

Team Name

Nasa GenAI Vanguard

Nasa, the first letters of each team member's name, Nivi, Amanda, Sofia, Alec
GenAI Vanguard, team that protects against harmful GenAI biases and behavior

Team Goals

Briefly describe your team's current understanding of the project and desired outcome.

UNDERSTANDING OF PROJECT PROMPT

This project aims to research the concept of "**Everyday Algorithm Auditing**" within **Generative AI systems, such as ChatGPT and Stable Diffusion**, with the overarching goal of empowering ordinary users to identify and address harmful algorithmic behaviors. Unlike traditional algorithm auditing methods led by technical experts, which often miss emergent biases, this project seeks to tap into the collective power of regular users who encounter AI systems in their daily interactions.

UNDERSTANDING OF THE DESIRED OUTCOMES

- Researching strategies to effectively involve and retain everyday users to help report emergent biases and harmful behaviours of popular online AI systems through creative UI/UX solutions
- Ideating strategies to involve users of different domain and technical knowledge to provide feedback on harmful behavior when it arises through innovative UI/UX methods
- Developing an automated, theoretical framework to assimilate the feedback collected from user audits on these tools in a easily shareable format for developers to make fixes

How do we define "success" in our work for the client/project?

- Successfully identifying and mitigating at least three harmful biases in Generative AI systems like ChatGPT, DALL-E, Microsoft Co-pilot, etc
- Suggest a simple feedback-loop framework to connect user audit feedback and developer fixes
- Contribute to the fairness in AI research community efforts

How do we define a successful learning experience?

- Gaining practical insights into the mechanisms of bias in AI and the methodologies for auditing algorithms through user-centered qualitative and quantitative research
- Learning mixed research methods and applying them to develop solutions that enable everyday user algorithm auditing in popular GenAI tools
- Active engagement from all the team members in achieving the course learning objectives, active participation on all assigned activities, basic expertise of practical UI/UX/Design skills and becoming an effective team player
- Sufficient skills to conduct basic HCI research and confidence to work on similar projects with labs or in the industry

We will be successful if we gain _____ from this project.

- UI/UX/Design Research Skills
- Tangible and actionable suggestions that make a difference to mitigating harmful AI bias in GenAI tools
- Learn effective team collaboration and project management skills
- Creating an enjoyable working environment, learning from different individuals and empathizing with diverse opinions
- Effective time management and delivering outcomes under strict schedules and deadlines

Roles

Describe the general roles needed for this project. Roles relate to administrative work needed to manage the project and the work to complete the assignments. Common roles often include project manager, meeting note taker, report designer, final deliverable compiler, QA/proofreader, etc. Discuss and create a list of roles appropriate for your team.

We strongly recommend that these roles should change from module to module, as you each should rotate in order to gain experience in each role as well as distribute the workload evenly.

[Copy this table for each upcoming module and complete it. The team can change roles and role assignments as the semester progresses - this document is a work in progress.] jianweic@andrew.cmu.edu

Module #2		
Data Analyst Quantitative Researcher [Fallback Project Manager]	Nivedhitha Dhanasekaran	<ul style="list-style-type: none">• Analyzes data collected from user testing/input• Programming, scriptings, digital toolkit curator• Basic background research and literature survey• Supports scheduling and other project management tasks
Design and UI/UX Specialist [Fallback Qualitative Researcher]	Jianwei Cheng	<ul style="list-style-type: none">• Manages overall project design and strategy• Prototyping and wireframing• Primary UI/UX and design consultant for the team
Communications Officer Qualitative Researcher [Fallback Data Analyst]	Sofia Reyes Franco	<ul style="list-style-type: none">• Manages communication and group chat logistics

		<ul style="list-style-type: none"> • Carries out the background research to drive the project • Ensures the team is aligned towards the objectives for each milestone by facilitating effective communication
Project Manager [Fallback Communications Officer]	Alec Chen	<ul style="list-style-type: none"> • Coordinates overall project, sets and enforces meetings • Sets project tasks, assigns individual project tasks • Tracks progress and co-ordinates communication

Communication

How will we communicate with each other/ keep each other aware of progress (especially to avoid surprises) and provide feedback to each other?

- **Meetings**
 - Weekly Digest on Friday or whenever availability matches on [When2Meet](#)
- **General Communication**
 - Slack channel **team-d1**
- **Collaborative Brainstorming**
 - Whiteboarding in-person
 - **Figjam**
 - iPad Notes, during sections
 - **Zoom** calls
- **Progress Tracking** (depends on ease of use for the team)
 - **Notion** (or)
 - **Trello**
- **Documentation**
 - [Google Drive](#)
 - **Slack**

How will we make decisions / reach alignment?

- Schedule a **all-hands meeting** with the Project Manager to debate and propose resolutions to conflicts as group over a call or in-person
- Deliberation or voting on difficult decisions, just a general consensus on everything else
- Make a priority matrix of tasks and try to maximize gains for everyone
- Shuffle individual effort requirement if someone is unable to contribute for a week by increasing workload for the next, or playing to each individual's strengths
- Getting things done is more important than achieving perfection, do iterative, rapid developments

What tools will we use to communicate?

- **Method:** We will primarily use Slack for day-to-day communication and Trello/Notion for tracking tasks and progress
- **Decision Making:** Decisions will be made democratically, with a vote if necessary. For deadlock situations, the Project Manager will set up a meeting to arrive at a consensus
- **Tools:** Slack, Trello, Google Drive for document sharing and collaboration
- **Availability:** Team members are expected to communicate their availability weekly and attend all scheduled meetings unless prior notice is given upto 20-30 mins before the scheduled activity

What are the team's expectations for availability? Are there levels to each person's availability?

- Moving forward, each team member should do their best to keep the scheduled meeting time (Friday 5PM) free, as to minimize time conflicts
- Fill up availability on When2Meet weekly, as and when their schedule changes
- Always suggest changes to meeting times by editing the Google Calendar event set up by the Project Manager
- Notify schedule changes on the Slack channel

Deliverables

How will we define our team's quality standards?

- All members should ensure **quality in each deliverable** by ensuring that they are data-driven, research-backed, user-tested, and actionable
- All members should take **ownership** by contribute to assignments, review changes and provide feedback for all deliverables
- All members should **contribute and collaborate** by attend scheduled brainstorming sessions to encourage creative solutions
- All members should be **accountable** and regularly check-in for updates in team meetings and sections on Monday

How will we support people in achieving those standards?

- We will make sure to maintain consistent communication/checking up on teammates to ensure things are getting done
- This not only keeps teammates accountable, but allows us to help out if something requires more effort than expected

How will we help people to improve their weaknesses?

- Identify their strengths and fill in for weaknesses
- Provide resources and support
- Emotionally back them up
- Prompt open communication
- Celebrate progress, successes, also failure
- Set clear guidance and expectation

How will we maintain accountability when standards aren't met?

- Provide constructive feedback
- Periodic internal peer performance evaluation
- Modify our goals and deadlines as needed
- Have backup plans and contingencies
- Keep our TAs and instructors in the loop for external reviews and feedback

Logistics

How will we work to organize ourselves as a team to accomplish our goals?

- **Organization:** Weekly meetings to review progress, adjust plans, and distribute tasks
- **Commitments:** Respect for each other's time and commitments by adhering to agreed deadlines
- **Tools:** Notion/Trello for task management, Google Drive for storage and sharing, Zoom for meetings
- **Work Sharing:** All documents and code will be stored on Google Drive/GitHub with access given to all team members

How will we work around other commitments?

- We will do our best to have a regular meeting time that all group members should keep free in their weekly schedules.
- For individual work, we will do our best to manage our time wisely, and let our other team members know if something requires more time and/or work than we currently have to complete.

What tools will we use to plan and collaborate?

- General Communication: Slack channel, Zoom
- Collaborative Brainstorming: Figjam, Whiteboards, in-person meetups, sections
- Progress Tracking: Notion/Trello
- Prototyping: Figma, Slides, etc.
- Documentation: Google Drive, GitHub, Google Docs, Google Slides

How will we store and share work created by individuals?

- Documentation: Google Drive, GitHub, Slack
- Adhere to our Logistical framework for managing work

Conflict

How will we address someone not contributing fully?

- Private conversations or coffee chat (not in an accusatory manner for sure)
- Express our concerns as a group amicably
- Understand their needs and difficulties before jumping to assumptions
- Offer support and help connect with course staff to help negotiate deadlines
- Follow up
- Involve course staff as a last resort

How will we handle team conflict? How do we make a decision when we disagree?

- Acknowledge the conflict
- Encourage open dialogues for our team members to express their perspectives and concerns
- Ensure everyone feels heard and understood by actively listening to their points of view
- Identify common ground
- Brainstorm and discuss alternative solutions or approaches to address the disagreements
- Aim for consensus rather than unanimity
- Involve course staff as a last resort

How will we handle a team member that needs support?

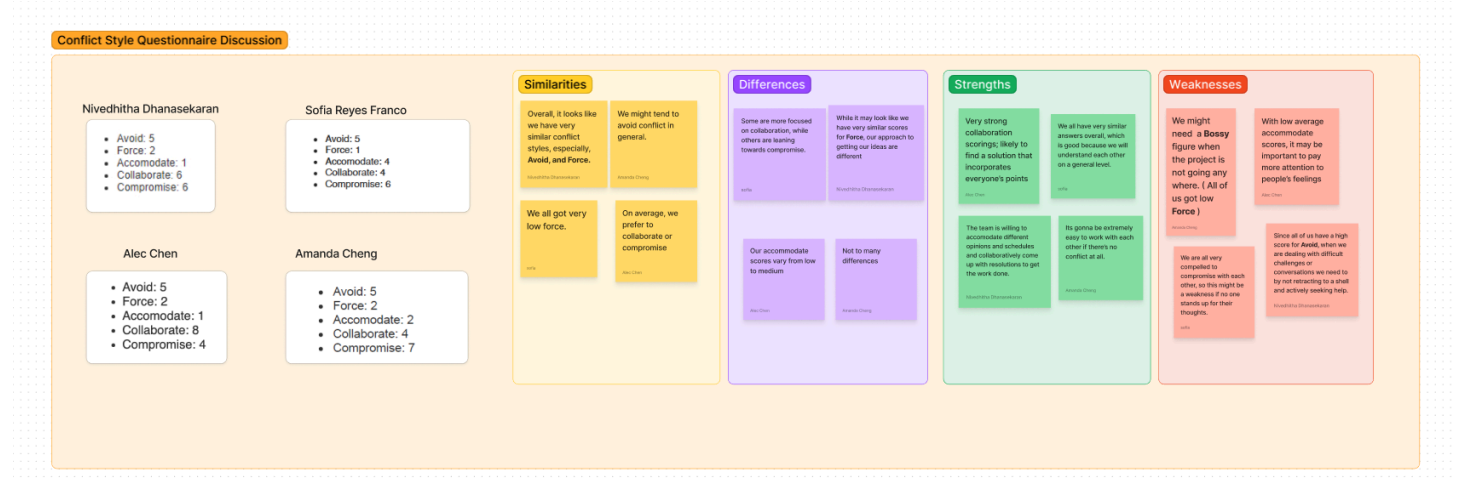
- Encourage the member to communicate openly about their challenges
- Listen actively to understand their needs and any obstacles they are facing due to coursework or if they are comfortable, personally
- Provide any form of resources they might need, connecting them to course staff or other professionals on-campus
- Regularly check-in to make sure everyone is on the same page and aligned towards meeting deadlines
- Be flexible and willing to adjust and accommodate changes

What other topics do we need to agree upon?

- Ownership and division of contributions if any of us want to put the project in our portfolio or if we choose publish the work online

Pre-Mortem Documentation — FigJam

- please click on the links to navigate to the tool pages



Conflict Style Discussions based on the Questionnaire Report

Pre-Mortem Ideas - What Might Go Wrong?

Internal Risks

Inadequate planning: Fail to establish a clear goal, roles or deadlines.

Amanda Cheng

Inadequate skills or expertise among team members for completing assigned course activities and project deliverables. This may become a difficult issue since none of are HCI majors.

Niveditha Dhanasekaran

Unresolved conflicts and disagreements: Conflicts and disagreement that are not resolved can create a tense or hostile working environment.

Amanda Cheng

All of us needing to compromise may lead to not getting things done as quickly as possible or as well as possible.

Yufei

If possible, always make sure to test everything: dedicating a lot of effort on something only to realize it doesn't work in practice is an indicator of insufficient testing

Alec Chen

Unequal contribution: not necessarily about members are not contributing equally but more of don't feel like the responsibility to do so.

Amanda Cheng

All of us avoiding the issue and hoping someone else will take over to address the problem.

Niveditha Dhanasekaran

Poor communication may lead to misunderstandings and failures in connecting divided work

Alec Chen

Being realistic about project development is important, otherwise we may miss deadlines

Alec Chen

No one person should be constantly leading discussions/meetings. Consequently, people that aren't contributing as much should be asked if they have things to share

Alec Chen

Worst-case scenario

Pursuing perfectionism and falling move on to next tasks.

Niveditha Dhanasekaran

Failing to seek help from TAs and course staff.

Niveditha Dhanasekaran

Procrastination ID and last minute rush.

Amanda Cheng

Failure to meet the deadline and not being able to at least invoke the revision policy (not recommended!!!!)

Niveditha Dhanasekaran

Repeatedly missing deadlines and failing to seek help.

Niveditha Dhanasekaran

We may also run into issues with software like QLI or Canvas timeouts, laptop breaking, loss of data, failing to back-up files.

Niveditha Dhanasekaran

Other factors like unexpected events or difficulties like a global pandemic or local incident.

Amanda Cheng

Dependence on external factors or third-party dependencies when conducting user studies and interviews for data collection.

Niveditha Dhanasekaran

Technical issues or third-party dependencies with stakeholders.

Amanda Cheng

Solutions

Schedule a all-hands meeting with the Project Manager to debate and propose resolutions to conflicts as group over a call or in-person.

Getting things done is more important than achieving perfection, do iterative, rapid developments.

Schedule time to get things done apart from sections and also consider mental health and workload of each member.

Work on documentation, plan tasks and sticking to deadlines.

Learning to seek help and keeping TA or course staff in the loop to ensure learning and skill development.

Make a priority matrix of tasks and try to maximize gains for everyone.

We will make sure to maintain consistent communication/checking up on teammates to ensure things are getting done.

Not all meetings need to be productive - some are just to celebrate accomplishments or review feedback for improvements.

Every member is responsible for ensuring quality of deliverables, taking ownership, contributing, and being accountable.

Planning a learning roadmap to get members up to speed if the assignment requires skills outside the teams' collective wisdom.

How do you feel about your team now?

- Helpful team members
- Effective communicators
- Open to discussion
- Flexible
- Encouraging and empathetic

What effect did doing the Project Pre-Mortem have on yourself? On others?

- Helpful in establishing the roles and expectations of each team member
- Understand group dynamics and work ethics
- Determine logistics and set up workflows

Pre-Mortem Ideas - What Might Go Wrong?

Inadequate planning: Fail to establish a clear goal, roles or deadlines.

Amanda Cheng

If possible, always make sure to test everything: dedicating a lot of effort on something only to realize it doesn't work in practice is an indicator of insufficient testing

Alec Chen

Poor communication may lead to misunderstandings and failures in connecting divided work

Alec Chen

Unresolved conflicts and disagreements: Conflicts and disagreement that are not resolved can create a tense or hostile working environment.

Amanda Cheng

No one person should be constantly leading discussions/meetings. Consequently, people that aren't contributing as much should be asked if they have things to share

Alec Chen

Unequal contribution: not necessarily about members are not contributing equally but more of don't feel like the responsibility to do so.

Amanda Cheng

Procrastination :D

Amanda Cheng

Dependence on external factors or third-party dependencies when conducting user studies and interviews for data collection.

Nivedhitha Dhanasekaran

All of us needing to compromise may lead to not getting things done as quickly as possible or as well as possible.

sofia

Being realistic about project development is important, otherwise we may miss deadlines

Alec Chen

Inadequate skills or expertise among team members for completing assigned course activities and project deliverables. This may become a difficult issue since none of are HCI majors.

Nivedhitha Dhanasekaran

Technical issues

Amanda Cheng

Other factors like unexpected events or difficulties.

Amanda Cheng

Pre-Mortem ideas organized by category

- **Red** - Insufficient preparations
- **Purple** - Unequal contributions
- **Green** - Lack of realism
- **Blue** - Insufficient communication / Human-related issues
- **Grey** - "Simply being unlucky" / Other

Project Management & Logistics

How do we schedule? — [When2Meet](#)

1. Each team member will key in the time when they are **AVAILABLE, every WEEK**, so that we can set up meetings to work on group projects
2. The teams, **Project Manager and Communication Officer** will send out **Google Calendar** invites to catch up on Zoom or in-person at a predetermined location on-campus
3. While we haven't decided on a weekly recurring schedule, we hope to get majority of our group work done on **Mondays during sections**

The screenshot shows the When2Meet interface for a group named "UCRE-Team-D1". On the left, there is a "Sign In" section with fields for "Your Name:" and "Password (optional):", a "Sign In" button, and instructions for new and returning users. On the right, there is a "Group's Availability" calendar. The calendar is a grid with days of the week (Sun, Mon, Tue, Wed, Thu, Fri, Sat) as columns and time slots (from 8:00 AM to 12:00 AM) as rows. The cells are colored in shades of green to indicate availability levels, with a legend showing "0/4 Available" (lightest green) and "4/4 Available" (darkest green). The calendar shows varying levels of availability across the week and time slots.

How will we manage our communication and links? — Slack

- **Slack thread:** team-d1 for communications
- **Resources folder in thread:** for organizing links and other resources

