Week 2

LLMs and Challenges related to Stereotypes and Biases

Large Language Models (LLMs)

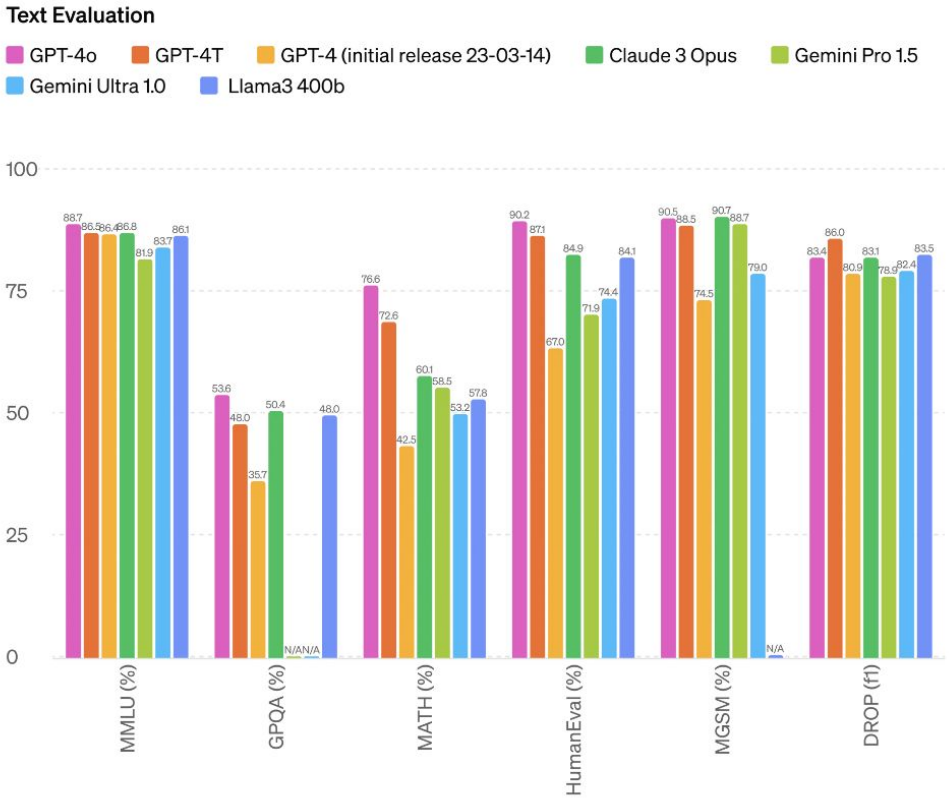
Large language models (LLM) are very **large deep learning models** that are **pre-trained on vast amounts of data**, often containing billions of words from diverse sources such as books, websites, and articles.. The underlying transformer is a set of **neural networks** that consist of an encoder and a decoder with self-attention capabilities. The encoder and decoder extract meanings from a sequence of text and understand the relationships between words and phrases in it.

LLMs can perform a wide range of language-related tasks, including translation, summarization, question-answering, and text generation.

Examples of LLMs include OpenAI's GPT-3, GPT-4, GPT-4.o, Claude 3 Opus, Google’s Gemini Pro 1.5, Gemini Ultra 1.0, Meta’s Llama3

https://aws.amazon.com/what-is/large-language-model/

Model evaluations - Text Evaluation



LLMs and Challenges related to Stereotypes and Biases

Deus Ex Machina and Personas from Large Language Models: Investigating the Composition of AI-Generated Persona Descriptions (CHI '24)

The Illusion of Artificial Inclusion (against substituting human participants with modern generative AI) (CHI '24)

Evaluating Large Language Models in Generating Synthetic HCI Research Data: a Case Study (CHI '23)

LLMs and Challenges related to Stereotypes and Biases

Deus Ex Machina and Personas from Large Language Models: Investigating the Composition of AI-Generated Persona Descriptions (CHI '24)

- The researchers investigated the **diversity and bias** in 450 personas generated by Large Language Models (LLMs), evaluated by internal evaluators and subject-matter experts (SMEs).
- The research findings **reveal biases in LLM-generated personas**, particularly in **age, occupation,** and **pain points,** as well as a **strong bias towards personas from the United States.**
- The findings suggest that **LLMs can generate consistent personas** perceived as **believable**, **relatable**, and **informative** while containing **relatively low amounts of stereotyping.**

LLMs and Challenges related to Stereotypes and Biases

Deus Ex Machina and Personas from Large Language Models: Investigating the Composition of AI-Generated Persona Descriptions (CHI '24)

Persona Generation

- 450 personas were generated using GPT-4, with **prompts specifying different genders** (male, female, one with specifying gender) **and addiction types** (alcohol, opioids, social media, online shopping, gambling).
- Then two-stage prompting strategies were used:
    - First a list of **30 "skeletal" personas** for each addiction-prompt type combination (skeletal in the sense they only contain basic information) was generated to ensure the persona descriptions are unique.
    - Then the skeletal persona were inputted back to the model, asking it to expand each persona description to create the **full persona descriptions** (i.e., "rounded personas") for analysis.

Notes: This might be a good method to avoid repetitive personas in persona generation process
The research did not compare the LLM persona with personas generated through traditional methods

LLMs and Challenges related to Stereotypes and Biases

The Illusion of Artificial Inclusion (against substituting human participants with modern generative AI) (CHI '24)

The paper surveys various "**substitution proposals**" and evaluates their arguments for **using AI instead of human participants in psychological science, user research, and AI development**. The authors find that while these proposals are motivated by reducing research costs and increasing data diversity, they ultimately **undermine the foundational values** of **work with human participants**: **representation, inclusion, and understanding**.

LLMs and Challenges related to Stereotypes and Biases

The Illusion of Artificial Inclusion (against substituting human participants with modern generative AI) (CHI '24)

- **Practical challenges** to the replacement of human participants:
  - Modern language models are **not yet ready to simulate human cognition and decision making**.
  - **"value lock-in"** (inability to update to reflect changing social norms)
  - Modern LLMs **struggle to model the wide range of opinions held across human communities,** especially **minority perspectives**. The training data used for LLMs echoes and reinforces the focus of psychology and HCI research on western, educated, industrialized, rich, and democratic people
  - User and psychology research rely on a variety of **nonlinguistic indicators** (reaction time, facial expressions, and even pupil dilation) to study and understand **human cognition and behavior**.

Notes: This paper presented a lot of constraints and challenges of LLM that we should be aware of in our research
Read references of this article

LLMs and Challenges related to Stereotypes and Biases

The Illusion of Artificial Inclusion (against substituting human participants with modern generative AI) (CHI '24)

- **Intrinsic challenges** to the replacement of human participants:
  - **The values of representation and inclusion.**
  - **The value of understanding**: Replacing human participants with AI **disrupts the intersubjectivity** between researcher and participant, undermining the goal of understanding.
  - **Intrinsic challenges across research and development** are deeply rooted in **the values of representation, inclusion, and understanding**. These challenges cannot be resolved through improved training or model performance alone, as they are fundamental to the participatory and intersubjective nature of scientific research and development .

Notes: This paper presented a lot of constraints and challenges of LLM that we should be aware of in our research
Read references of this article

LLMs and Challenges related to Stereotypes and Biases

Evaluating Large Language Models in Generating Synthetic HCI Research Data: a Case Study (CHI '23)

- This research explored the potential of large language models (LLMs) in **generating synthetic user research data**.
- **OpenAI's GPT-3** model was used to **generate open-ended questionnaire** responses about experiencing **video games as art**, a topic not tractable with traditional computational user models. Then the researchers tested whether synthetic responses can be distinguished from real responses.
- The research concluded that GPT-3 can, in this context, **yield believable accounts of HCI experiences**.
- Given the low cost and high speed of LLM data generation, synthetic data should be useful in **ideating and piloting new experiments**, although any **findings must obviously always be validated with real data.**

Notes: Can we work with small data and LLMs to generate a story (synthetic research data?) ?

LLMs and Challenges related to Stereotypes and Biases Summary

Bias and Stereotyping

- All papers highlight the presence of biases in LLM-generated content, whether in user personas, synthetic research data, or creative ideas.
- Biases often reflect societal stereotypes, such as gender roles and US-centric perspectives, which can skew the data and insights generated by LLMs.

Validation and Human Oversight

- AI-generated contents and persona should be validated with real human data to ensure accuracy and reliability.
- Involving subject-matter experts and adopting iterative evaluation processes can help address biases and improve the quality of AI-generated content.

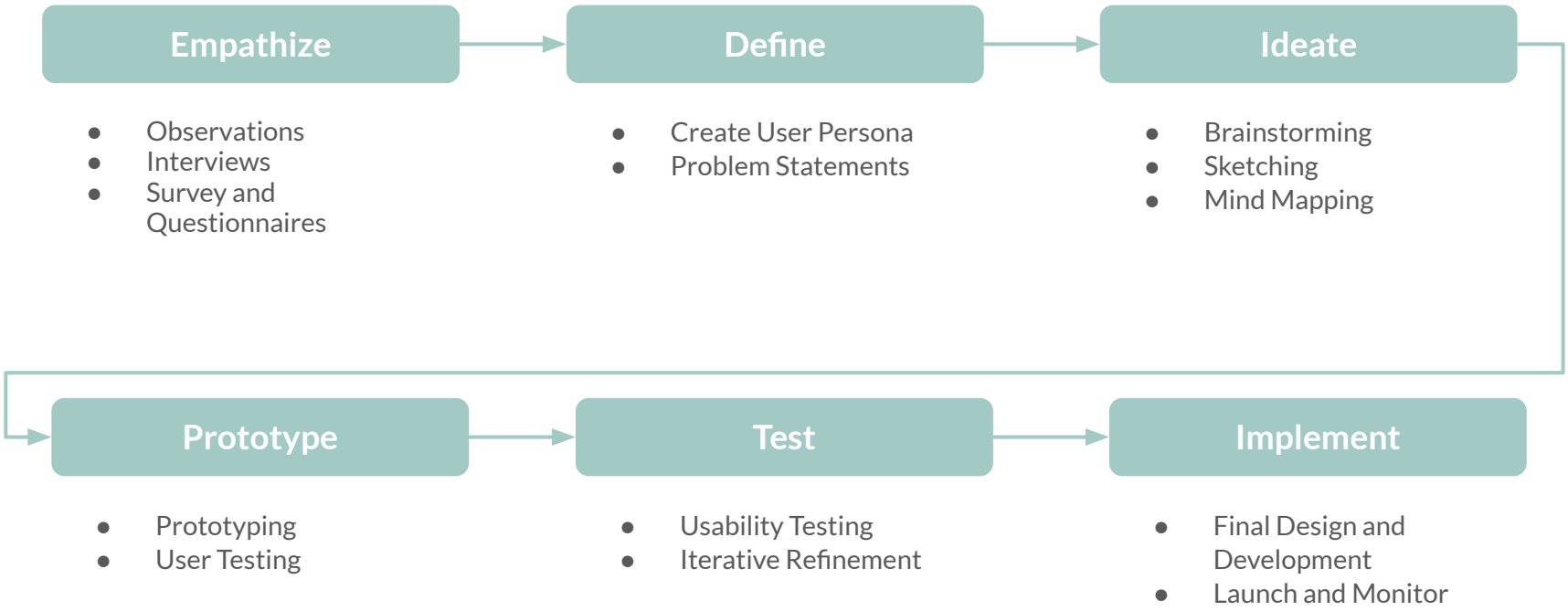
Human-centered design (HCD) phases and use of persona tool in those phases

HCD (Human-centered design)

Human-centered design (HCD) is an approach that puts human needs, capabilities, and behavior first, then designs to accommodate those needs, capabilities, and ways of behaving.

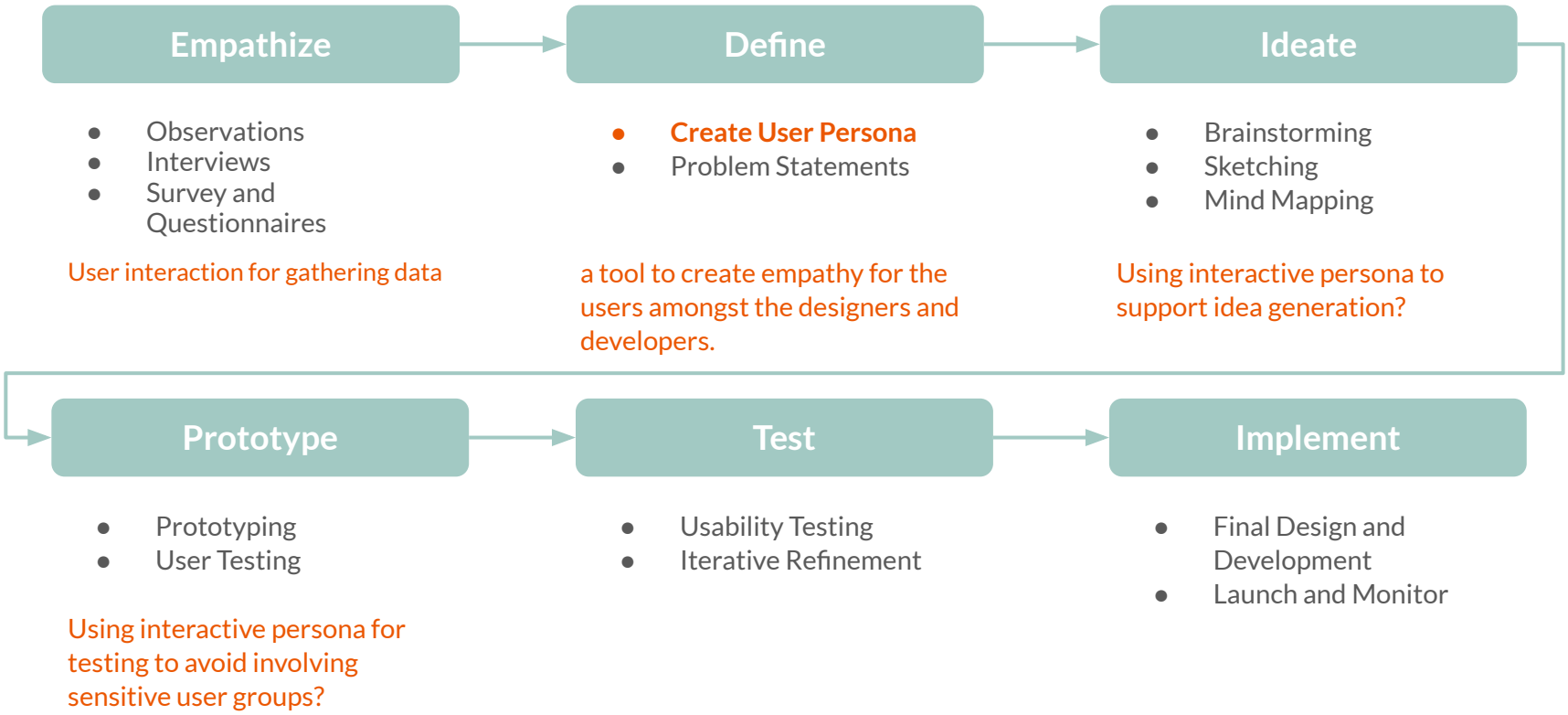
- Norman, Don. The Design of Everyday Things. Revised and Expanded Edition, Basic Books, 2013.

HCD Phases



Norman, Don. The Design of Everyday Things. Revised and Expanded Edition, Basic Books, 2013.
https://www.interaction-design.org/literature/topics/human-centered-design
https://www.designkit.org/

HCD Phases + Use of Persona Tool



Use of Persona

Design

Communication

Decision making

Persona in HCD - Research Papers

Personas in action: ethnography in an interaction design team (NordiCHI '02)

The mystique of numbers: belief in quantitative approaches to segmentation and persona development (CHI EA '10)

How do designers and user experience professionals actually perceive and use personas? (CHI '12)

Personas and decision making in the design process: an ethnographic case study (CHI '12)

Stereotypes and Politics: Reflections on Personas (CHI '16)

"What does your Agent look like?": A Drawing Study to Understand Users' Perceived Persona of Conversational Agent (CHI EA '19)

Personas and Analytics: A Comparative User Study of Efficiency and Effectiveness for a User Identification Task (CHI '20)

Personas: New Data, New Trends (CHI EA '22)

AVOCUS: A Voice Customization System for Online Personas (CHI EA '23)

Persona Co-Design for Improving Digital Accessibility (CHI EA '23)

Unraveling The Complexity: A User-Centered Design Process For Narrative Visualization (CHI EA '23)

Personas: practice and theory (DUX '03)

The mystique of numbers: belief in quantitative approaches to segmentation and persona development (CHI EA '10)

- This paper examines the tension between quantitative market research and qualitative user-centered design (UCD) research. It highlights the risks associated with relying excessively on quantitative segmentation for persona development and the dangers of deferring too much to quantitative methods.

Persona in HCD

How do designers and user experience professionals actually perceive and use personas? (CHI '12)

This is a study of how experienced user-centered design (UCD) practitioners with prior experience deploying personas, use and perceive personas in industrial software design.

- Practitioners used personas almost exclusively for **communication**, but not for design. It is used to **build support for a chosen design** or more generally **to advocate user needs**.
- Participants identified four problems with personas, finding them **abstract, impersonal, misleading and distracting**.
- Personas cannot replace **immersion in actual user data**. And rather than focusing on creating engaging personas, **it is critical to avoid persona attributes that mislead or distract**.
- 14 experienced practitioners—10 designers and 4 UX professionals from one company were interviewed (relatively small sample size)

Persona in HCD

How do designers and user experience professionals actually perceive and use personas? (CHI '12)

Reasons why designers **do not use personas** for their own design work:

1. **Personas are abstract** – it is hard to understand the abstraction process from user data to persona, so personas come across as lacking critical detail.
2. **Personas are impersonal** – the personifying details in personas fail to provide a sense of empathy.
3. **Personifying details mislead** – it is difficult to select personal details that do not create false constraints on the design problem.
4. **Personifying details distract** – personifying details make it hard to focus on the aspects of a persona that are critical for the design problem.

To avoid these problems, the practitioners wanted firsthand experience with users, or personal access to user study data. They viewed this as necessary to derive the rich understanding of users required for design.

Notes: How can we make interactive persona more useful and desired? Can we make it less abstract, and provide reasoning for persona details, providing additional context upon request?

Persona in HCD

Personas and decision making in the design process: an ethnographic case study (CHI '12)

- This discourse analysis of the decision-making sessions of designers at a top tier design firm reveals that although the designers dedicate much time researching, developing, and refining personas, personas themselves make **relatively few appearances in the designers' language during decision-making sessions**.
- Functions of Personas in Decision-Making Meetings: **Role-playing** (47.5%), **Focusing** (33.9%), **Meeting Maintenance**(10.9%), **Empathy** (3.3%), **Clarification** (2.7%), **Approximation** (1.6%)
- A benefit of personas is that they provide a "common language" for designers and clients alike "to talk about users meaningfully". However, based on the research, **at no time did a client ever refer to a persona.**
- Designers who research, develop, and refine personas **do not use those personas as a primary persuasive mechanism** during design decision making meetings and rely much more prominently on their own opinions about the design and vaguely referenced hypothetical stories.
- Involving the **entire design team** in the **persona creation process** may increase their effectiveness and frequency of use.

Notes: Can interactive persona be engaging in the design and decision making process?

Persona in HCD

Stereotypes and Politics: Reflections on Personas (CHI '16)

- While personas might help **focus on the audience, prioritize, challenge assumptions, and prevent self-referential design**, the success of the method depends on how and on what basis the persona descriptions are developed, perceived, and employed .
- Personas run the risk of **re-inscribing existing stereotypes** and following more of an I-methodological than a user-centered approach.
- The study involves **semi-structured interviews** with usability experts to explore their perceptions and how they navigate these controversies.
- **Personas can also replace the direct communication with the users when they are not available** though the direct communication with the users would always be preferred.
- Some experts highlight the importance of **empirical data** in creating valid personas, while others acknowledge the creative value of fictitious elements.

Notes: It is important to avoid stereotype and not reinforcing existing stereotypes during the persona creation process.
Working with sensitive group, persona can replace the direct communication when users are not available.
Stereotype exist in personas that were created with traditional methods. Comparison between personas that are created by LLMs and traditional quantitative methods?

Persona in HCD

Tomasz Miaskiewicz and Kenneth A Kozar. 2011. **Personas and user-centered design: How can personas benefit product design processes?** Design Studies 32, 5: 417-430.

The five most significant benefits of persona use identified in this study were:

1. focus on audience and their goals (rather than the specific limitations or opportunities presented by technology)
2. prioritize product requirements and help to determine if the right problems are being solved
3. prioritize audiences and bring about a focus on the most important audience(s)
4. challenge assumptions by bringing them to the surface and challenge long-standing (and often incorrect) organizational assumptions about the users/customers
5. prevent selfreferential design by helping individuals realize how the users are different from themselves.

Persona in HCD

Persona Co-Design for Improving Digital Accessibility (CHI EA '23)

- This paper examines approaches to **developing authentic accessibility personas** that can help improve the design of products **for persons with disabilities**.
- It describes a **nested co-design approach** involving persons with disabilities at multiple stages of the design process. This method aims to create personas that **enhance empathy and understanding**, thereby improving digital accessibility in higher education.
- Co-Design Methodology: The study employs a co-design approach, involving PwD **at all stages of the design process**. This includes **initial data collection** through surveys and interviews, **design thinking workshops**, and **iterative feedback loops**. This method helped ensure that the **resulting personas portray authentic experiences.**
- Additionally, development and evaluation will continue on incorporating interactive persona artefacts that help build a virtual encounter for target users.

Notes: Is it possible to use interactive persona in all stages of the design process to increase authentic experience?
Can the interactive personal be co-designed to enhance the authenticity?

Persona in HCD

Personas: practice and theory (DUX '03)

‘Personas’ is an interaction design technique with considerable potential for software product development. In three years of use, our colleagues and we have extended Alan Cooper’s technique to make Personas a powerful complement to other usability methods.

- The researcher outlined the psychological theory that explains **why Personas are more engaging than design based primarily on scenarios**.
- **Personas can engage team members very effectively**. They also provide a **conduit for conveying a broad range of qualitative and quantitative data**, and **focus attention on aspects of design** and use that other methods do not.

Notes: Auto-Generated Personas: Enhancing User-centered Design Practices among University Students (CHI '24) has shown that there is no significant difference in collaboration and creativity (engagement) between persona that are generated by LLM and traditional approach.

Persona in HCD | Grounding Personas

Persona cases: a technique for grounding personas (CHI '11)

“Personas are a popular technique in User-Centered Design, however their validity can be called into question. While the techniques used to developed personas and their integration with other design activities provide some measure of validity, a persona's legitimacy can be threatened by challenging its characteristics. This note presents Persona Cases: personas whose characteristics are both grounded in, and traceable to their originating source of empirical data. This approach builds on the premise that sense-making in qualitative data analysis is an argumentative activity, and aligns concepts associated with a Grounded Theory analysis with recent work on arguing the characteristics of personas. We illustrate this approach using a case study in the Critical Infrastructure Protection domain.”

Persona in HCD Summary

Personas are valuable tools in the human-centered design process, providing a means to understand and empathize with users.

Benefits of Personas:

- **Communication**: Personas are effective tools for communicating user needs and preferences to different stakeholders within a project
- **Focus on User Needs**: They help keep the design team focused on the users' goals and challenges, preventing self-referential design decisions
- **Empathy Building**: Personas foster empathy by humanizing user data, making it easier for designers to relate to and understand users' perspectives

Persona in HCD Summary

Challenges and Considerations:

- **Stereotyping**: Poorly crafted personas can reinforce stereotypes and create biased assumptions about users
- **Abstraction**: Personas sometimes fail to generate empathy if they are too abstract or lack detailed user insights
- **Limited Use in Decision-Making**: Despite their benefits, personas are often underutilized in direct decision-making processes, with designers relying more on their own opinions or other methods like user scenarios
- **Co-Design**: Involving users in the persona creation process (co-design) can enhance the authenticity and relevance of personas

Research Questions

Data

- Could LLM be used to create persona when big data are not available or sensitive to obtain?

Interactive Persona

- Can LLM be used to create interactive persona that can engage and interact with designers and address some of the issues researchers see in persona adoption?

Evaluation and Effectiveness

- What metrics and methodologies can be used to evaluate the quality and effectiveness of interactive LLM-based personas in representing people with disabilities?

Bias and Stereotype

- How to detect and monitor stereotype and bias in LLM created personas (especially towards sensitive user groups)?
- How does LLM generated persona compare to personas created from traditional methods?
- How can subject experts or users be engaged as co-creators in the persona creation process to ensure authenticity and validation?

Next Steps

Draft literature review on Overleaf

Refine research questions

Revist relevant articles in depth

Research Proposal and 1st draft of Report

Questions

How to design framework and mid-fidelity prototype?

Would the NLP team working on the design and implement of the LLM model? How does my work contribute to their process?