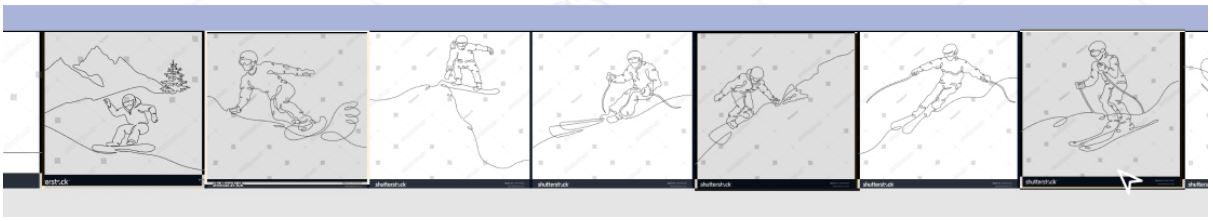

Figure 8. Multiple but not continuous images selected in List view

➔ Drag and drop files and folders to add into the collection

Choose “selected” image: Navigate to the image list at the bottom of the interface, use mouse to choose, drag and drop the image into the Navigator frame locating on the upper left. Now we can see the image displaying in the frame is the current “selected” image

➔ Activate automatic comparison

Navigate to the Navigator frame, right-click on the “selected” image and choose “Compare with images in the currently selected folder”. After the comparison is complete, the similar images will be displayed on the list view at the bottom of the interface.

Figure 9. Activate automatic comparison

➔ Activate manual comparison (display selected image and candidate image side-by-side)

Manual comparison is usually done after automatic image comparison so that users can take out the best images of the bunch based on their preference. By clicking on the  icon located on the right side. When manual comparison is activated, the “selected” image and the “candidate” image will be displayed side-by-side.

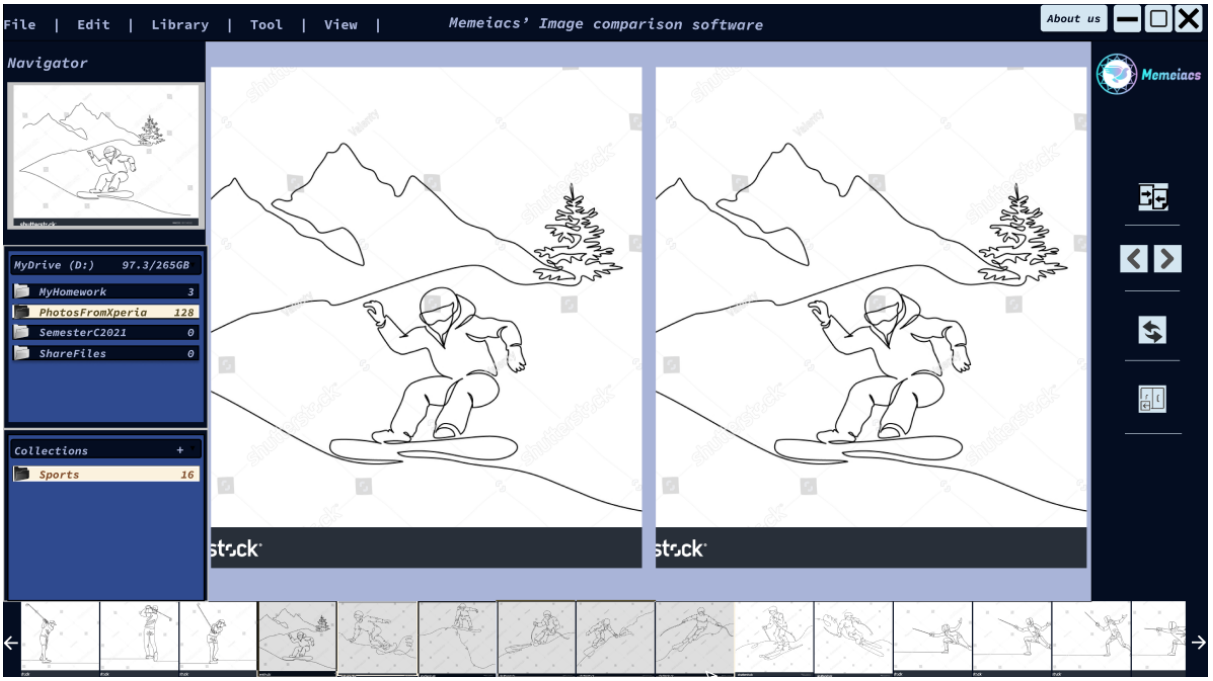

Figure 10. Activate manual comparison

Introduce the functionalities when activating manual comparison

➔ Browsing through the “candidate” images using the arrow icons: 

➔ Switch the view side of the “selected” and “candidate” images by clicking on the  icon. By default, the “selected” image is on the left side.

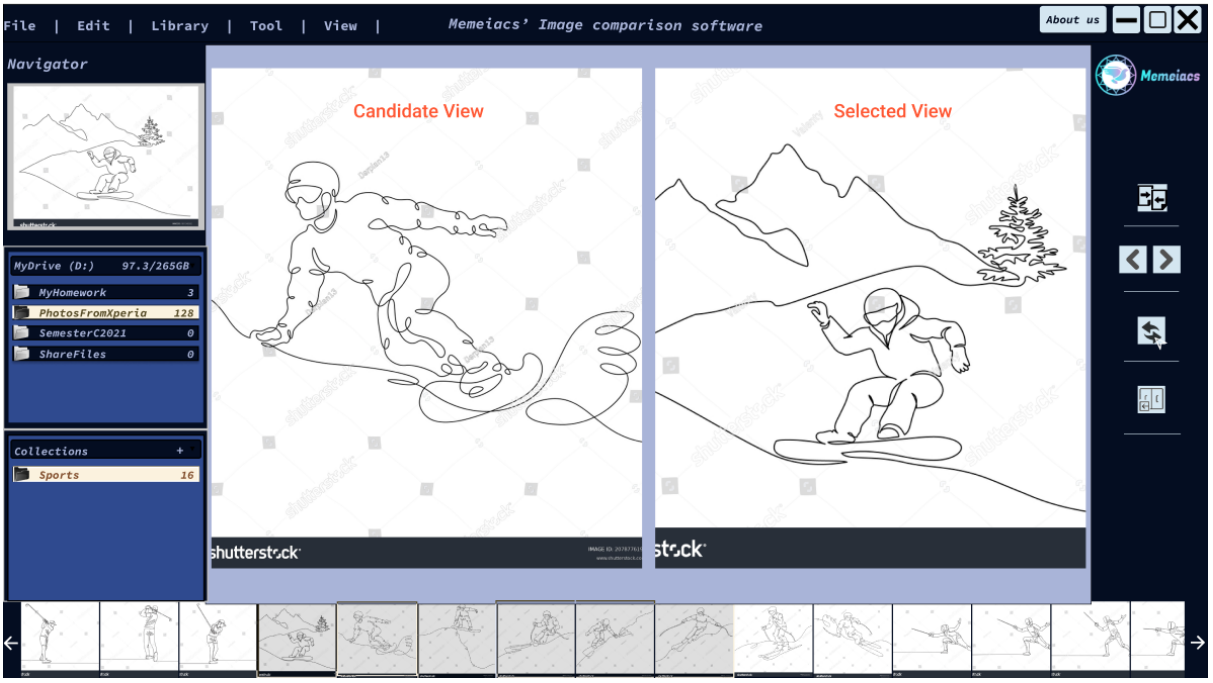

Figure 11. Swap "candidate" and "selected" view side

➔ Replace the "selected" image with the current "candidate" image by clicking on the  icon located on the right side. This function is used when the user finds a "candidate" that is better than the "selected", replaces the "selected" and keeps looking for a better "candidate". Do not worry because this function does not remove any of your original images.

4. Tools and Technology

➔ **Python:** a programming language that is tremendously popular right now because it is considered easy to work with amongst other programming languages. It takes less lines of code in Python to accomplish many common tasks from reading a website to processing text or data format like JSON or crunching numbers. Moreover, Python is powerful enough that robust enterprise-grade software can be built with it such as Pinterest, Dropbox or Quora use Python to build their offering and for automation management. Install Python to your computer is crystal clear as simple instructions are stated in this link.

➔ **An Integrated Development Environment (IDE)** that can work with Python: in this project, our team decided to use Spyder - an open source environment that is built for developing Python applications and providing a good debugging tool. Spyder can be downloaded through the instructions in this link.

- ➔ **OpenCV:** stands for Open-Source Computer Vision - one of the most popular open source libraries for computer vision and image processing tasks for not only Python but C++, Java or MATLAB community. OpenCV can be installed through the instructions in this link.
- ➔ **Difference hash:** one of the hashing algorithms. This hashing algorithm is well known for its implementation in comparing similar images as they generate exactly the same or closely just a few different bits of hashcode when perceptually similar (Chris Currier 2019).
- ➔ **tkinter:** a standard library to build graphical user interfaces for applications in Python. Python developers consider tkinter to be the easiest and fastest way to create the user interface. Note that 'tkinter' and 'Tkinter' are 2 different versions of the library that support different Python versions. 'Tkinter' module only supports Python 2.x and the 'tkinter' module only supports Python 3.x and above. In this project our team decides to use Python 3.8 so the instruction to install tkinter can be found in this link.

5. Skills Required

- ➔ **Basic Python programming:** basic knowledge about installing extensions, importing libraries, defining a method or writing a formula to perform a calculation is needed.
- ➔ **Basic knowledge about data structure:** working with images requires basic knowledge about arrays, matrices such as defining a one or multiple dimensional array (usually 2D), performing traverse, search, insert, delete and perform calculations with the elements.
- ➔ **Read and understand the code**: Doing the project requires a lot of online research. Therefore, this is the most important skill to understand the idea of the referenced codes so that we can modify the work of others to fit with our requirements.

6. Feedback

(Requirement) Give and receive feedback (contribution, attitude, cooperation, strength/ weakness,...). One paragraph needed to combine the whole team's feedback,analyse how cooperation between members benefits the future work.

	to Thuy	to Minh	to Quan	to Khoa
Thuy		Minh is mainly in charge of project ideas and he has done a great job working on it. He has a very enthusiastic	Quan is in charge of the Ideal Jobs section as well as other ones like IT Technologies. He is on time, often early, with his work.	Ethan (Khoa) is mainly in charge of designing and coding the website and has done a fantastic job on the design and

		character and is always ready to work, all leader qualities. He is extremely cooperative.	He is responsible and diligent. He is open to opinions and is extremely cooperative.	functionality. He updates the team frequently and is also extremely cooperative.
Minh	Thuy (Lala) is in charge of the IT Work and IT Technology section. She is also mainly responsible for managing the whole report and coming up with plans for the team. She is well known for her sense of humor, conscientious and rapid response to complete the tasks and support other members. Moreover, she cooperates with the team very well.		Quan is in charge of the Ideal Job and IT Technology section. He is also mainly responsible for keeping track of the teamwork process which his critical thinking is helpful when giving feedback and solving/ pointing out hidden problems that the team has overlooked. He is a hard-working and disciplined person who never overdues the assigned tasks. Moreover, he cooperates with the team very well.	Khoa (Ethan) is in charge of the Tools section and designing the webpage. He is also mainly responsible for guiding everyone through how to work with GitHub. He is a very meticulous person who cares about every detail of his design, so it takes quite a lot of time on his tasks but the team agrees with being best outcome oriented. Moreover, he cooperates with the team very well.
Quan	Thuy is a really reliable teammate. She makes sure our team does not have any mistake in our report and also ensure our team progression	Minh is a very careful person with his work. He is very careful with the details but sometimes takes too much time to do his work. Overall, he is a reliable person.		Khoa is also someone who is very careful with his work. He always focuses on what his job and position needs to do and tries his best to fulfil it.
Khoa	Thuy(LaLa) is a trustworthy person when giving many brilliant ideas to help teammates finish their work perfectly. Beside her work, she was also very active in asking every member to support when she had finished her work.	Minh is actually an active person when asking me “Do you need any help, please ask”. Because he is responsible for the core section of the assignment, he must spend more time to guarantee his work will claim as high a score as possible.	Quan is very serious when working with the team by encouraging everyone to finish their work on time. In addition, he not only contributed ideas to the team project but also supported efficient teammates to have good outcomes.	

References

n.d, *meme noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com*, Oxford Dictionary, viewed 18 November 2021, <https://www.oxfordlearnersdictionaries.com/definition/english/meme>.

Simply Explained 2017, *How does a blockchain work - Simply Explained*, video recording, viewed 27 November 2021, <https://www.youtube.com/watch?v=SSo_EIwHSd4>.

pwc n.d, *Making sense of bitcoin, cryptocurrency and blockchain*, PWC, viewed 27 November 2021,

<https://www.pwc.com/us/en/industries/financial-services/fintech/bitcoin-blockchain-cryptocurrency.html>.

n.d, *How secure is blockchain really?*, MIT Technology Review, viewed 27 November 2021, <https://www.technologyreview.com/2018/04/25/143246/how-secure-is-blockchain-really/>.

2021, *Future of Blockchain: How Will It Revolutionize The World In 2022 & Beyond! - The European Business Review*, The European Business Review, viewed 27 November 2021, <https://www.europeanbusinessreview.com/future-of-blockchain-how-will-it-revolutionize-the-world-in-2022-beyond/>.

George Strawn 2019, 'Blockchain', *IT professional*, January, vol. 21, no. 1, pp. 91-92, viewed 27 November 2021, 10.1109/MITP.2018.2879244 database.

Gayvoronskaya, T & Meinel, C 2020, *Blockchain: Hype or Innovation*, Springer International Publishing AG, Cham.

Botscrew 2017, *What Are Bots? How Do Chatbots Work? - BotsCrew*, BotsCrew, viewed 28 November 2021, <https://botscrew.com/blog/what-are-bots/>.

n.d, *What is Natural Language Processing?*, SAS, viewed 28 November 2021, <https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html>.

IBM Cloud Education n.d, *Natural Language Processing (NLP)*, IBM, viewed 28 November 2021, <https://www.ibm.com/cloud/learn/natural-language-processing>.

2020, *Hot Technologies in Cybersecurity - Cyber Degrees*, Cybersecurity Degrees | Cybersecurity Degrees Online, viewed 26 November 2021, <https://www.cyberdegrees.org/resources/hot-technologies-cyber-security/>.

2020, *Hot Technologies in Cybersecurity - Cyber Degrees*, Cybersecurity Degrees | Cybersecurity Degrees Online, viewed 26 November 2021, <https://www.cyberdegrees.org/resources/hot-technologies-cyber-security/>.

2021, *Artificial Intelligence In 2021 - The Developments So Far*, viewed 27 November 2021, <https://www.analyticsinsight.net/artificial-intelligence-in-2021-the-developments-so-far/?fbclid=IwAR0elxshjSHQL1CfMtQkxU62Ct6nJzY-MZpA12qQJ6qr3x0z0w2hQHlRSi8>.

2021, *The Current State of AI 2021: Report Now Available*, Appen, viewed 27 November 2021, <https://appen.com/blog/state-of-ai-and-machine-learning-2021-report/>.

Adrian Rosebrock 2017, *Image hashing with OpenCV and Python - PyImageSearch*, PyImageSearch, viewed 4 December 2021, <https://www.pyimagesearch.com/2017/11/27/image-hashing-opencv-python/>.

Coskun, B, Sankur, B & Memon, N 2006, 'Spatio–Temporal Transform Based Video Hashing', *IEEE Transactions on Multimedia*, December, vol. 8, no. 6, pp. 1190-1208, viewed 4 December 2021, 10.1109/TMM.2006.884614 database.

Coskun, B, Sankur, B & Memon, N 2006, 'Spatio–Temporal Transform Based Video Hashing', *IEEE Transactions on Multimedia*, December, vol. 8, no. 6, pp. 1190-1208, viewed 4 December 2021, 10.1109/TMM.2006.884614 database.

Muhammad Waiz Khan 2021, *Convert Image to Grayscale in Python*, blog, viewed 4 December 2021, <https://www.delftstack.com/howto/python/convert-image-to-grayscale-python/#convert-an-image-to-grayscale-in-python-using-the-cv2.imread-method-of-the-opencv-library>.

2013, *Kind of Like That - The Hacker Factor Blog*, The Hacker Factor Blog, viewed 4 December 2021, <http://www.hackerfactor.com/blog/index.php?/archives/529-Kind-of-Like-That.html>.

Chris Currier 2019, *Finding Similar Pictures with dHash Values - MSAB*, MSAB, viewed 4 December 2021, <https://www.msab.com/blog/finding-similar-pictures-with-dhash-values/>.

Appendices