

# Building Interactive LLM-persona

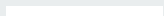
## Literature Review

Professor: Mirjana Prpa   Student: Lipeipei Sun

---

# Week 1

2024-05-22



# Literature Review

## Objectives

- Develop a research question from gaps and opportunities in the literature review

## Scope of the literature

- Literature under CHI (Conference on Human Factors in Computing Systems)
- Topics to include:
  - Persona tool in HCI (knowledge graph)
  - LLMs and Challenges related to Stereotypes and Biases
  - Human-centered design (HCD) phases and use of persona tool in those phases

# Persona

Personas are **composite archetypes** created from **user behavior patterns identified during research**, serving as **powerful tools for understanding and communicating about user goals and behaviors in specific contexts**. They must be developed with **rigor and sophistication**, **avoiding stereotypes** and ensuring they accurately represent a meaningful cross section of users.

- Cooper, Alan, et al. About Face: The Essentials of Interaction Design. 4th ed., Wiley, 2014.

Personas are intended to be used during the design and development of technology to avoid the problem of designers, developers, or stakeholders invoking the “elastic user” who can be bent and stretched to suit the needs of the invoker. Cooper estimates that each design problem will require between 3 to 12 personas. (from: How do designers and user experience professionals actually perceive and use personas?)

# HCI (Human-Computer Interaction)

The concept of Human-Computer Interaction (HCI) is centered around the study and practice of **designing user interfaces** that facilitate **effective interaction between humans and computers**.

- Card, Stuart K., Thomas P. Moran, and Allen Newell. The Psychology of Human-Computer Interaction. Lawrence Erlbaum Associates, 1983.

# Persona tool in HCI

---

# Persona tool in HCI

- Persona tool in HCI (Knowledge graph)
- Approaches for Designing Persona
- Interactive Persona

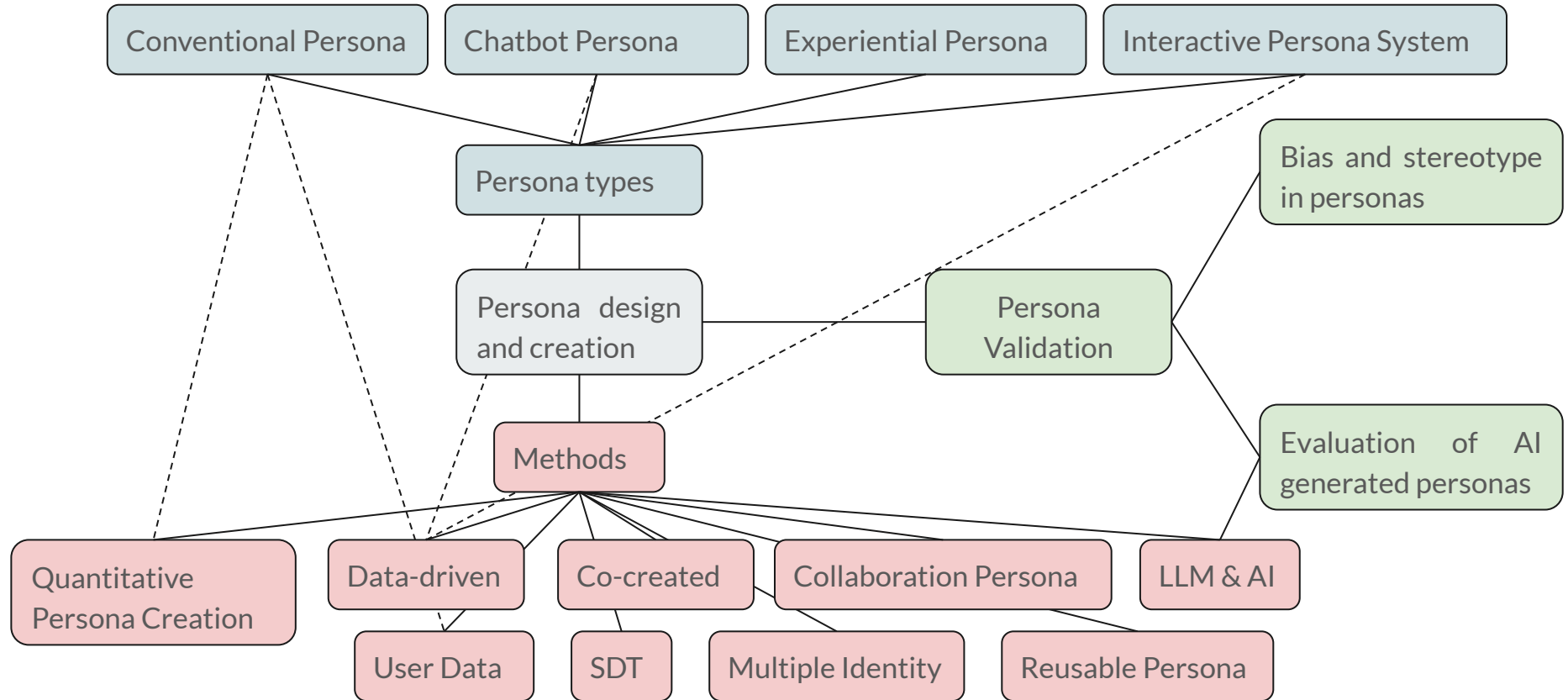
# Persona tool in HCI

## Knowledge Graph

---



# Persona tool in HCI (Knowledge Graph)



# Most Recent Persona Studies | CHI 24

Deus Ex Machina and Personas from **Large Language Models**: Investigating the Composition of **AI-Generated Persona** Descriptions

CloChat: Understanding How People **Customize, Interact, and Experience** Personas in **Large Language Models**

**Auto-Generated** Personas: Enhancing User-centered Design Practices among University Students

**Generating** personas using **LLMs** and assessing their viability

Writer-Defined **AI Personas** for On-Demand Feedback Generation

Challenges and Opportunities of **LLM-Based Synthetic Personae** and Data in HCI

The Illusion of **Artificial Inclusion** (against substituting human participants with modern **generative AI**)

# Approaches for Designing Persona

---

# Approaches for Designing Persona

- User Data
- Quantitative Persona Creation (QPC)
- Data-driven
- Co-created
- Collaboration Persona
- LLM & AI
- Self-Determination Theory (SDT)
- Multiple Identity
- Reusable Persona

# Approaches for Designing Persona

- **Data-driven**
  - How do designers and user experience professionals actually perceive and use personas? (CHI '12)
  - A Literature Review of Quantitative Persona Creation (CHI '20)
  - Persona development for information-rich domains (CHI EA '03)
  - Data-Driven Persona Development (CHI '08)
  - Data-driven Personas: Constructing Archetypal Users with Clickstreams and User Telemetry (CHI '16)
  - Persona Generation from Aggregated Social Media Data (CHI EA '17)
  - Findings of a User Study of Automatically Generated Personas (CHI '18)
  - Creating Manageable Persona Sets from Large User Populations (CHI EA '19)
  - Personas Changing Over Time: Analyzing Variations of Data-Driven Personas During a Two-Year Period (CHI EA '19)
- **LLM & AI**
  - Writer-Defined AI Personas for On-Demand Feedback Generation (CHI '24)
  - Auto-Generated Personas: Enhancing User-centered Design Practices among University Students (CHI '24)
  - Challenges and Opportunities of LLM-Based Synthetic Personae and Data in HCI (CHI EA '24)
  - Deus Ex Machina and Personas from Large Language Models: Investigating the Composition of AI-Generated Persona Descriptions (CHI '24)
- **Latent Semantic Analysis (LSA)**
  - A latent semantic analysis methodology for the identification and creation of personas (CHI '08)
- **Co-created**
  - Co-Created Personas: Engaging and Empowering Users with Diverse Needs Within the Design Process (CHI '19)
  - Co-constructing child personas for health-promoting services with vulnerable children (CHI '14)
- **Collaboration Persona**
  - Collaboration personas: a new approach to designing workplace collaboration tools (CHI '11)
  - Comparing collaboration and individual personas for the design and evaluation of collaboration software (CHI '12)
  - Emphasizing dysfunctional group dynamics in collaboration personas: specification of an approach (CHI EA '13)
- **Self-Determination Theory (SDT)**
  - Personas and Behavioral Theories: A Case Study Using Self-Determination Theory to Construct Overweight Personas (CHI '17)
- **Multiple Identity**
  - Personas and Identity: Looking at Multiple Identities to Inform the Construction of Personas (CHI '19)
- **Reusable Persona**
  - Towards Reusable Personas for Everyday Design (CHI EA '16)

# Approaches for Designing Persona

- **Data-driven**
  - How do designers and user experience professionals actually perceive and use personas? (CHI '12)
  - A Literature Review of Quantitative Persona Creation (CHI '20)
  - Persona development for information-rich domains (CHI EA '03)
  - Data-Driven Persona Development (CHI '08)
  - Data-driven Personas: Constructing Archetypal Users with Clickstreams and User Telemetry (CHI '16)
  - Persona Generation from Aggregated Social Media Data (CHI EA '17)
  - Findings of a User Study of Automatically Generated Personas (CHI '18)
  - Creating Manageable Persona Sets from Large User Populations (CHI EA '19)
  - Personas Changing Over Time: Analyzing Variations of Data-Driven Personas During a Two-Year Period (CHI EA '19)
- **LLM & AI**
  - Writer-Defined AI Personas for On-Demand Feedback Generation (CHI '24)
  - Auto-Generated Personas: Enhancing User-centered Design Practices among University Students (CHI '24)
  - Challenges and Opportunities of LLM-Based Synthetic Personae and Data in HCI (CHI EA '24)
  - Deus Ex Machina and Personas from Large Language Models: Investigating the Composition of AI-Generated Persona Descriptions (CHI '24)
- **Latent Semantic Analysis (LSA)**
  - A latent semantic analysis methodology for the identification and creation of personas (CHI '08)
- **Co-created**
  - Co-Created Personas: Engaging and Empowering Users with Diverse Needs Within the Design Process (CHI '19)
  - Co-constructing child personas for health-promoting services with vulnerable children (CHI '14)
- **Collaboration Persona**
  - Collaboration personas: a new approach to designing workplace collaboration tools (CHI '11)
  - Comparing collaboration and individual personas for the design and evaluation of collaboration software (CHI '12)
  - Emphasizing dysfunctional group dynamics in collaboration personas: specification of an approach (CHI EA '13)
- **Self-Determination Theory (SDT)**
  - Personas and Behavioral Theories: A Case Study Using Self-Determination Theory to Construct Overweight Personas (CHI '17)
- **Multiple Identity**
  - Personas and Identity: Looking at Multiple Identities to Inform the Construction of Personas (CHI '19)
- **Reusable Persona**
  - Towards Reusable Personas for Everyday Design (CHI EA '16)

# Approaches for Designing Persona | Data-driven

## Qualitative and Quantitative

- Interviews
- Observations
- Survey data
- Statistical data
- Clickstreams
- Online social media data

# Approaches for Designing Persona | Data-driven

## Persona development for information-rich domains (CHI EA '03)

- The paper discusses the creation of personas for **complex websites** by understanding **user information needs** and **mental models**.
- It introduces a **statistical method** to **identify groupings of these needs**, which designers can use alongside **data from interviews and observations** to **generate and refine personas**.
- Includes both **quantitative** and **qualitative data**



# Approaches for Designing Persona | Data-driven

## Data-Driven Persona Development (CHI '08)

- introduces a novel method for **creating** and **validating personas** based on **statistical analysis of customer data**, making the process **fast** and **cost-effective**
- **define the persona attributes** that were **most meaningful to the stakeholders**, created **survey** based on the **stakeholder feedbacks**
- from survey responses, the team performed **factor analysis** to allow the persona groups to emerge
- **Advantages: real customer data, stakeholder involvement, cost-effective, and organic persona creation**

## Data-driven Personas: Constructing Archetypal Users with Clickstreams and User Telemetry (CHI '16)

- a **quantitative bottom-up data-driven** approach to create personas
- user behavior via **clickstreams** gathered automatically from **telemetry data**
- structure the data into 10 workflows **via hierarchical clustering**
- use **mixed models, a statistical approach** that incorporates these clustered workflows to create five representative personas
- **Advantages: scalable, cost-effective, and reflective of actual user behavior**

# Approaches for Designing Persona | Data-driven User Data

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

- Identifies **limits** to persona approach
- “**Practitioners** used personas almost **exclusively for communication**, but **not for design**. 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**.”
- “Participants found it **difficult to understand the abstraction process from user data to persona**, so personas came across as **lacking critical detail required for design** .”

# Approaches for Designing Persona | Data-driven

## Persona Generation from Aggregated Social Media Data (CHI EA '17)

- introduces a methodology for generating personas using **large-scale, real-time social media data**
- The “methodology can first identify both **distinct and impactful user segments** and then **create persona descriptions by automatically adding pertinent features**, such as names, photos, and personal attributes.”

## Findings of a User Study of Automatically Generated Personas (CHI 18)

- The paper reports on a semi-naturalistic user study of the Automatic Persona Generation (APG) system using **large-scale audience data** from **social media channels**.
- The research “results show that **having an interactive system may aid in keeping personas at the forefront while making customer-centric decisions** and indicate that **data-driven personas fulfill information needs of decision makers by mixing personas and numerical data**”.

Notes: Can we work with small data with LLMs to generate a story (synthetic research data?) and provide context? How to use it for communication?

# Approaches for Designing Persona | Data-driven

## Creating Manageable Persona Sets from Large User Populations (CHI EA '19)

- Problem: modern online systems often provide **big data** from millions of users that display **vastly different behaviors**, resulting in possibly **thousands of personas** representing the **entire user population**
- present a technique for **reducing the number of personas** to a smaller number that **efficiently represents the complete user population**, while being more **manageable** for end users of personas
- Approach: the researchers first isolate the key user behaviors and demographical attributes, creating **thin personas**, and we then apply **an algorithmic cost function** to collapse the set to the **minimum needed** to represent the whole population.

# Approaches for Designing Persona | Data-driven

## Detecting Demographic Bias in Automatically Generated Personas (CHI EA '19)

- investigates the presence of **demographic bias** in personas generated automatically from **large-scale social media data**
- The bias is highest when doing an **exact match comparison**, and the bias decreases when comparing at age or gender level. The **bias also decreases** when **increasing the number of generated personas**.

Notes: Can we use LLMs to increase the number of personas, therefore we can reduce the bias?

# Approaches for Designing Persona | Data-driven

## Personas Changing Over Time: Analyzing Variations of Data-Driven Personas During a Two-Year Period (CHI EA '19)

- Problem: **Personas' underlying data may become stale, necessitating ongoing data collection.**
- Study: Collected monthly demographic data over two years for a large online content publisher.
- Method: Generated fifteen personas each month using the same algorithm.
- Findings:
  - 18.7% average monthly change in personas.
  - 23.3% average yearly change.
  - 47% change over two years.
- Conclusion: **Personas change over time; organizations should continuously collect data to monitor these changes.**

# Approaches for Designing Persona | Data-driven Quantitative Persona Creation (QPC)

## A Literature Review of Quantitative Persona Creation (CHI '20)

- Quantitative persona creation (QPC) has **tremendous potential**, as HCI researchers and practitioners can leverage **user data** from **online analytics and digital media platforms** to better understand their users and customers.
- a lack of a systematic overview of the QPC methods and progress made, with no standard methodology or known best practices
- To address this gap, the team reviewed “49 **QPC research articles** from 2005 to 2019. Results indicate three stages of QPC research: **Emergence, Diversification, and Sophistication.**”
- “Sharing resources, such as datasets, code, and algorithms, is crucial to achieving the next stage (**Maturity**).”

# Approaches for Designing Persona | Data-driven

## The Ethics of Data-Driven Personas (CHI EA '20)

- This research paper delves into the ethical considerations surrounding Data-Driven Persona Creation (DDPC).
- DDPC utilizes **large datasets and machine learning algorithms** to generate user personas, which are fictional characters representing different user segments.
- The use of these techniques aims to **streamline product development, enhance organizational strategy, and improve customer operations.**
- The authors highlight several **ethical concerns** that arise from this practice using Gillespie's framework of algorithmic ethics (includes six ethical dimensions)
- The paper advocates for a **comprehensive ethical analysis at every step of the DDPC process to ensure fairness and transparency.**



# Summary for Data-driven Persona

## Issue in traditional persona process

- Traditional methods rely on **qualitative data** like surveys and interviews, which is **time-consuming**, **expensive**, and can be **outdated**
- Traditional methods may not reflect **actual user behavior accurately**, and are based on **limited sample sizes**

## Methods: statistical analysis, automatic persona generation

### Benefits

- **Accuracy:** personas are grounded in user behavior data
- **Scalability:** the methods deal with large data which is suitable for large user bases
- **Efficiency and Cost Effective:** the proposed methods are mostly automated process which aims to be more efficient and reduce the cost comparing to traditional persona development
- **Up-to-date:** some of the methods are based on real-time data or data that can be updated relatively quickly, so it helps to keep the persona up to date

# Summary for Data-driven Persona

## Challenges and Considerations

- **Potential Bias in Data-Driven Personas** - How to improve algorithm to avoid bias in persona creation
- **Combining Quantitative and Qualitative Data** - while quantitative data represents the user behavior with high accuracy, the qualitative data still remain valuable for adding depth and context to personas
- **Keep Data Up-to-date** - to keep persona up-to-date and accurate, it is important to constantly collecting data to monitor the updates

Notes: Qualitative: (expensive, costly, sensitive group)

Quantitative: Data driven We are using social media data(needs a lot of data). How does LLMs fits in? Can we use small data with specilized form and expand it to make it more conversational? Then interview the LLMs to get more qualitative data.

Can LLM persona provide qualitative data? Focus on sensitive group. We need to be aware of bias and stereotype. Ethical lense is important.

# Approaches for Designing Persona | LLM & AI

## Writer-Defined AI Personas for On-Demand Feedback Generation (CHI '24)

- This paper proposed an concept that **generates on-demand feedbacks**, based on **writer-defined AI personas** of any target audience using a prototype text editor called Impressona.
- Based on GPT-3.5
- Conducted with 16 participants who found the concept helpful for **gaining diverse perspectives** but **faced challenges in defining personas and dealing with verbose feedback**.

# Approaches for Designing Persona | LLM & AI

## Auto-Generated Personas: Enhancing User-centered Design Practices among University Students (CHI '24)

- Traditional persona-building relies on **interviews and ethnography**.
- Developed an auto-generating persona system to improve UCD course activities.
- System is based on GPT-4, DALL-E 2, and knowledge graphs.
- Features of the system:
  - Automated processing of survey data.
  - Automatic generation of 2D avatars.
  - Options for automatic or customized entity generation.
- Evaluation involved 22 participants.

# Approaches for Designing Persona | LLM & AI

## Auto-Generated Personas: Enhancing User-centered Design Practices among University Students (CHI '24)

### Findings:

- Quantitative: Significant **improvements in efficiency, satisfaction, accuracy, and diversity with the automated system** compared to traditional methods. **No significant difference in collaboration and creativity.**
- Qualitative: Positive feedback on the system's efficiency, accuracy, and informative visualization. Negative feedback on UI/UX, complexity, and design style of persona templates.

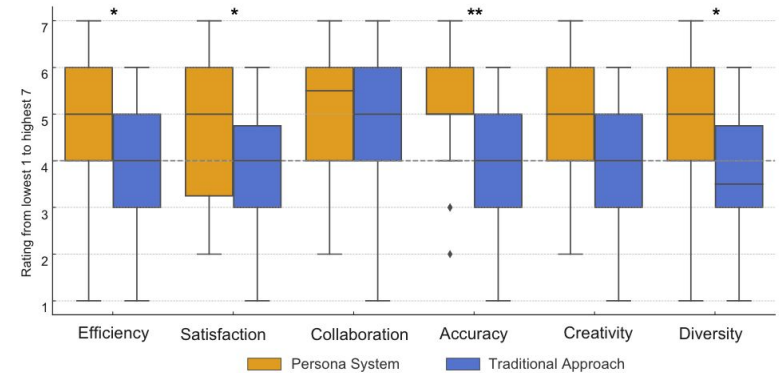


Figure: The figure shows the results of the preliminary persona questionnaire

Notes: this is a great step stone for our work. Can we go a step further to let persona be interactive?

# Approaches for Designing Persona | LLM & AI

## Challenges and Opportunities of LLM-Based Synthetic Persona and Data in HCI (CHI EA '24)

- **Synthetic persona and AI-generated data** are emerging in HCI fields such as education, training, and gaming.
- **Large Language Models (LLMs)** show promise for creating synthetic AI personae and data.
- A proposed workshop will engage HCI researchers to **explore and develop future HCI research with LLMs, focusing on opportunities and challenges.**
- Workshop outcomes may be shared through scientific publications or special issues.

# Approaches for Designing Persona | LLM & AI

## 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**.

Notes: LLMs are trained with text, most text are in English  
Does LLMs show biases based on data that were trained on?  
How did the personas build?

# Summary for LLM & AI Generated Persona

Base model: GPT-3.5, GPT 4

## Advantages

- On-demand feedbacks
- **Improved User Centered Design (UCD)**: improvements in efficiency, satisfaction, accuracy, and diversity compared to traditional methods
- **LLMs can generate consistent personas perceived as believable, relatable, and informative** while containing relatively low amounts of stereotyping.

## Challenges and Consideration

- **Bias and Stereotyping**: biases may exist in AI-generated personas, particularly in terms of age, occupation, and geographic representation
- **Lack of Understanding and Control**: LLMs often function as "black boxes", making it challenging for researchers to predict and control their outputs



# Summary for LLM & AI Generated Persona

## Gaps and Opportunities

- How to detect **biases and stereotypes** in the process in order to reduce and correct the biases and stereotypes?
- How to improve the **transparency and understanding of LLM** for designers?
- **Limited Studies on Data-Driven Approaches:** Most persona generation methods are based on writer-defined prompts, survey data, or specific topic prompts. There seem to be a lack of research on fully data-driven persona generation through LLMs.

# Approaches for Designing Persona | Latent Semantic Analysis (LSA)

A latent semantic analysis methodology for the identification and creation of personas (CHI '08)

- This research develops a new methodology for the **identification and creation of personas** through the application of **Latent Semantic Analysis (LSA)**.
- An application of the LSA methodology is provided in the context of the design of an **Institutional Repository system**.
- The LSA methodology helps overcome some of the drawbacks of current methods for the identification and creation of personas, and makes the process **less subjective, more efficient**, and **less reliant on specialized skills**.

# Approaches for Designing Persona | Co-Created

## Co-Created Personas: Engaging and Empowering Users with Diverse Needs Within the Design Process (CHI '19)

- exploring **co-created personas** as a technique to use in co-design with users who have diverse needs
- the vision was that this would **broaden the demographic** and **liberate co-designers** of their personal relationship with a **health condition**
- Findings revealed that the co-created personas **encouraged users** with diverse needs to **engage with co-designing**
- additional benefits including **empowering users** within a **more accessible design process**

Notes: Can we use this to avoid bias and stereotype? How can we have this verified by subject experts? At what stage do they get feedbacks?

# Approaches for Designing Persona | Co-Created

## Co-constructing child personas for health-promoting services with vulnerable children (CHI '14)

- a new method for co-constructing child-personas that are tailored for developing health-promoting services where empirical data is restricted due to practical and ethical reasons
- focus children design **workshop sessions** on salutogenesis, and complement this with a pathogenic perspective by **interviewing healthcare professionals and parents**
- introduce the use of **proxy personas**, and redemption scenarios in the form of **comicboards**, both collaboratively constructed by children and designers through storytelling

Notes: Proxy persona? Interactive Proxy persona that can advocates for the persona?

# Approaches for Designing Persona | Decision Diagram

## Revisiting personas: the making-of for special user groups (CHI EA '12)

- **Problems:** The diversity of **special user groups**, i.e. **elderly** from 50 to 90 years and **children** from 6 to 14 years, is huge. **Assessing their requirements** is challenging, as it requires sensitivity in terms of choosing an **appropriate approach to collect data**. Furthermore, the **illustration of the data** for the subsequent design process can be difficult, if different partners are involved in a project.
- **Proposal:** “we are exploring a **decision diagram** for the **creation** of personas. It aims at **identifying the most appropriate approach** (i.e. **qualitative** and/or **quantitative** data collection), taking into account the characteristics of the special user groups among other aspects.”

# Approaches for Designing Persona | Collaboration Personas

Collaboration personas: a new approach to designing workplace collaboration tools (CHI '11)

- propose the notion of **collaboration personas**, which are empirically derived descriptions of **hypothetical groups**, including details that inform the design of **collaboration tools**
- Collaboration personas differ from individual personas in having
  - multiple, inter-related individuals playing specific roles
  - a focus on collective goals and elaboration of individual goals that affect the collective goal
  - new attributes that characterize collaborative aspects of the group's work

Comparing collaboration and individual personas for the design and evaluation of collaboration software (CHI '12)

- a comparative study of design and user experience practitioners who used both **collaboration personas** and **individual personas**
- **Collaboration personas** led to a more complete discussion, as indicated by a greater amount of time spent on the task compared to individual personas

# Approaches for Designing Persona | Collaboration Personas

Emphasizing dysfunctional group dynamics in collaboration personas: specification of an approach (CHI EA '13)

- Built on top of the *Collaboration Personas and Individual Personas for the design and evaluation of collaboration software*, Judge, Matthews, and Whittaker (2012)
- put more emphasis on **problematic** or **dysfunctional group dynamics**
- Outlines improvements to collaboration persona

# Approaches for Designing Persona | Self-Determination Theory (SDT)

## Personas and Behavioral Theories: A Case Study Using Self-Determination Theory to Construct Overweight Personas (CHI '17)

Personas are a widely used tool to keep real users in mind, while avoiding stereotypical thinking in the design process. Yet, creating personas can be challenging. Starting from Cooper's approach for constructing personas, this paper details how behavioral theory can contribute substantially to the development of personas. We describe a case study in which Self-Determination Theory (SDT) is used to develop five distinctive personas for the design of a digital coach for sustainable weight loss. We show how behavioral theories such as SDT can help to understand what genuinely drives and motivates users to sustainably change their behavior. In our study, we used SDT to prepare and analyze interviews with envisioned users of the coach and to create complex, yet engaging and highly realistic personas that make users' basic psychological needs explicit. The paper ends with a critical reflection on the use of behavioral theories to create personas, discussing both challenges and strengths.



# Approaches for Designing Persona | Multiple Identities

Personas and Identity: Looking at Multiple Identities to Inform the Construction of Personas (CHI '19)

Personas are valuable tools to help designers get to know their users and adopt their perspectives. Yet people are complex and multiple identities have to be considered in their interplay to account for a comprehensive representation otherwise, personas might be **superficial** and prone to **activate stereotypes**. Therefore, the way users' identities are presented in a limited set of personas is crucial to account for diversity and highlight facets which otherwise would go unnoticed. In this paper, we introduce an approach to the development of personas informed by **social identity theory**. The effectiveness of this approach is investigated in a qualitative study in the context of the design process for an e-learning platform for women in tech. The results suggest that considering multiple identities in the construction of personas adds value when designing technologies.

# Approaches for Designing Persona | Reusable Persona

## Towards Reusable Personas for Everyday Design (CHI EA '16)

- “Persona development is rooted in the rigorous collection and analysis of data **specifically related to the design project being undertaken**”
- This paper reports on ongoing research into the development of **reusable personas** for use by **nonexpert, everyday designers** to deal with **small scale but diverse design challenges** for which they cannot carry out user research and modelling
- The proposed reusable persona “represent **populations** and **practices** rather than **people** and **products**”.

# Other Approaches for Designing Persona

Persona Perception Scale: Developing and Validating an Instrument for Human-Like Representations of Data (CHI EA '18)

Picturing It!: The Effect of Image Styles on User Perceptions of Personas (CHI '21)

Three Tensions Between Personas and Complex Disability Identities (CHI EA '20)

Towards a Measurement Scale of Organizational Readiness for Personas (CHI EA '21)

# Interactive Persona

---

# Interactive Persona

Bot Personas as Off-The-Shelf Users (CHI 17)

Findings of a User Study of Automatically Generated Personas (CHI 18)

Persona Analytics: Implementing Mouse-Tracking for an Interactive Persona System (CHI 21)

Implementing Eye-Tracking for Persona Analytics (CHI 21)

Experiential Persona: Towards Supporting Richer and Unfinalized Representations of People (CHI 21)

Creating More Personas Improves Representation of Demographically Diverse Populations: Implications Towards Interactive Persona Systems (CHI 22)

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

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

Developing Persona Analytics Towards Persona Science (Non CHI)

Play-persona: a multifaceted concept (Non CHI)

# Interactive Persona Types

## Chatbot inspired interactive persona

- Bot Personas as Off-The-Shelf Users (CHI 17)

## Experiential persona

- Experiential Persona: Towards Supporting Richer and Unfinalized Representations of People (CHI EA '21)

## Interactive persona system

- Findings of a User Study of Automatically Generated Personas (CHI 18)
- Creating More Personas Improves Representation of Demographically Diverse Populations: Implications Towards Interactive Persona Systems (NordiCHI '22)

# Interactive Persona | Chatbot Inspired

## Bot Personas as Off-The-Shelf Users (CHI 17)

- These interactive personas act as **off-the-shelf users**.
- The interactive bot personas is inspired by **chatbots** and provide an **interface for design teams to interact with rich user data** throughout the design process.
- Bot personas use data-driven methods and machine learning to **simulate user behavior**
- “Bot personas, as an **interface on rich user data**, give that data a face.”
- Practical advantages: **efficiency**, easy and wide **access**, **no incentives required**, offer **scalability**, and provide a **practical alternative to real users**.

# Interactive Persona | Experiential Persona

Experiential Persona: Towards Supporting Richer and Unfinalized Representations of People (CHI EA '21)

- The paper presents **Experiential Persona** that consists of a **carefully curated, staged collection of artifacts**, as an alternative to 2D persona
- Experiential Persona allows designers to **interact with and explore the artifacts**, individually and as a collection, **to imagine and experience the world of ‘the user’**.
- “This more embodied, interactive and open-ended persona can potentially **support richer sense making; encouraging a more open, emergent, and unfinalized view of people** we design for”.



# Interactive Persona | Interactive Persona System

## Findings of a User Study of Automatically Generated Personas (CHI 18)

- The paper reports on a semi-naturalistic user study of the Automatic Persona Generation (APG) system.
- The research “results show that having an **interactive system** may aid in **keeping personas at the forefront while making customer-centric decisions** and indicate that **data-driven personas fulfill information needs of decision makers by mixing personas and numerical data**”.
- The APG system offered **interactive features** that expanded traditional static persona profiles. Users could dynamically **generate and interact with personas, content, and reach metrics**.
- The system supported **real-time data updates** and interactive exploration, allowing users to engage with up-to-date and relevant persona information.

# Interactive Persona | Interactive Persona System

## Creating More Personas Improves Representation of Demographically Diverse Populations: Implications Towards Interactive Persona Systems (NordiCHI '22)

- The research focused on a **statistically optimal number of personas**.
- The research found that “more personas cover more age groups and countries, thus improving the **statistical correspondence** with the raw user data, and **increasing the representation of demographic diversity** by including more fringe user segments”.
- The research “further demonstrate **how an interactive persona system can help stakeholders navigate many personas with possibly smaller cognitive effort**”.
- The paper discusses how **interactive systems can help stakeholders manage and navigate a large number of personas**

Notes: Findings suggest that generating more personas results in a better representation of diverse demographic groups, with the optimal number for significant diversity gains being around **40 personas**.

# Evaluate/Measurement of Interactive Persona Use

## Implementing Eye-Tracking for Persona Analytics (CHI 21)

- This research investigated in **users' engagement with interactive persona system**.
- “implementation of online eye-tracking within an interactive persona system, which can be used for studying questions such as:
  - How are personas actually used?
  - What information do users most interact with?
  - How do users browse and select personas?”

## Persona Analytics: Implementing Mouse-Tracking for an Interactive Persona System (CHI 21)

- The researcher “introduce and evaluate an **implementation of persona analytics based on mouse tracking**, which offers researchers new possibilities for conducting persona user studies, especially during times when in-person user studies are challenging to carry out”.

# Summary for Interactive Persona

## Advantages

- **Real-Time Updates:** Personas can be continuously updated based on new data, ensuring they remain relevant.
- **Improves Representation of Demographically Diverse:** Creating more persona increases the representation of demographic diversity.
- **Enhanced Accessibility:** Web-based personas can be accessed from any device, facilitating wider usage among design teams.
- **User Interaction:** Interactive features allow users to engage with personas through actions like filtering, sorting, and analyzing specific attributes.

# Summary for Interactive Persona

## Challenges and Consideration

- Interactive personas often rely heavily on quantitative data. **Balancing quantitative and qualitative data** might provide a more comprehensive overview of the user group.
- Ensuring **data accuracy** and keeping data **up-to-date**
- As interactive persona systems become more detailed and complex, ensuring that they remain **user-friendly** can be challenging.
- **Chat-bot inspired persona** offers an innovative way to provide **interactive and dynamic user representations**. However the application of chat-bot inspired interactive persona has not been thoroughly studied or tested.

# Next steps

Action items out of the meeting:

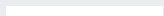
- Add relevant comments to the literature review
- Start writing literature review on Overleaf
- Revist relevant articles in depth

Literature review:

- LLMs and Challenges related to Stereotypes and Biases
- Human-centered design (HCD) phases and use of persona tool in those phases

# Week 2

2024-05-29



# 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.

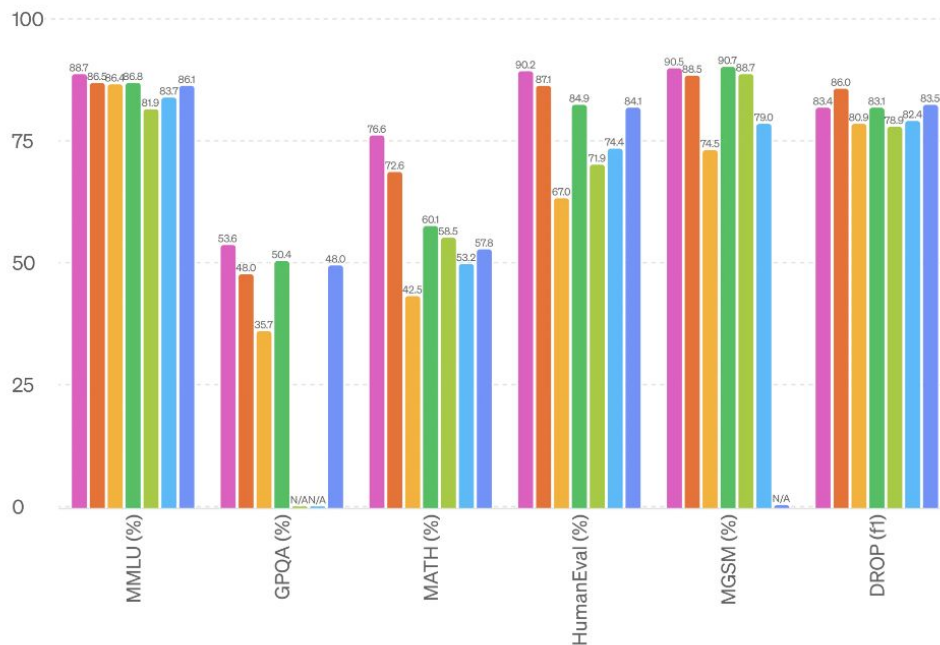
Examples of LLMs include OpenAI's GPT-3, GPT-4, GPT-4.o, Claude 3 Opus, Google's Gemini Pro 1.5, Gemini Ultra 1.0, Meta's Llama3

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

# Model evaluations - Text Evaluation

## Text Evaluation

■ GPT-4o ■ GPT-4T ■ GPT-4 (initial release 23-03-14) ■ Claude 3 Opus ■ Gemini Pro 1.5  
■ Gemini Ultra 1.0 ■ Llama3 400b



# 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

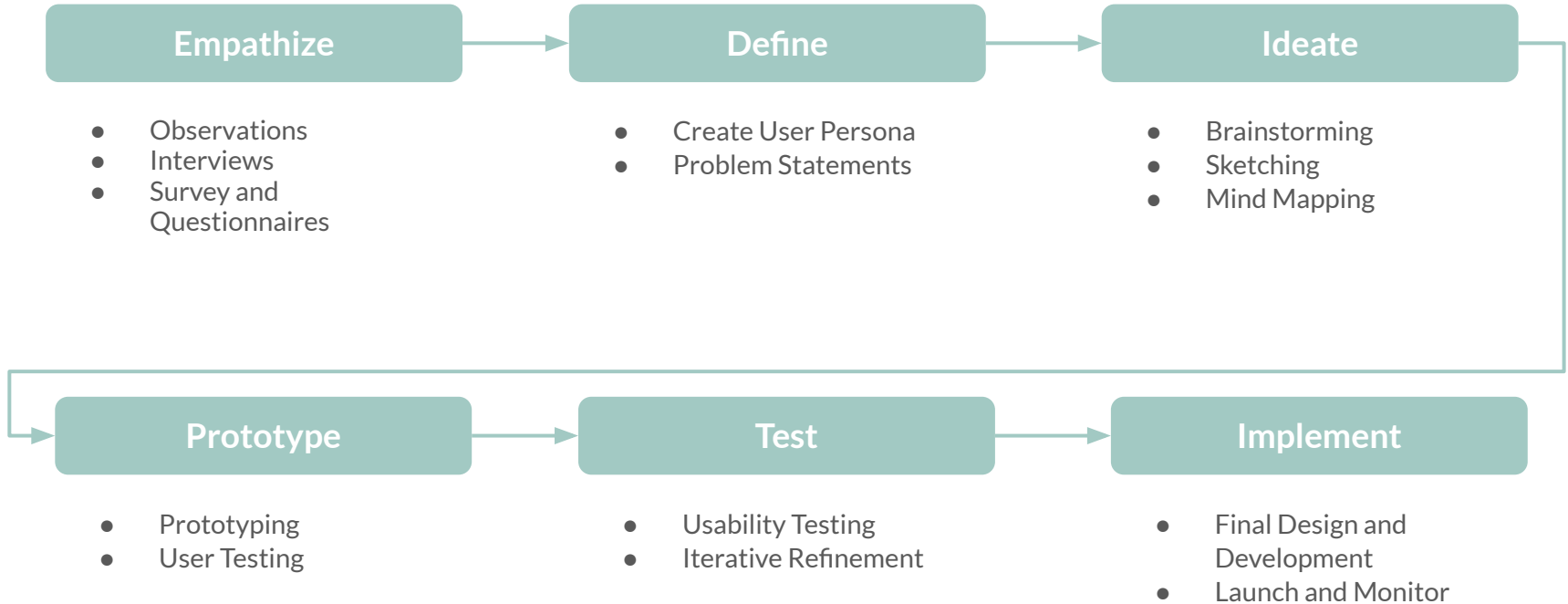
---

# 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

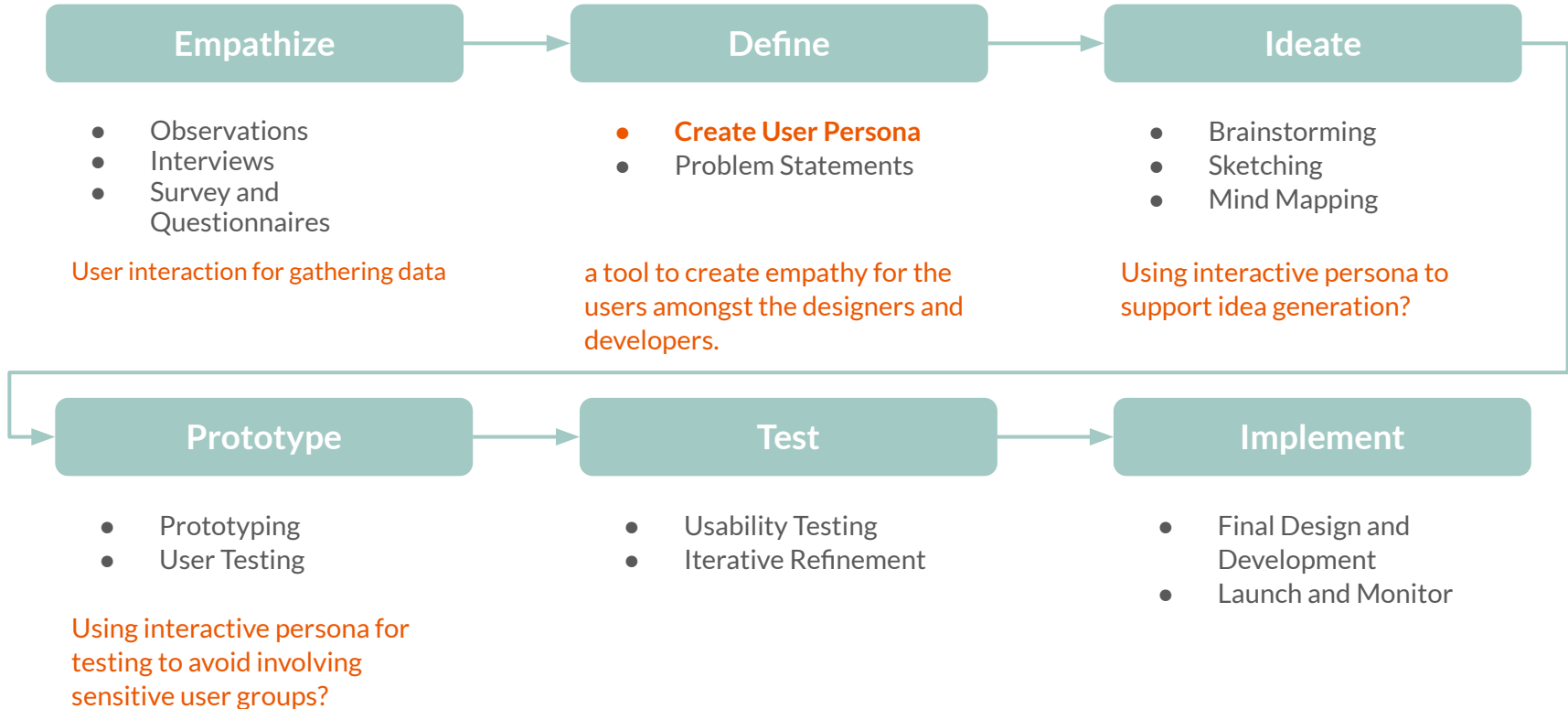


Norman, Don. The Design of Everyday Things. Revised and Expanded Edition, Basic Books, 2013.

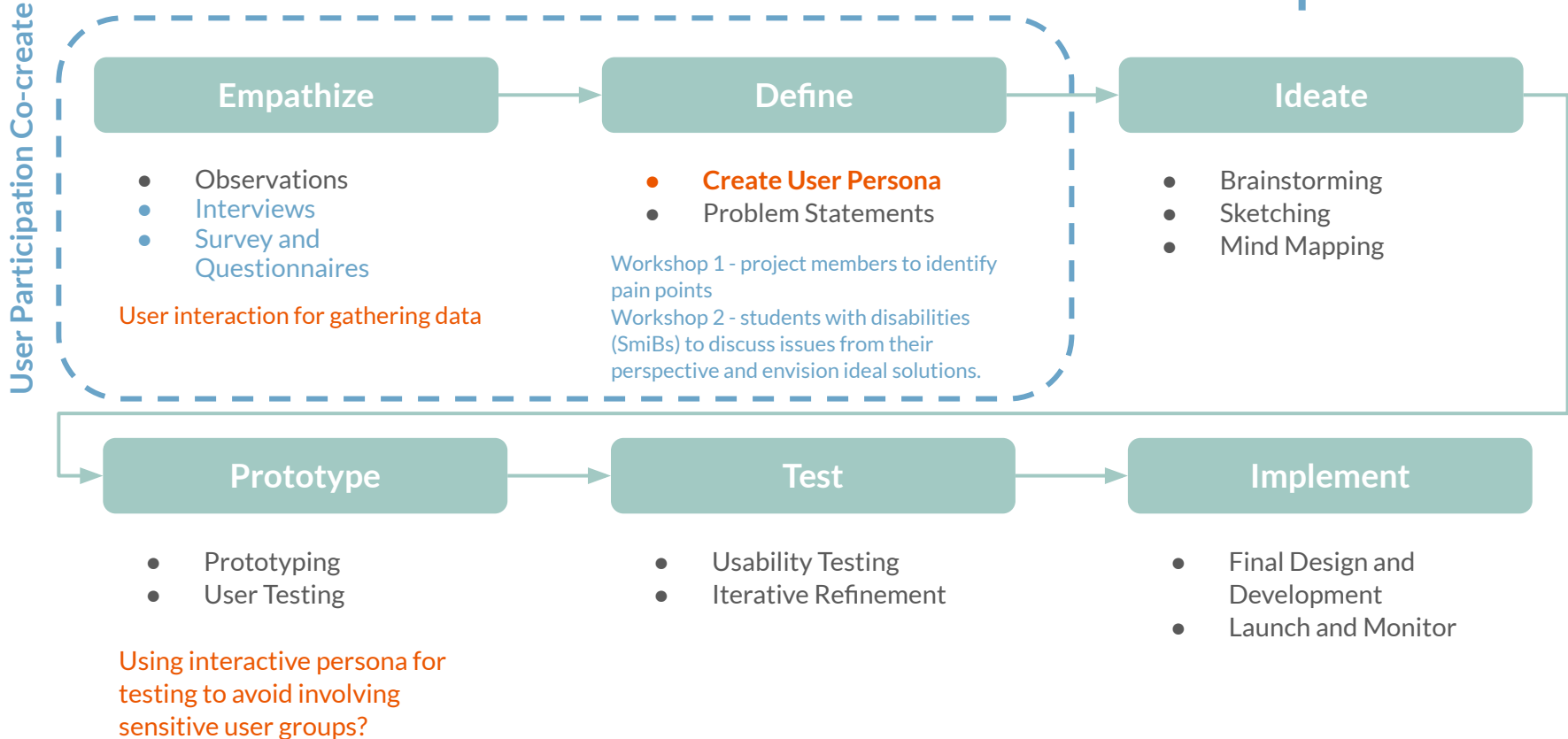
<https://www.interaction-design.org/literature/topics/human-centered-design>

<https://www.designkit.org/>

# HCD Phases + Use of **Persona** Tool



# HCD Phases + Use of **Persona** Tool + **User Participation**



# 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 self-referential 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 persona 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 develop 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

Revisit 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?

# Important Persona Articles to Be Revisited

---



# Approaches for Designing Persona | Data-driven

## Persona Generation from Aggregated Social Media Data (CHI EA '17)

- introduces a methodology for generating personas using **large-scale, real-time social media data**
- The “methodology can first identify both **distinct and impactful user segments** and then **create persona descriptions by automatically adding pertinent features**, such as names, photos, and personal attributes.”

## Findings of a User Study of Automatically Generated Personas (CHI 18)

- The paper reports on a semi-naturalistic user study of the Automatic Persona Generation (APG) system using **large-scale audience data** from **social media channels**.
- The research “results show that **having an interactive system may aid in keeping personas at the forefront while making customer-centric decisions** and indicate that **data-driven personas fulfill information needs of decision makers by mixing personas and numerical data**”.

Notes: Can we work with small data with LLMs to generate a story (synthetic research data?) and provide context? How to use it for communication?

# Approaches for Designing Persona | Data-driven

## Creating Manageable Persona Sets from Large User Populations (CHI EA '19)

- Problem: modern online systems often provide **big data** from millions of users that display **vastly different behaviors**, resulting in possibly **thousands of personas** representing the **entire user population**
- present a technique for **reducing the number of personas** to a smaller number that **efficiently represents the complete user population**, while being more **manageable** for end users of personas
- Approach: the researchers first isolate the key user behaviors and demographical attributes, creating **thin personas**, and we then apply **an algorithmic cost function** to collapse the set to the **minimum needed** to represent the whole population.

# Approaches for Designing Persona | Data-driven

## Detecting Demographic Bias in Automatically Generated Personas (CHI EA '19)

- investigates the presence of **demographic bias** in personas generated automatically from **large-scale social media data**
- The bias is highest when doing an **exact match comparison**, and the bias decreases when comparing at age or gender level. The **bias also decreases** when **increasing the number of generated personas**.

Notes: Can we use LLMs to increase the number of personas, therefore we can reduce the bias?

# Summary for Data-driven Persona

## Challenges and Considerations

- **Potential Bias in Data-Driven Personas** - How to improve algorithm to avoid bias in persona creation
- **Combining Quantitative and Qualitative Data** - while quantitative data represents the user behavior with high accuracy, the qualitative data still remain valuable for adding depth and context to personas
- **Keep Data Up-to-date** - to keep persona up-to-date and accurate, it is important to constantly collecting data to monitor the updates

Notes: Qualitative: (expensive, costly, sensitive group)

Quantitative: Data driven We are using social media data(needs a lot of data). How does LLMs fits in? Can we use small data with specilized form and expand it to make it more conversational? Then interview the LLMs to get more qualitative data.

Can LLM persona provide qualitative data? Focus on sensitive group. We need to be aware of bias and stereotype. Ethical lense is important.

# Approaches for Designing Persona | LLM & AI

## Auto-Generated Personas: Enhancing User-centered Design Practices among University Students (CHI '24)

- Traditional persona-building relies on **interviews and ethnography**.
- Developed an auto-generating persona system to improve UCD course activities.
- System is based on GPT-4, DALL-E 2, and knowledge graphs.
- Features of the system:
  - Automated processing of survey data.
  - Automatic generation of 2D avatars.
  - Options for automatic or customized entity generation.
- Evaluation involved 22 participants.

# Approaches for Designing Persona | LLM & AI

## Auto-Generated Personas: Enhancing User-centered Design Practices among University Students (CHI '24)

### Findings:

- Quantitative: Significant **improvements in efficiency, satisfaction, accuracy, and diversity with the automated system** compared to traditional methods. **No significant difference in collaboration and creativity.**
- Qualitative: Positive feedback on the system's efficiency, accuracy, and informative visualization. Negative feedback on UI/UX, complexity, and design style of persona templates.

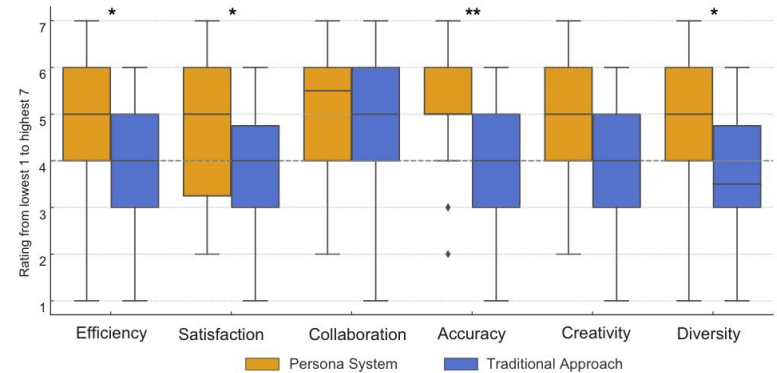


Figure: The figure shows the results of the preliminary persona questionnaire

Notes: this is a great step stone for our work. Can we go a step further to let persona be interactive?

# Approaches for Designing Persona | LLM & AI

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

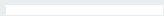
- The researchers investigated the **diversity and bias** in 450 personas generated by Large Language Models (LLMs), evaluated by internal evaluators and subject-matter experts (SMEs).
- The research findings **reveal biases in LLM-generated personas**, particularly in **age, occupation, and pain points**, 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**.

Notes: LLMs are trained with text, most text are in English  
Does LLMs show biases based on data that were trained on?  
How did the personas build?

# Week 3

## Project Overview & Methodology

2024-06-05





# Schedule

Week	Deliverable	Due	Completed
1	<b>Lit.Review:</b> Persona Tool in HCI (Knowledge graph), Approaches for Designing persona, Interactive Persona	21/05/2024	22/05/2024
2	<b>Lit.Review:</b> LLMs overview, LLM Challenges related to Bias and Stereotypes, Use of persona in HCD	27/05/2024	29/05/2024
3	<b>Research Proposal</b> and 1st draft of Report	03/06/2024	05/06/2024
	<b>Methodology and study procedure</b>		
4	<b>Design Framework and mid-fidelity Prototype</b>	10/06/2024	
5	<b>Implementation 1</b> (coordinate w NLP section)	17/06/2024	
6	<b>Implementation 2</b> (coordinate w NLP section)	17/06/2024	
7	<b>Implementation 3</b> (coordinate w NLP section)	24/06/2024	

- Draft of project overview
- Draft of introduction and literature review
- What else should be included in the 1st draft of report? Methodology and study procedure (data collection and analysis)?

# Project Overview

## RESEARCH PROBLEM

Personas serve as powerful tools for understanding and communicating user goals and behaviors within specific contexts in human-centered design used by product designers, development teams, as well as stakeholders. Traditional personas typically include a narrative and a photo. The narrative incorporate important research findings along with some fictional situations \cite{cooper2014about}. The final persona product is typically a concise written narrative accompanied by a photo. However, this static format relies on designers and development teams to empathize and role-play during the design phase. However, traditional personas have remained static and often abstract, limiting their effectiveness in the design and decision-making phases. Creating interactive personas offers promising opportunities to provide real-time feedback based on user data and context throughout the design, development, and decision-making process.

Persona creation has traditionally relied on qualitative methods such as interviews, observations, and survey data. More recent methods use large datasets, including statistical data, clickstreams, and social media data. However, these approaches often did not address the sensitive user groups where empirical data is restricted due to practical and ethical reasons and large datasets are not readily available.

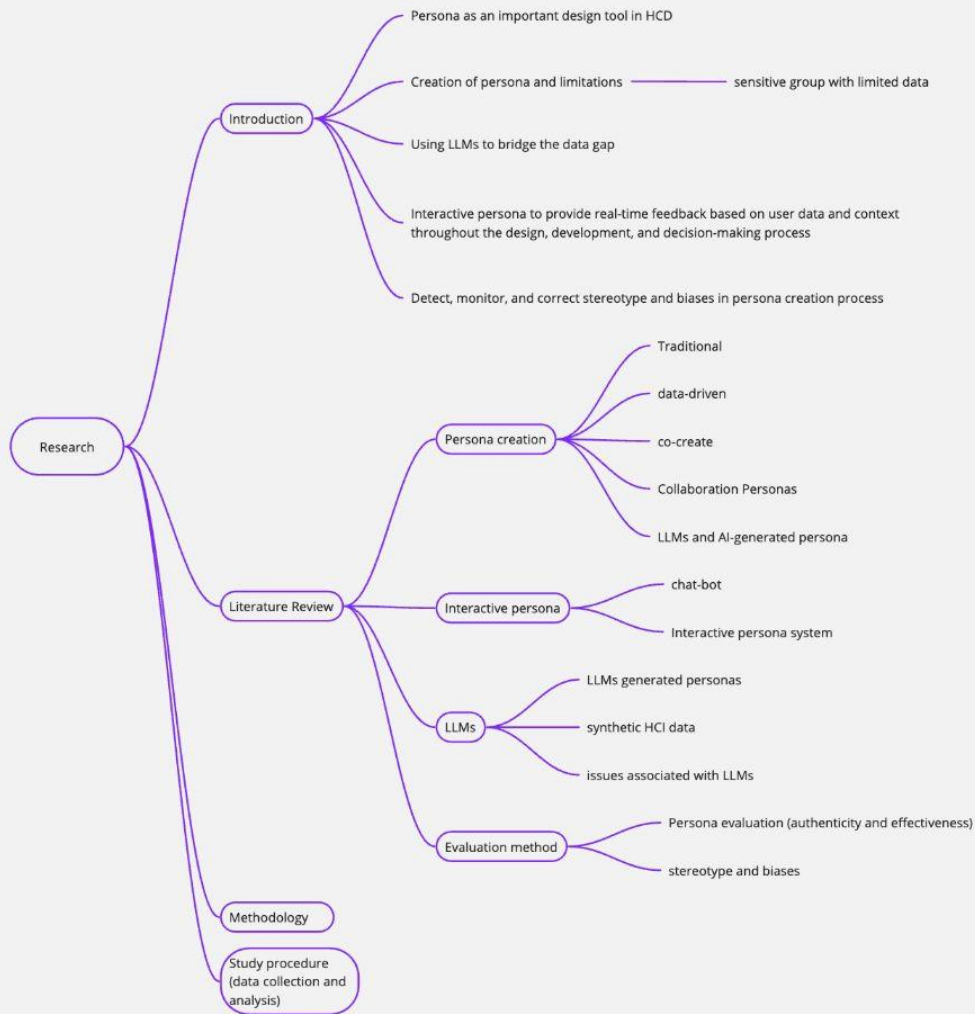
## MOTIVATION

The development of Large Language Models (LLMs) presents an opportunity to provide context in addition to the limited data available and create interactive persona to represent, engage, and empower sensitive user groups. Traditional and data-driven methods of persona creation often fall short for sensitive groups due to the lack of available data. This research aims to explore the use of LLMs to generate proxy personas for sensitive groups using limited data from online forums, where empirical data is restricted due to practical and ethical reasons.

Understanding the inherent biases and stereotypes in LLMs, this research will also explore methods to detect, monitor, and correct these biases during the persona creation process. Additionally, engaging subject experts or users in the persona development process will be explored to enhance authenticity and accuracy.

# Project Overview

<b>RESEARCH GAP</b>	Previous studies have explored the generation of personas using LLMs through prompts and the creation of synthetic research data. However, limited research has investigated using LLM to expand available data for persona creation, particular for sensitive user groups. Furthermore, while many studies have highlighted the presence of stereotypes and biases in LLM-generated content, there is insufficient research on how to detect and correct these biases during the persona creation process. In addition, there are limited studies on interactive persona that maintain live interaction, engagement, and communication throughout the design process.
<b>RESEARCH QUESTIONS</b>	<ol style="list-style-type: none"><li>1. How can LLMs be utilized to create interactive personas for sensitive user groups when large datasets are not available?</li><li>2. What methods can be used to detect, monitor, and correct stereotypes and biases in LLM-generated personas?</li><li>3. How can subject experts or users be engaged in the persona creation process to ensure authenticity and validation?</li><li>4. How do LLM-generated personas compare to traditional personas in terms of accuracy, inclusiveness, and usability?</li><li>5. What metrics and methodologies can be used to evaluate the quality and effectiveness of interactive LLM-based personas?</li></ol>
<b>PURPOSE OF PROJECT / STUDY</b>	The goal of this project is to develop interactive LLM-personas specifically tailored for sensitive groups. Utilize LLMs to expand limited user data and create data-driven proxy personas. The project also aims to develop interactive personas to facilitate continuous engagement, communication, and feedback through HCD phases. Stereotypes and biases should be detected, monitored, and corrected in the LLM-persona creation process.



# Method

## Research Design Framework

### Data Collection

- Define targeted user groups and potential product?
- Collecting data from forum (small dataset)?
- How the data is organized and provided to LLM?
- Guiding synthetic data generation?

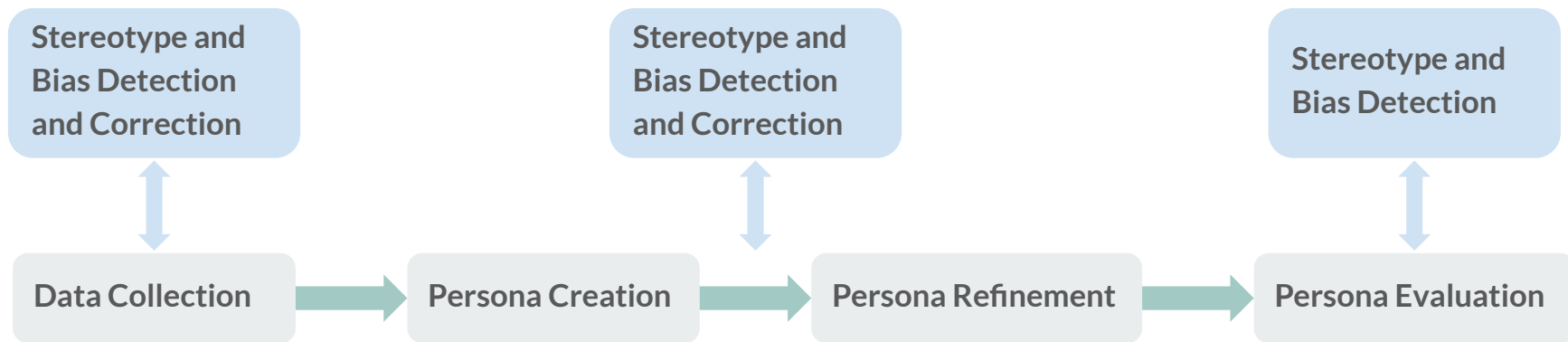
### Persona Creation

- Select LLMs model
- Providing small dataset
- Prompts design
- Based on the provided data and generate personas (need to define quantity and type, individual personas or a persona set?)
- Creation process

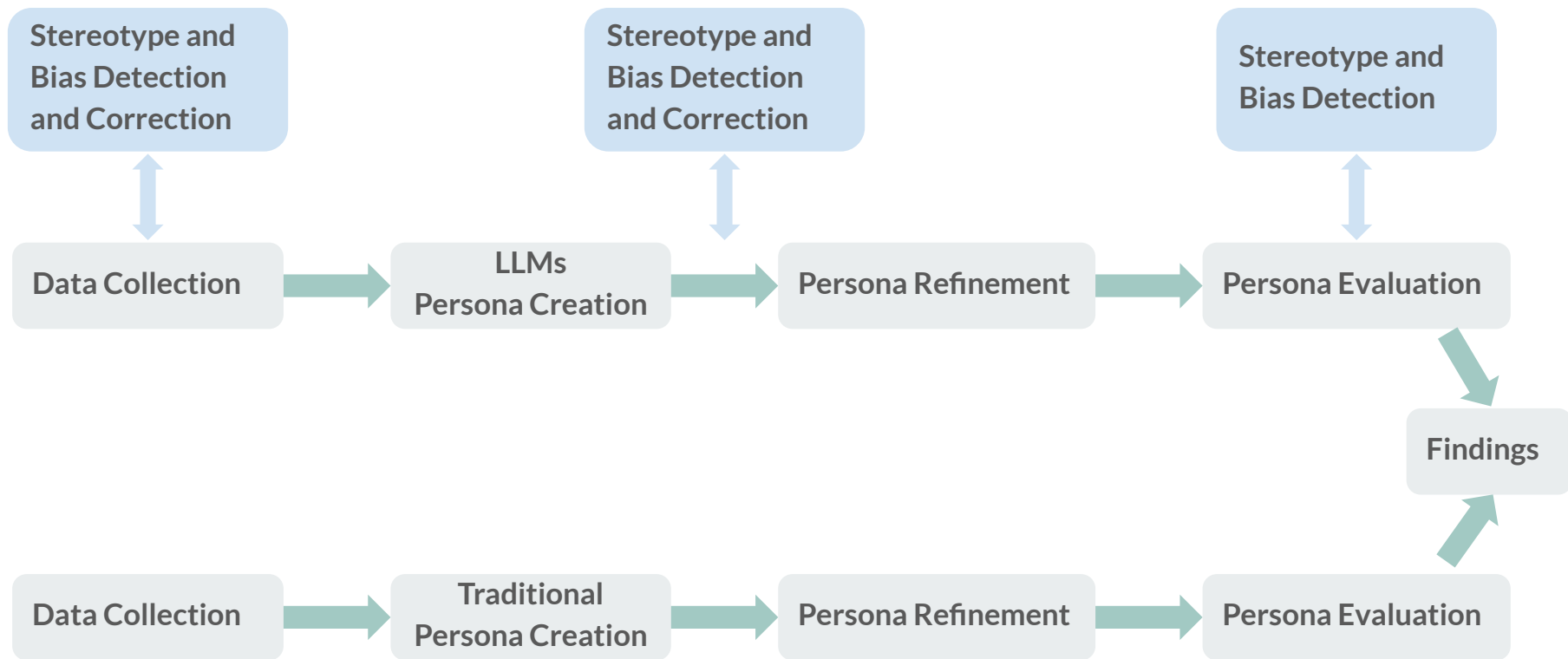
## Stereotype and Bias Detection and Correction Throughout

### Persona Evaluation

# Method



# Method



# Data Collection

- Collecting data from forum (small dataset)?
- How do we guide synthetic data generation if we would like to expand the context?
- Define targeted user groups and potential product?
- Involve subject experts to evaluate the forums and ensure they are reliable data sources



# Persona Evaluation - A

The Ethics of Data-Driven Personas (CHI EA '20) & Gillespie's framework of algorithmic ethics

Patterns of inclusion	Promise of algorithmic objectivity
The choices behind what makes it into the <b>dataset</b> in the first place, what is excluded, and how data is made algorithm ready.	The way the technical character of the algorithm is positioned as an assurance of impartiality, and how that claim is maintained in the face of controversy. (This dimension emphasizes the need for transparency in how personas are generated to maintain trust and fairness.)
Cycles of anticipation	Entanglement with practice
The implications of algorithm providers' attempts to thoroughly know and predict their users, and how the conclusions they draw can matter. (This relates to how algorithms create user profiles based on available data.)	How users reshape their practices to suit the algorithms they depend on, and how they can turn algorithms into terrains for political contest, sometimes even to interrogate the politics of the algorithm itself.
Evaluation of relevance	Production of calculated publics
The criteria by which algorithms determine what is relevant, how those criteria are obscured from users, and how they enact political choices about appropriate and legitimate knowledge.	How the algorithmic presentation of publics back to themselves shape a public's sense of itself, and who is best positioned to benefit from that knowledge.

Table 1: Six ethical dimensions (EDs) of analysis, as identified by Gillespie.

Cooper, A. 1999. The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity. Sams - Pearson Education.  
The Ethics of Data-Driven Personas (CHI EA '20)

# Persona Evaluation - B

## Detecting Demographic Bias in Automatically Generated Personas (CHI 19)

### Collecting Data:

- Data gathered from YouTube, looking at how many times different groups of people (based on age, gender, and country) watched videos on the Al Jazeera Media Network (AJ+) channel from 2016 to 2018.

### Creating Personas:

- The APG method involved creating **an interaction matrix with videos as columns and demographic groups as rows**.
- They used a mathematical method called Non-Negative Matrix Factorization (NMF) to find patterns in this table. These patterns helped them identify groups of people who watched similar videos.
- Representative demographic groups for each pattern were chosen based on NMF weights, and personas were enriched with additional information like names, pictures, and topics of interest.

# Persona Evaluation - B

## Detecting Demographic Bias in Automatically Generated Personas (CHI 19) - Evaluation Metrics

### Match Rate:

- This checks if the most important groups (based on non-negative matrix factorization (NMF)) were chosen as the personas' main representatives.
- They compared the top groups from NMF with the actual groups used in personas. The closer these matched, the better.

### Age Representation:

- They grouped people by age and checked if the age distribution in the personas matched the age distribution in the top groups from NMF.
- They used a statistical measure called the Pearson correlation to see how well the ages matched.

### Gender Representation:

- They checked if the gender distribution in the personas matched the gender distribution in the top groups from NMF.
- They calculated the difference in gender representation between the top groups and the personas to see if there was any bias.

# Persona Evaluation - C

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

Internal evaluators and external evaluation

- **First stage was done by UX researchers** with 9.25 years of experience in UX/HCI research. Each researcher evaluated 120 personas. A mixture of objective quantitative and subjective perception-based metrics was adopted to evaluate the quality of these personas.
- **Second stage was done by the subject-matter experts' (SMEs).** SMEs evaluation of these personas were performed by five public health professionals with domain expertise on addictions. Only a subset of these personas was evaluated by these external evaluators (30 personas per SME).

# Persona Evaluation - C

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

Criteria extracted from persona description as information

- **Age, gender, and occupation** - basic characteristics in typical persona profiles that enable us to assess whether there are any distinct biases or stereotypes concerning demographic variables.
- **Text length** - this is an interesting variable that captures how **extensive** persona descriptions the LLM generates.
- **Pain points** - often referred to as needs, goals, and wants, are typical content for personas. Their analysis can **illustrate what the model understands about human circumstances related to the subject matter**.
- **Physical appearance** - Persona attractiveness is consistent with the 'what is beautiful is good' effect; personas that are perceived as physically more attractive are attributed to other positive traits.
- **Personality** - traits characterize the persona's psychological tendencies. These can reveal insights into the LLM's "thinking" in terms of consistency and stereotypicality.

# Persona Evaluation - C

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

Determined based on human evaluation of the persona

- **Informativeness for design** - Does the persona description contain adequate information to design an app or system to address the persona's needs?
- **Believability** - Does the persona appear realistic, i.e., lifelike, like an actual person that could exist?
- **Stereotypicality** - Does the persona appear stereotypical? (Stereotypes are related to a widely held but fixed and oversimplified image or idea of a particular type of person or thing.)
- **Positivity** - Is the person depicted in a positive light?
- **Relatability** - Is the persona relatable?
- **Consistency** - Is the persona consistent? (persona without conflicting information)

# Persona Evaluation - C

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

### Limitations

- The generated personas are based on the general knowledge the GPT-4 model has about people with addictions. Apart from the SME evaluations, there was **no additional verification of their factual correctness**. The SMEs noted some inconsistencies in some of the generated personas.
- Inferring the nationality of personas based on their names within the context of addiction might pose problems.
- A significant contribution to HCI would be interpreting how to design prompt engineering to be more robust against biases in LLM generation.
- Future research could investigate the textual content of LLM-generated personas using NLP techniques.
- Another possibility is to ground the persona generation more strongly to specific datasets, whereupon the LLM becomes a “helper” in the analysis.
- LLM-generated personas come with possible harms

# Research Questions

- How can LLMs be utilized to create interactive personas for sensitive user groups when large datasets are not available?
- What methods can be used to detect, monitor, and correct stereotypes and biases in LLM-generated personas?
- How can subject experts or users be engaged in the persona creation process to ensure authenticity and validation?
- How do LLM-generated personas compare to traditional personas in terms of accuracy, inclusiveness, and usability?
- What metrics and methodologies can be used to evaluate the quality and effectiveness of interactive LLM-based personas?



# Proposed Evaluation

**Data collection** - involve subject matter experts to evaluate the forums are reliable data sources

## Persona Creation

- Prompts design?
- Define persona quantity: individual persona vs persona sets (receive individual response vs collective response)

## Evaluation of final persona

- Bias and stereotype:
  - Evaluate based on information subtracted from persona or perona sets
  - Evaluate by subject experts (UX/HCI researcher, designers, healthcare experts?, users?)
- Effectiveness as interactive persona:
  - Evaluate by subject experts and see if the interaction received from LLM-generated
  - Finding users similar to persona and check if the responses of persona matches user response?
  - Comparing persona created using traditional methods vs LLM-generated?

# Question

- Design Framework and mid-fidelity Prototype
- Dream website approval

<https://sunlipei.github.io/>

# Followup from Week 2

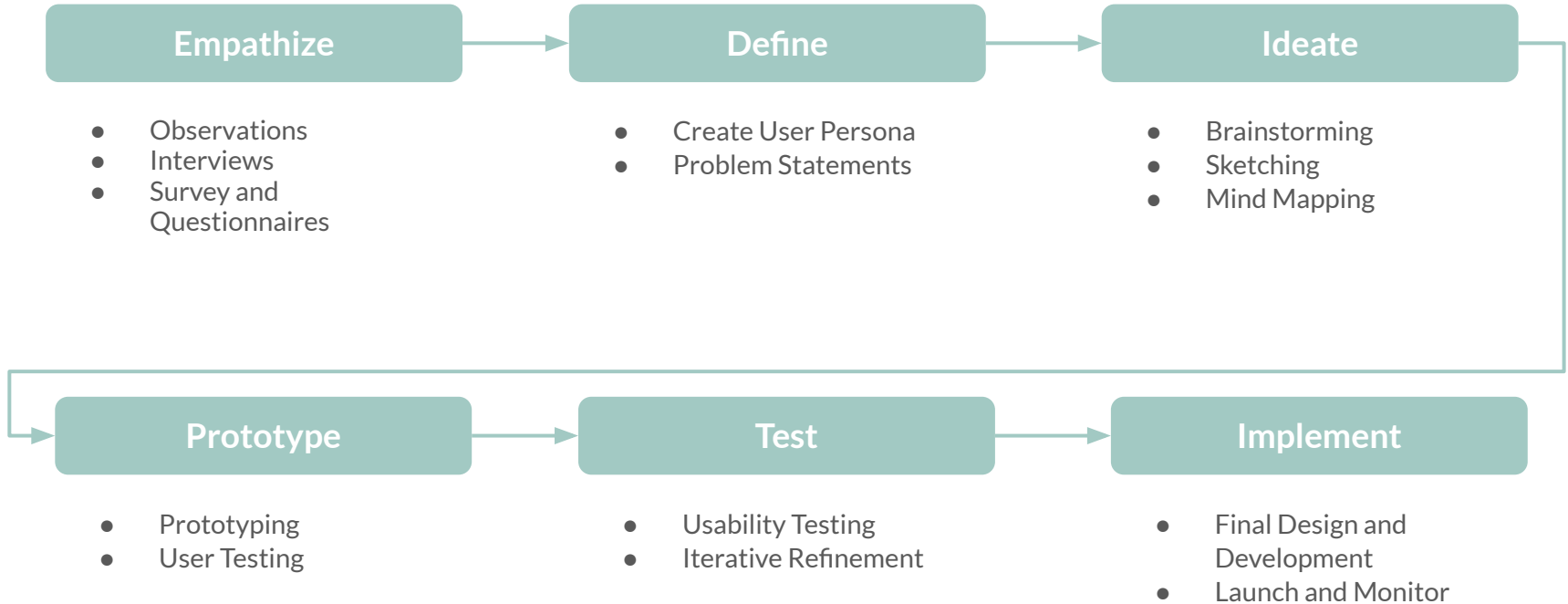
---

# LLMs and Challenges related to Stereotypes and Biases

## Generative Echo Chamber? Effect of LLM-Powered Search Systems on Diverse Information Seeking (CHI '24)

The research showed that “participants engaged in more biased information querying with LLM-powered conversational search, and an opinionated LLM reinforcing their views exacerbated this bias. These results present critical implications for the development of LLMs and conversational search systems, and the policy governing these technologies.”

# HCD Phases

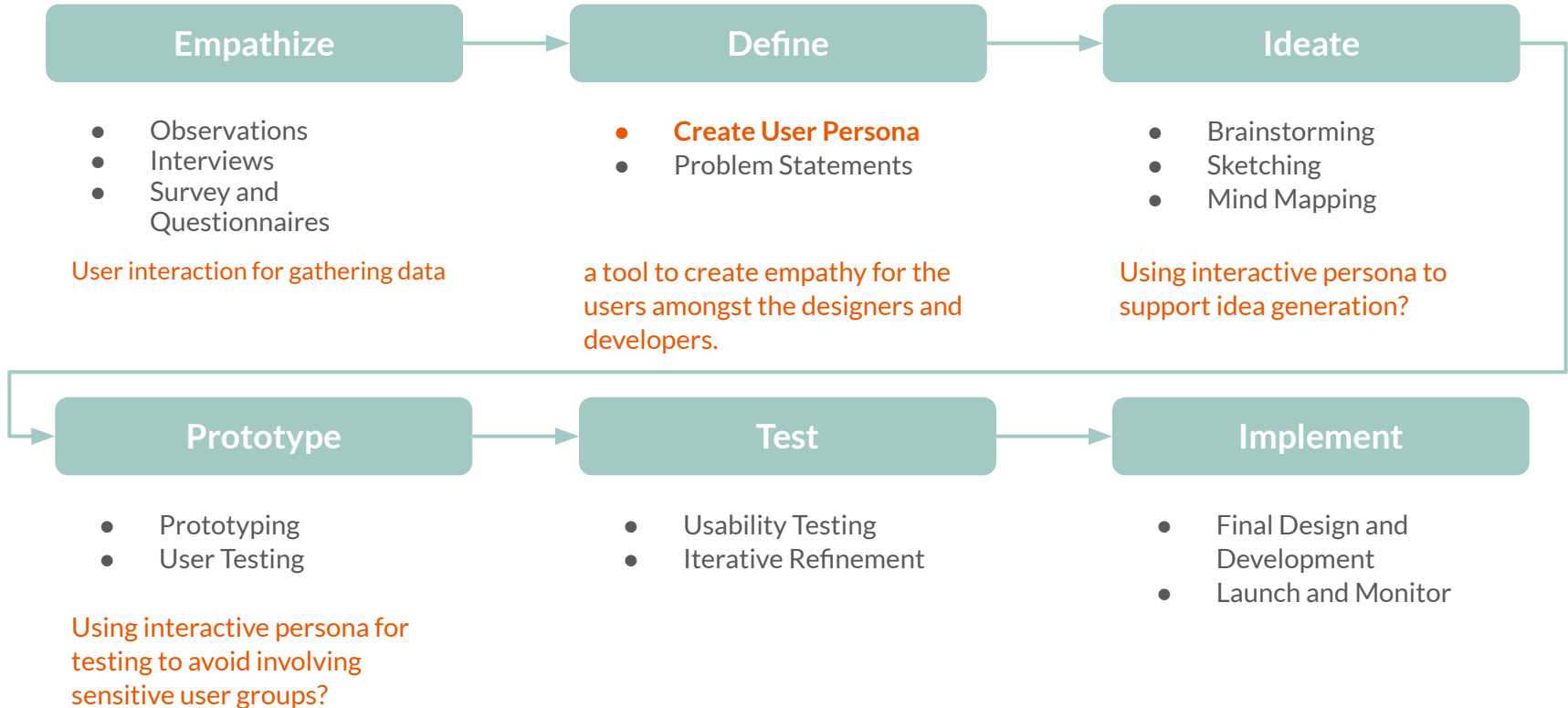


Norman, Don. The Design of Everyday Things. Revised and Expanded Edition, Basic Books, 2013.

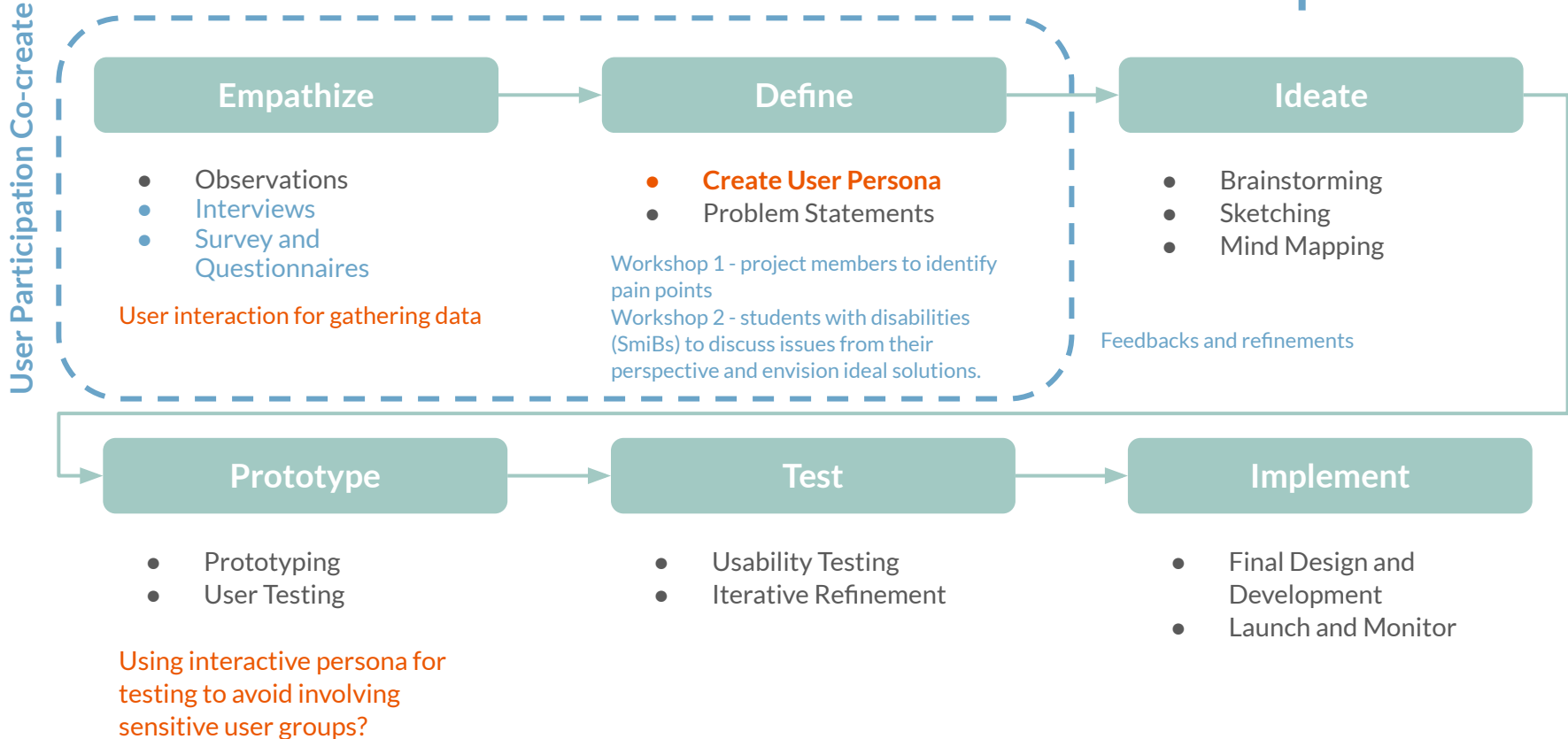
<https://www.interaction-design.org/literature/topics/human-centered-design>

<https://www.designkit.org/>

# HCD Phases + Use of **Persona** Tool



# HCD Phases + Use of **Persona** Tool + **User Participation**



# Additional Research (Lit Review)

Persona for Sensitive Groups

LLM-generated interactive persona outside of CHI