

ChatOverflow

Misinformation and Disinformation in ChatGPT

CHEN XI DIONG | 32722656

YI BIN SIA | 33363129

BRANDON LAU | 32184654

LANA HUSSAYNI | 33267871

HONG BO KANG | 32684673

SHUNNOSUKE TAKEI | 32843909

23 MAY, 2023

FIT1055
ASSIGNMENT 2B
INTRODUCTION

Chen Xi

We are a team of 6, and we've been working on the topic of Misinformation and Disinformation in ChatGPT.

(introduce members) and we'll be presenting in this order.

Agenda

- **Introduction**
 - Background
 - Problem Statement
- **Prototype Methodology**
 - Prototype Objectives
 - Prototype Method
 - Prototyping Tools
 - Prototype Development Stages
- **Proposed Solution**
 - Prototype Implementation Workflow
 - Task Delegation
 - Integration of Output
 - Reflection
- **Prototype Demonstration**
 - Prototype Preview
 - Live Demonstration @Figma
- **Conclusion**

FIT1055
ASSIGNMENT 2B
INTRODUCTION

Chen Xi

I will be starting off the presentation with an introduction, followed by Yi Bin and Shun, who will be sharing about our Prototype Methodology. Brandon and Lana will be in charge of the proposed solution section, whereas Hong Bo will demonstrate our prototype. Finally, we will wrap up our presentation with a conclusion and Q&A.

Background

- What is ChatGPT?
 - An AI language model known as Generative Pre-trained Transformer (GPT)
 - Released for public use in November 2022
 - Has over 100 million users (Duarte, 2023)
- How is ChatGPT built?
 - Using ML techniques such as NLP and NLG
 - Training data obtained from web crawling, prominently Common Crawl Corpus
 - Trained with Unsupervised Learning
- What is it used for?
 - Anything! Give it a prompt and it will return you a response

FIT1055
ASSIGNMENT 2B
INTRODUCTION

Chen Xi

ChatGPT is a language model which possesses Artificial Intelligence. It understands human language prompts and provides human-like responses.

Due to its user-friendly interface, easy accessibility and quick responses, it has gained a vast user base in a matter of months upon release.

Since ChatGPT is an AI, it is trained using Machine Learning techniques such as Natural Language Processing and Natural Language Generation. NLP helps it to understand our natural language, and NLG allows it to generate responses in that form.

Its training data is obtained mainly from the web, which is full of inaccurate, unsanitised information, which will lead to our problem statement later on.

The model is trained using unsupervised learning. Put simply, the model is required to make its own analysis and learn patterns from unlabelled data.

ChatGPT can be asked about practically anything, a simple question such as what to eat for lunch, theory related questions such as what is the Ethical Reasoning Framework, as well as write a speech script.

Problem Statement

- What are exactly the ethical issues pertaining to ChatGPT?
 - Misinformation
 - Disinformation
- Why is it a significant issue that requires a shift in perspectives?
 - Given its wide user base and ease of use
 - Can cause large scale damage with relatively small amount of effort
- How can ChatGPT lead to ethical issues from the perspectives of algorithms?
 - people will continue to contribute fake sources of information to ChatGPT's dataset
 - abused by people with malicious intent to make false accusations, impersonate authorities, and other ethical misconducts

FIT1055
ASSIGNMENT 2B
INTRODUCTION

Chen Xi

We will focus on the ethical issues of misinformation and disinformation using ChatGPT.

Misinformation is false or inaccurate information that is shared unintentionally, often due to a lack of knowledge or incorrect assumptions. Since ChatGPT is trained using data from the web, and such data is mostly not checked for its credibility, misinformation can happen quite frequently.

Disinformation is false or misleading information that is deliberately spread with the intention of deceiving people or manipulating public opinion. This is a devastating issue as it can be used to create convincing texts used for phishing, creating public confusion and more.

These issues cannot be taken lightly as it has over 100 million users, and any of them could be misusing ChatGPT to perform malicious acts. Similarly, any of them could be exposed to misinformation due to the unsanitised data set ChatGPT is trained on.

With just a few prompts, ChatGPT can create an extremely convincing block of text. It can also be used to mimic a certain person's writing style or speech tone. This can easily invoke widespread confusion, or defame a certain person.

As there is no guarantee that all users of ChatGPT are purely good-willed, there will always be the risk of ChatGPT being abused for unethical conducts. People can embed ChatGPT within other software or websites to trick their users, or even just use policy-bypassing prompts to exploit

ChatGPT for malicious uses.

In misinformation context, people's online posts are not necessarily true and ChatGPT doesn't check for its validity. As long as people contribute to the web, there is always a potential risk of ChatGPT providing false information.

Prototype Objectives

- **Objective of our prototype**
 - General demonstration of how the final product would look like
 - Easier to make changes than final product
- **Our aims**
 - Easy to use
 - Clean and comfortable design
 - Useful functions
 - Privacy-focused
- **Prototype method chosen**
 - High fidelity prototype

FIT1055
ASSIGNMENT 2B
PROTOTYPE METHODOLOGY

Yi Bin

The objective of us making a prototype is to demonstrate to the audience of how our product would look like, and make possible adjustments or improvements as efficiently as possible. We understand that error checking and debugging after the final product is released will have a heavier workload.

Our prototype aims to create a user-friendly, visually appealing, and functional product. We aim to achieve the following key goals:

Easy to use: We want to ensure that our prototype is intuitive and easy for users to navigate and interact with.

Clean and comfortable design: The prototype will have a visually appealing and modern design, providing users with a comfortable and engaging experience.

Useful functions: Our prototype will showcase the core features and functions of the final product, allowing users to understand its value and usefulness.

Privacy-focused: We will prioritize the privacy and security of user data throughout the prototype development process.

Prototype Method Justification

Pros of High Fidelity

- Realistic representation:
 - High fidelity prototype provides a visually appealing and interactive representation of the final product, giving stakeholders a clear understanding of its look and feel.
- User experience testing:
 - The high fidelity prototype allows us to simulate user interactions and conduct usability testing, ensuring a seamless and intuitive user experience.
- Feedback and improvements:
 - The high fidelity prototype enables us to gather valuable user feedback early in the development process, facilitating iterative improvements and reducing the risk of costly revisions later on.

FIT1055
ASSIGNMENT 2B
PROTOTYPE METHODOLOGY

Yi Bin

Our team has opted for a high fidelity prototype for our project. This is because some of our members have prior experience in making prototypes, therefore we are able to progress smoothly during this process.

Realistic representation: It is visually and interactively represent the final product, helping stakeholders understand its look and feel.

User experience testing: We can simulate user interactions and conduct usability testing, gaining valuable insights into the user experience.

Feedback and improvements: Gathering user feedback early enables iterative improvements, reducing the risk of costly revisions later on.

Prototyping Tools

- Figma

- it is a cloud-based design and prototyping tool used for creating user interfaces, wireframes, and interactive design prototypes.
- Users can create and edit designs directly in the browser without the need for additional software installation



- Zoom

- Zoom is a widely used video conferencing and communication application that enables users to connect and collaborate remotely. It offers a range of features that facilitate online meetings, webinars, and virtual events.
- Allows us to communicate in real time



FIT1055
ASSIGNMENT 2B
PROTOTYPE METHODOLOGY

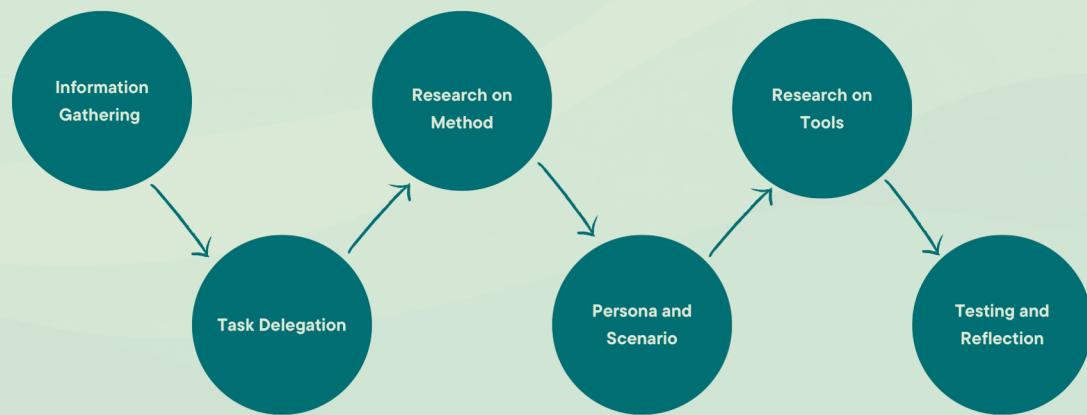
Shun

When making our prototype, we utilized two programs: Figma and Zoom.

Figma is a cloud-based design and prototyping tool used for creating user interfaces, wireframes, and interactive design prototypes. It allows users to create and edit designs directly in the browser without the need for additional software installation.

While Zoom was used for remote communication so we could discuss in real time while working on the prototype.

Prototype Development Stages



FIT1055
ASSIGNMENT 2B
PROTOTYPE METHODOLOGY

Shun

These are the stages of how we developed our prototype.

First, we gathered the necessary information for our project scope from various sources, such as searching the web for user reviews, conducting market research, etc.

We then defined tasks and divided them among our team to work in parallel for efficient collaboration.

After that, we researched on prototyping methods, namely low and high fidelity. We decided to produce a high fidelity prototype, as justified earlier.

Finally, we searched for prototyping tools and decided to use Figma for creating our prototype.

We also tested the prototype and reflected on what could be improved. Subsequently, we made changes and repeated this process.

Proposed Solution

- **Goal:** Create a user-friendly website that enables interactive communication and engagement with users and developers of ChatGPT
- **Key Features:**
 - Direct interaction with developers
 - Ability to create threads.
 - Credibility system
 - Dedicated page for patch notes
 - Actively moderated

FIT1055
ASSIGNMENT 2B
PROPOSED SOLUTION

Brandon

The proposed solution is to create a user-friendly website that enables interactive communication and engagement between users and developers of ChatGPT.

Users will have a direct channel to interact with developers, share feedback, and express suggestions or concerns.

The website will feature thread creation for community engagement and discussions on ChatGPT-related topics.

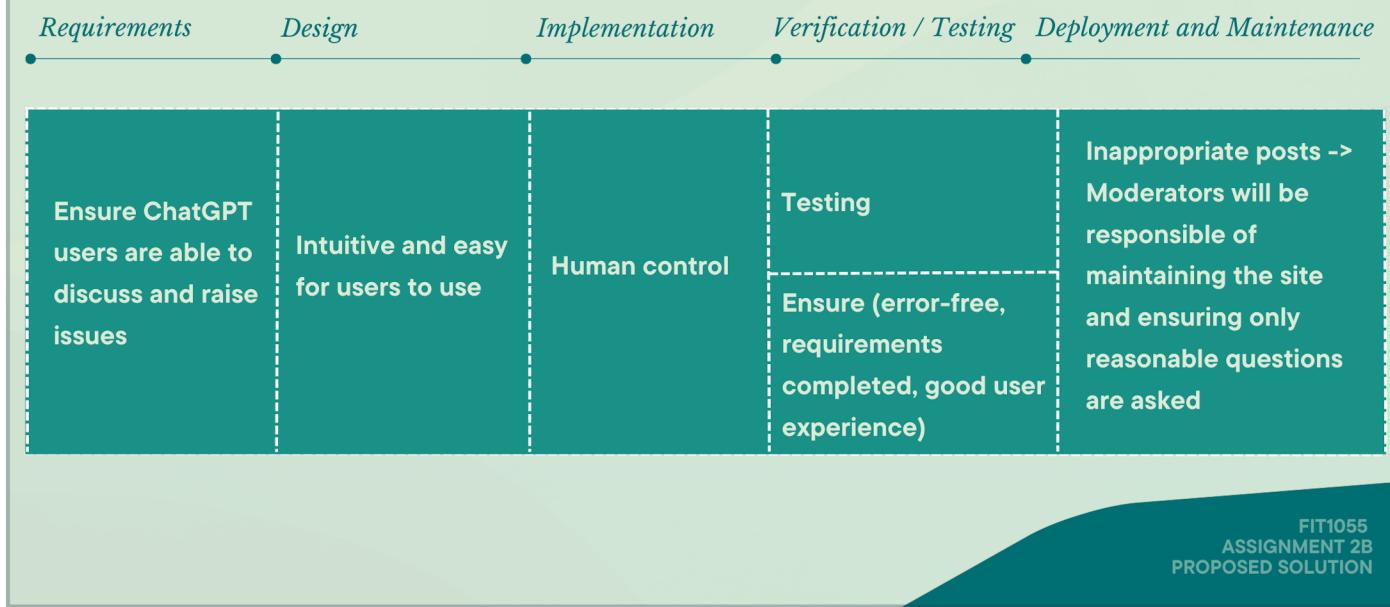
A credibility system will be implemented to establish trust and reward valuable contributions from users.

A dedicated patch notes page will provide transparency, keeping users informed about updates and improvements in ChatGPT.

The website will be actively moderated to maintain a safe and respectful environment.

User-friendly design, intuitive navigation, and a visually appealing layout will enhance the user experience.

Prototype Implementation Workflow



Brandon

Requirements: Enable ChatGPT users to discuss and raise issues to promote user engagement and create a feedback loop

Design: The prototype needs to be intuitive and easy for the users to interact with

Implementation: We decided to introduce human control into the loop of ChatGPT usage to better control the responses it gives, better filtering out inaccurate information, as well as flag malicious uses of the tool.

Testing: We tested the prototype to ensure that it does not have errors during the demonstration process to better showcase the overall user experience. We also checked whether the prototype covers all of our intended features.

Deployment and Maintenance: We also wanted to demonstrate that it is possible to adhere to problems in the website(e.g. inappropriate user posts and comments) as soon as possible.

Prototype Implementation Workflow



FIT1055
ASSIGNMENT 2B
PROPOSED SOLUTION

Brandon

Whenever a user finds a potential bug or issue with ChatGPT, they can follow the steps here.

First, the user needs to log in to his account at the login page, located at an intuitive corner of the website.

After that, the user is able to post a thread or problem, then it would be reflected on the main page. This aims to seek help from other people, especially the Developers of ChatGPT.

Then, when a Developer sees the problem, they can address it or inform the user that it is not a major problem.

Last but not least, the Developer can make appropriate updates to ChatGPT and release patch notes on the platform.

Task Delegation

Information Gathering Everyone	Research on Method Yi Bin	Research on Tools Brandon and Shun	Testing and Reflecting Lana
<ul style="list-style-type: none">Conduct Market ResearchGather User RequirementsCombine InformationDefine Project Scope	<p>Task Delegation Leader</p> <ul style="list-style-type: none">Divide tasks to each memberOpinions and preferences are taken into account	<p>Persona and Scenario Chen Xi and Hong Bo</p> <ul style="list-style-type: none">Two types: low fidelity and high fidelityLow fidelity: Easy to create, but not elaborate enoughHigh fidelity: Relatively time and resource consuming, but provides better visualisation	<ul style="list-style-type: none">Adobe XD, Sketch, Figma, etc.Figma: clean UI, useful tools and external libraries

FIT1055
ASSIGNMENT 2B
PROPOSED SOLUTION

Lana

Everyone conducted market research and gathered user requirements to the problem statement. We then combined all the information to define the project's scope.

Chen Xi, our group leader, then divided the tasks to each of the team members. In this process group members' opinions and preferences about the tasks are taken into account to ensure fairness.

There are two types of prototypes, namely low fidelity prototype and high fidelity prototype. Low fidelity prototypes are easy to make but may not be extensive enough to demonstrate our solution. Since a member has prior experience in prototyping, therefore Yi Bin proposed that we opt for high fidelity.

Hong Bo and Chen Xi created three personas, namely an Average User, a Developer, and a Moderator. We made up a scenario where the user makes a post about a ChatGPT issue, the Dev responds and makes adequate updates, and the Mod filters out inappropriate content.

After conducting thorough research, Shun and Brandon have concluded that Figma is the ideal prototyping tool for our project as it had the capabilities to precisely visualize the essential elements of a forum page in a remarkably clear and intuitive manner. Also due to the fact that a team member has experience working with the application, we decided to use Figma for creating the prototype.

Having built a working prototype, Lana conducted tests and provide comments as well as possible improvements that could be made. This is a process where we iteratively refine our prototype until it reaches its finalised form.

Integration of output into prototype

- **Research on Methods**

- High fidelity prototype is used to provide a better demonstration, which is better for the audience to grasp our concept.

- **Research on Tools**

- Use Figma's design features to create the visual elements of the prototype, including screens, components, and interactions.

- **Prototype Implementation Workflow**

- Provides the fundamental functionality of the website
 - Conceptualization of idea into product

- **UI Design**

- Clean and simple interface allows for easy visualization and direct interaction to convey how our prototype should work

- **Persona & Scenario**

- For the audience to better understand who our target users are, and in what case would they use our product

FIT1055
ASSIGNMENT 2B
PROPOSED SOLUTION

Lana

Now, lets look into the integration of output into our prototype.

First off, our research on methods. When looking for what kind of prototype we should go for, we found that a high fidelity prototype to provide a better demonstration, which is better for the audience to grasp our concept.

For research on tools we decided on using Figma's design features to create visual elements of the protype. This includes screens, components, and interactions.

For our prototype implementation workflow we made sure to provide the fundamental functionality of the website and conceptualized our idea into a product. In our case, a website.

We are aiming for our UI design to be clean and simple for easy visualization and direct interaction to convey how our prototype should work.

Lastly for our persona and scenario, we have prepared a scenario where a user has encountered a problem on ChatGPT and posted the issue onto our site. In our scenario we have three different perspectives of different user roles: normal user, developer and moderator.

Reflection

- What could we have done better?
 - Prototyping and Design:
 - Each page could be less congested
 - Tag developer feature
 - Profiles for Developers
 - Edit post feature for poster
 - Teamwork
 - Punctuality

FIT1055
ASSIGNMENT 2B
PROPOSED SOLUTION

Lana

To reflect on our process, there are few things we could've done better.

In terms of prototyping and design, our pages could be less congested as it might be difficult for users to navigate with so many things on screen all at once.

We could also add other features to our website such as a "Tag Developer" which would give users the choice to directly notify developers that would be most capable of resolving their issues.

To make that process easier, we could have also added a feature profiles for the Developers so that users are able to identify which developer should they reach out so that their issue is solved.

Lastly an edit post feature for the poster to update their question if more issues were to arise.

In terms of teamwork, in general we worked well there was just the minor issue with punctuality. However these never escalated into anything major as they were very small delays.

Prototype Preview

The screenshot shows the main interface of the Chat Overflow prototype. At the top right, there are three user profiles with arrows pointing from them to their respective detailed views below:

- John Jonn...** (Normal User)
- Sam Altman** (Dev)
- Chen Xi Diong** (Mod)

The main content area includes:

- Recent Patch Notes:** Two sections: "New releases (May 12)" and "New releases (May 3)".
- Popular Questions:** A section with a "Make a thread" button and filters for "Recent", "Popular", and "Bounded".
- Recent Tags:** A list of 11 tags: Tag 1, Tag 2, Tag 3, Tag 4, Tag 5, Tag 6, Tag 7, Tag 8, Tag 9, Tag 10, Tag 11.
- Example question here:** Three examples with placeholder text and a "tag" button.
- Advertisement:** An advertisement for Shopee Puffo with the text "GET UP TO RM100* FOR FREE".

At the bottom right of the main screen, there is a footer with the text: FIT1055 ASSIGNMENT 2B PROTOTYPE DEMONSTRATION.

Hong Bo

We will be demonstrating our prototype using the three personas in the scenario mentioned earlier.

Here is a preview of what the main page would look like, and we can log in with different users, which are the Normal User, the Developer and the Moderator through the upper right corner.

Prototype Demonstration

<https://www.figma.com/proto/03e2QwK8sIpesLQXZgbjzH/ChatOverflow?type=design&node-id=162-1923&scaling=min-zoom&page-id=0%3A1&starting-point-node-id=162%3A870>

FIT1055
ASSIGNMENT 2B
PROTOTYPE DEMONSTRATION

Hong Bo

For demonstrating the prototype I will be using Figma as the platform. So first of all we will assume there is 3 scenario for using this prototype. In the first scenario, a user found out that there is some content that is quite interesting or maybe triggered some ethical problems the user can come to this website and upload the problem they met. So I will be going to demonstrate how a user can post a thread on this website. Before doing this I will have a brief introduction on the feature on our website. First, you can see that on the right side of this website, there is a patch note where the user can see the updates on ChatGPT or OpenAI. And for the body part, there will be some popular threads that are posted by other users. And we can see that we can list the thread in popular bountied and recent. So for the scenario, the users need to log in as him and make a thread by clicking the button "Make A Thread" on the main page and we will be on the new. So what if a user doesn't know how to post a thread we also add the FAQ section which is in the search bar. Which might help the user to solve some common questions that are asked by other users. For the next scenario where is for the developer of the ChatGPT/OpenAI.

Next up is for the developer to answer this new thread first we can log in as a developer and now we can see that there is a developer tag below the profile, and on the main page we also can see that the patch note part is different from the user as the developer of ChatGPT can update and edit the patch note. Moving on to answering the thread as a developer we can click on the thread and click the comment box to add a comment and post it we can see that the new comment will be priorly posted on the first comment in the thread as it is a developer.

And for the last persona is where we as the website developer will be the moderator of this website which have the capability to unseen the message from user, delete thread, and edit the thread. So I will first log in the as a moderator and u can see the main page have the icon for unseen, delete, and edit the on the end of the box of the thread. So by clicking the eye Icon on the third thread we can see the post will jump out the red text which make says "Hidden from users" then by clicking the delete icon the thread will be deleted. Next is where we want to edit the first thread. We click the pencil icon on the first thread and we will jump into the thread page and click the title of the thread to highlight it and (click backspace) to delete the original title then (press spacebar to have the new text written in Figma) write or edit the new text.

And that all for my part and I will pass it to Shun for the Conclusion.

Conclusion

- **The problem:**
 - Spread of misinformation/disinformation by ChatGPT
 - Significant impact due to large user base
 - Abusive usage with malicious intent
- **Our solution:**
 - A dedicated forum page for anything ChatGPT-related
 - a collaborative space for developers and users
 - establish a closer connection with users/increase transparency
 - Features:
 - Threads and discussions, interaction with devs, patch notes, etc.
- **Building our prototype:**
 - High fidelity prototype
 - Applications used: Figma, Zoom
- **Future Works:**
 - Establish a fully working, continuously improving website

FIT1055
ASSIGNMENT 2B
CONCLUSION

Shun

To summarize, our ethical problem was that ChatGPT was spreading a lot of misinformation and disinformation, which can be severely impactful because there are millions of people around the world using it as a “reliable” and “credible” source of information. Because of how many people are unaware of the situation, other people can abuse the capabilities of ChatGPT for malicious intent, whether that's for financial gain, or to impersonate someone, or creating fake rumours and accusations, all of which are already a big problem in the internet today. So, the question was how can we solve this? We decided that it was best if we built a platform where users and developers can work together to uncover and report any issues and work towards a fix as a community. This will allow the developers to be more transparent with their users, and establish a closer connection with them. This solution also gives an opportunity to help raise awareness on the potential ‘dangers’ of ChatGPT and other AI-powered technology as well. As explained in more detail earlier, some features of our forum page are the ability to create threads and discussions, more interaction with the developers, a dedicated page for recent patch notes, and others. Then finally, to accurately represent and visualize what our forum page would look like, we created a high fidelity prototype, using applications such as figma and zoom.

Future works: Our team has high hopes for this prototype's future development. We envision expanding the site's features and capabilities to better serve their users and meet their needs. We will continuously be improving the user experience and incorporating feedback to make the website even more user-friendly and intuitive. With these efforts, we hope to establish this website as a leading platform for discussing and resolving ChatGPT's ethical issues pertaining to

misinformation and disinformation.

So that concludes our presentation, and now we will be moving on to our qna.

Q&A

FIT1055
ASSIGNMENT 2B
Q&A

References

Duarte, F., 2023 "Number of ChatGPT Users (2023)" Exploding Topics
<https://explodingtopics.com/blog/chatgpt-users>