Zeya Chen | Working in Progress, 2024 Spring
PhD Program, Institute of Design, Illinois Tech

Designing for AI's "Last Mile":
A Behavioral Approach for Hybrid Intelligent Interventions Design in AI-Empowered Socio-Technical Systems

0. Brief Introduction + Research Questions

Recent advancements in artificial intelligence (AI) have enabled remarkable capabilities, with AI systems now driving cars, translating languages, and discovering new drugs. Despite the widespread presence of AI in everyday products and services, research indicates that over 85% of AI projects fail to create value for users or deliver viable services (Ermakova et al., 2021; Joshi et al., 2021; Weiner, 2020). Many of these failures stem from a lack of human-centered design, as design research is often not involved until after the decision of what to innovate has already been made (Kross and Guo, 2021; Nahar et al., 2022; Piorkowski et al. 2021). Practitioners repeatedly experience AI project failures due to addressing the wrong problems – developing solutions that do not meet real user needs (Yildirim et al., 2023). One article aptly labeled this issue as the "Last Mile Problem" of AI (Berinato, 2019), which describes the challenge of producing data-evidenced insights but failing to communicate them effectively, leading to wasted or misapplied information (Logg, 2019). This problem can be caused by not only technology limitations, but also biases and heuristic of AI practitioners and users.

AI failures can trace back to problem selection and formulation stages (Yildirim et al., 2023). Data science and developer teams fail to systematically define needs from domain experts and product managers, envisioning AI systems users do not want (Kross and Guo, 2021; Lam et al., 2023; Yang et al., 2019; Yildirim et al., 2023). Conversely, product role practitioners (e.g., designers, product managers) lack understanding of AI's reasonable capabilities, conceptualizing unbuilt AI solutions (Dove et al., 2017; Yang et al., 2019; Yang et al., 2020). Teams overlook low-hanging fruit where simple AI could improve user experience (UX) (Yang et al., 2020), and engaging domain stakeholders/users early in AI development remains challenging (Kross and Guo, 2021). These issues may be largely due to designers lacking good mental models for what AI can and should do, which presents a significant challenge.

In recent years, resources such as human-AI guidelines and design patterns have become available (Google PAIR, 2019; Apple, 2019). However, practitioners report that these guidelines primarily assist with prototyping and iteration (Buxton, 2020) – “making the thing right” . What designers and product managers lack most are resources to aid in problem framing and uncovering leverage point in complex AI service systems(Yildirim et al., 2023) – “making the right thing.” To address these challenges, this research focuses on the following research questions:

<u>Main research questions:</u>

How can designers leverage hybrid intelligence to identify and address the "last mile problem" of AI-empowered products/services by designing and implementing effective and ethical behavioral interventions within complex socio-technical systems?

This main question can be broken down into two interconnected parts *(Fig.1)*:

- ***Design <u>for</u> AI***: The primary part focuses on problem framing, designing, and implementing behavioral interventions to address the "last mile problems" that arise between AI systems, AI practitioners, and end-users within complex socio-technical systems.
- ***Design <u>with</u> AI***: The supplementary part studys how hybrid intelligence (collaboration of human intelligence and machine intelligence) can be leveraged during the design process to enhance effectiveness and ethicality of design solutions.

The dual-focus feature of this research is similar to the dual force nature of behavioral design – when behavioural design tackles complex challenges, it has a dual focus on: 1) achieving behavioural effect(s) (e.g., reducing users' over-trust behaviors on algorithms); facilitated by 2) designing and implementing intervention(s) (e.g., information campaigns for users or speed bumps for developers) (Khadilkar & Cash, 2020). Balancing behavioral principles (design) and behavioral changing techniques (practices) is crucial for successful interventions (Nielsen et al., 2024). In this research, I propose to balance the dual focus by centering the problems framing, design and implementation of behavioral interventions for AI's last-mile challenges **(design for AI) as the primary, major forces**. To enhance the effectiveness and ensure the ethical considerations of these interventions, I will investigate and employ strategies of hybrid intelligence between AI and designers **(design with AI) as a secondary, supplementary focus**. By evaluating intervention results and reflecting on the design processes of addressing AI's last-mile problems, I aim to contribute to the growing body of design knowledge *for* AI.

Based on the above logic, I propose the following sub-research questions to lead my research:

1. **Exploring Designer-AI Collaboration:**

- *How do designers employ AI tools in design projects, and how does this change traditional design processes?*

This sub-question begins by investigating the "design with AI" aspect, focusing on how designers currently work and should work with AI, and aims to establish a foundational understanding of the relationship between designers and AI in the new era. The goal is to gather insights and strategies for hybrid intelligence working mode, which can later inspire the design process, methods, or tools for solving "the last-mile problem."

2. **Identifying the Problem and Designing Interventions:**

- *How could behavioral science principles and theories aid designers in identifying human/machine behavioral patterns and uncovering "the last mile problem" between AI practitioners, AI users, and AI products?*
- *How to design and implement ethical and effective human-AI behavioral interventions within complex socio-technical systems for AI products/services?*

These sub-questions, which span the dual focus, form the central and most critical part of my research. The first sub-question explores how behavioral knowledge can help in the human-AI collaborative design process to understand AI's last mile challenges from

multi-stakeholder systematic perspective (bridging 'design with AI' and 'design for AI'). The second question focuses on introducing changes by designing effective behavioral interventions for AI-enabled innovations ('design for AI'). Together, these sub-questions establish a solid behavioral lens and set a clear direction for the next steps in the following sub-research questions, Q3 and Q4.

3. **Evolving Design Methods and Tools:**
- *What adaptations or evolutions are necessary for traditional design methods to support design for AI-empowered products/services? What new design techniques and tools are missing but needed?*

This sub-question builds upon the previous sub-questions by investigating the necessary changes in design methods and tools and suggesting the integration of behavioral perspectives and big data power to address AI's "last-mile challenges." It serves as a bridge between the two parts of the main question. The methods and tools developed in response to this sub-question could primarily support "design for AI" by updating design knowledges for tackling wicked AI problems. Additionally, these methods and tools might also have the potential to assist "design with AI" by equipping designers with new techniques to easily collaborate with AI in the design process.

4. **Envisioning Role and Responsibility of Designers:**
- *What emergent roles, responsibilities, and ethical considerations do designers face in the context of emerging hybrid intelligence systems?*

This final sub-question circles back to the beginning from a higher-level perspective. It summarizes and synthesizes the findings from the previous sub-questions, looking forwards to the emergent roles, responsibilities, and ethical considerations of designers in hybrid intelligence futures.

By connecting the sub-questions to specific aspects of the main research question, I aim to demonstrate a clear logical flow that addresses the complex challenge of employing hybrid intelligence for designing behavioral interventions to tackle the "last-mile problem" of AI-empowered innovations *(Fig.1)*. Although each sub-question may appear to be a broad research topic on its own, the following figure illustrates how these questions are interconnected and focused on the specific research area at the intersection of designing with/for AI. Through this research, I hope to contribute to the growing body of knowledge at the intersection of design, behavioral science, and human-computer interaction, providing valuable insights and interactive frameworks to guide designers in tackling wicked problems between AI products/services and users.

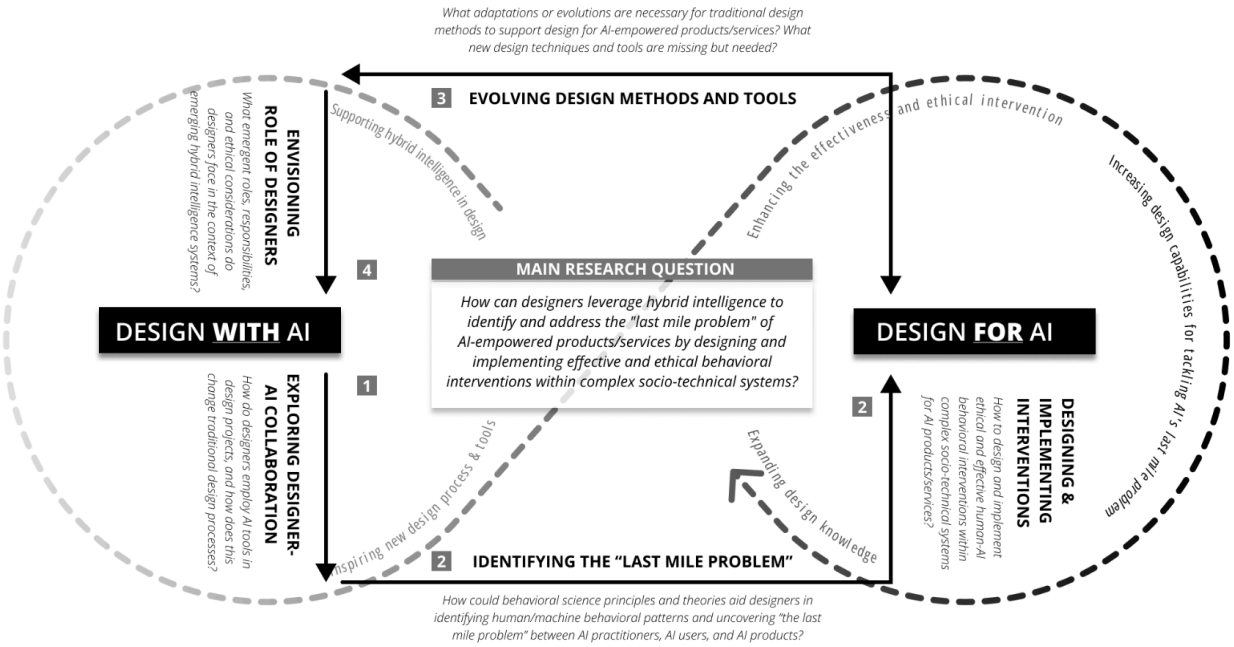


Figure 1. Relationship between main research question and sub-Qs, and the dual-focus of this research with major focus on designing for AI.

1. Philosophical Context of Design Research

1.1 New Paradigm of Hybrid Intelligence - Why designing both with/for AI?

Hybrid Intelligence serves as the foundational philosophical stone supporting my research, addressing the fundamental question of why this research involves both designing *with* AI and *for* AI.

As AI technologies transition from labs to real-world applications, their benefits are often accompanied by negative impacts on individuals and societies, stemming from both the technologies' limitations and how they are deployed. While current approaches mainly include technical solutions improving AI models, and governance solutions developing ethical regulations and policies (Guszcza et al., 2022), Guszcza et al. (2022) propose a complementary Hybrid Intelligence approach, integrating machine and human intelligence to overcome technical limitations. This paradigm builds intelligent systems augmenting human capabilities, leveraging our strengths while compensating for weaknesses, considering ethical and societal factors (Akata et al., 2022).

The hybrid approach informs my research perspectives by emphasizing the importance of ensuring human rights, needs, and values are integral to the design process, and highlighting the need for meaningful humanity-centered design and preventing centralized power over technology development (Guszcza & Schwartz, 2020). Given the inevitable evolution of AI in our daily lives, designers should actively seek ways to develop human-AI hybrid intelligence for responsible innovations rather than attempting to avoid AI technology altogether.

However, Hybrid Intelligence lacks systematic development as an applied, translational field (Guszcza et al., 2022). Such a field would require a broader development approach than traditional machine learning engineering. As Guszcza et al. (2022) suggests, it should be ***design-focused***, enabling crucial human needs and values to be addressed early in the design process rather than reactively; it should involve ***multi-stakeholder collaboration***, reflecting diverse perspectives and local contexts in the design process; additionally, it should be ***interdisciplinary***, integrating concepts and methods from the **behavioral sciences**, **human-computer interaction (HCI)**, **human-centered design**, and **applied ethics of data and information sciences**. Therefore, by anchoring my research in the philosophical framework of Hybrid Intelligence and focusing on the AI-empowered products/services design process, guided by the theories and concepts from the previously mentioned four disciplines, I aim to contribute to the systematic development of this applied, translational field.

1.2 ANT Theory & Autonomous Behavior - *Why a behavioral perspective? Does AI have behavior?* [HCI/Behavioral science]

To illustrate the importance and necessity of behavioral knowledge, we must first ask a fundamental question: Does AI have behavior?

The idea that AI systems can exhibit behaviors traditionally associated with humans and AI anthropomorphism has sparked philosophical debate. For example, Actor-Network Theory (ANT) (Latour, 2005) suggests that both human subjects and non-human objects, such as physical designs, exercise agency that affects actions. Behavioral science scholars suggest AI engages in an intelligent process similar to humans – perceiving stimuli, reasoning about potential actions, and selecting actions expected to achieve objectives (Mills and Sætra, 2022). However, distinctions between AI and human behavior have been raised, such as AI's reliance on intensive training data and reinforcement learning algorithms, contrasting with human autonomous learning and self-reflection (de Vos 2020; Watson 2019). Researchers have critically examined the implications of granting AI systems a form of autonomous "motive power" (Marx 2013 [1867]) and decision-making ability, which has conventionally been a distinguishing feature of human moral consideration (Gunkel 2020; Turkle 2004 [1984]).

These concepts inform my research of hybrid intelligence for design in below several ways:

- **Non-human agency**: ANT posits that non-human actors like technologies and artifacts can exert force and influence within heterogeneous networks. This perspective encourages me to consider AI systems as active participants in the design process, shaping and being shaped by interactions with human and non-human actors.
- **Power dynamics**: Zuboff's (2019) examination of power dynamics when "inorganic entities" possess autonomous decision-making capacity resonates with ANT's focus on how non-human actors can wield power and reconfigure networks. This informs my research by highlighting the need to critically examine the power dynamics between AI systems, AI practitioners, and AI users during both the development and usage processes.
- **Behavioral influences**: Admitting the behavioral influences of AI on AI practitioners and users during both development and usage processes aligns with ANT's flattened ontology challenging human agency centrality. This perspective encourages me to consider the dynamic collaborative relationship between intelligent non-human agents (AI) and human agents (AI practitioners, users, and other potential stakeholders) from systematic perspective in the research.

By incorporating the philosophical contexts of ANT Theory and Autonomous Behavior Theory at the intersection of HCI and Behavioral Science, my research aims to provide a more comprehensive understanding of the complex interactions between AI systems, AI practitioners and users.

1.3 Feminist AI - *What are/should be the roles of AI?* [HCI/Ethic]

Recognizing AI's behavioral influence from a larger system perspective expands the discourse on the human-AI relationship. To dive deeper into understanding the existing relationship between human-AI interaction and the potential role of AI in hybrid intelligence futures, I refer to the field of applied ethics study from computational and social science to investigate principles and theories to guide ethical practices in human-AI collaboration.

Feminist critiques of AI have argued for the inclusion of marginalized forms of knowledge (Harding, 2008; Liboiron, 2021) and questioned objectivity in science while critiquing hegemonic masculinity in technoculture (Forsythe, 1993; Wajcman, 1991). These critiques lead me to examine and reconsider the representation of AI technologies. Furthermore, they serve as a reminder to exercise caution when utilizing AI as design materials in collaboration, given the non-neutral features, lack of representative datasets, and untransparent algorithmic "black box" of AI systems. Drawing from Actor-Network Theory (ANT) and autonomous behavioral theory mentioned above, my research frames AI as intelligent non-human agents possessing their own biases and stereotypical behaviors.

The concept of situated knowledge (Haraway, 1988) and Alison Adam's feminist critique of AI (Adam, 1995) inform my research perspective by highlighting the importance of considering diverse and marginalized voices in the development and analysis of AI systems. This insight suggests that the discrepancy between AI's capabilities and users' needs may also be rooted in the problems faced by marginalized groups and their unheard requirements. It highlights a potential research direction involving the inclusion of diverse voices during the initial stages of AI product/service development, specifically the problem framing and ideation phases. Moreover, it underscores the necessity of considering inclusivity from the perspectives of both AI inside workers and outside end-users.

The term "feminist artificial intelligence (FAI)" encompasses diverse manifestations of feminisms, including intersectional (Combahee River Collective, 1979), Black feminist (hooks, 2015), decolonial (Lugones, 2010), and liberal aspects. The diverse roles and purposes of AI within feminist contexts inform the exploration of multiple possibilities for conceptualizing AI. Building upon Toupin's research (2023), the diagram below situates my research agenda within this context.

AI as design	This category focuses on the design aspects of AI systems, emphasizing the importance of including diverse voices (women, queer, trans, and BIPOC individuals) in the design process and considering the cultural context in which AI tools are developed.	Bardzell, 2010 Costanza-Shock, 2018 D'lgnazio and Klein, 2020 Meinders, 2017, 2020
AI as discourse	This category refers to the use of AI as a signifier or placeholder to describe critical work addressing the relationship between gender,	Adam, 1997 Noble, 2018 Benjamin, 2019

	race, class, and technology. In this role, FAI discourse has the power to create new imaginaries and engage women and girls in AI.	Avila, 2021
AI as culture	This role emphasizes the cultural factors that shape the relationship between gender and technology. AI as culture perspective argues that changing cultural norms associated with technology, such as social, physical, and cognitive biases, can lead to the development of FAI.	Wajcman, 1991 Wellner and Rithman, 2020

Drawing from the philosophical analysis of FAI theory, I aim to position my research at the intersection of "AI as design," "AI as discourse," and "AI as culture." By engaging with the concerns and principles of "AI as design" and "AI as discourse," the research seeks to incorporate inclusivity in the AI's design and development process. This ensures that not only underserved needs can be fulfilled by AI systems but also that diverse voices can be present and seen through AI products. Informed by the "AI as culture" perspective, the research aims to develop interventions in human-AI interaction by challenging and transforming the cultural norms associated with technology, such as social, physical, and cognitive biases. This approach has the potential to facilitate the development of more equitable and socially responsible AI systems, aligning with the primary focus of the research – "designing for AI."

1.4 Universalism vs Pluralism of Human-centered AI - *What design philosophies help?* [Design/Ethics]

Inspired by the FAI theory, several design ethics and philosophies — including Human-centered AI (HCAI), Inclusive Design, Design Justice, and Pluriversal Design — can inform responsible design both with and for AI.

The concept of HCAI has emerged as a response to issues such as biased datasets, discrimination, and privacy threats in AI systems. HCAI emphasizes the importance of understanding user needs and ensuring human control (Sio & Hoven, 2018) while integrating traditional HCI/design methods to create valuable, reliable, and trustworthy systems that benefit society (Shneiderman, 2020). As AI failures continue to rise globally, responsible design theories, including Inclusive Design and Design Justice, have become increasingly crucial for mitigating bias and incorporating diverse perspectives.

Inclusive Design and Design Justice share a philosophical foundation that recognizes human diversity and rejects the notion of a universal "normal standard" (Bianchin & Heylighen, 2017; Persson et al., 2015). Inclusive Design's goal of facilitating equal opportunities and societal participation for all (Bendixen & Bentzon, 2015) provides a clear direction for engaging diverse stakeholders' voices in identifying problems, including but not limited to

AI practitioners, AI users, AI regulators, and AI-impacted non-users. Design Justice's critical perspective, which acknowledges systemic power asymmetries and the potential for technology to create barriers for marginalized communities (Costanza-Chock, 2018), supports my research by providing structured principles and checklists for examining problems from a wider social-technical systematic perspective.

Pluriversal Design theory posits that there is no universally desirable design and that designs from diverse cultural contexts have unique merits and beauties in terms of their own worldviews (Escobar, 2018; Noel et al., 2023). This theory informs my research by fostering a more inclusive and context-sensitive approach to AI design, regarding each AI problem and human-machine collaboration process as a distinctive challenge, considering the unique merits and beauties of designs from various stakeholders and diverse social contexts.

Inspired by the FAI theory, these design ethics and philosophies collectively address the need for approaches that deal with different degrees of access, ability, awareness, and diverse perspectives in responsible AI design. Drawing from HCAI, Inclusive Design, Design Justice, and Pluriversal Design theories does not imply that this research will consider all of these theories simultaneously. Instead, these philosophical contexts provide a strong foundation, clear ethical direction, and flexible frameworks and approaches for the research, guiding the responsible AI design process and solutions that benefit society as a whole. Similar to but adding on to FAI theory, these design ethics not only inform the research's nuanced and context-sensitive exploration of human-machine collaboration on responsible AI design but also bring resourceful methods and tools to potentially lead ethical envision into practical design activities.

1.5 Rationalistic vs Humanistic perspectives - *What are roles of design research?*

[Behavioral Science/Design]

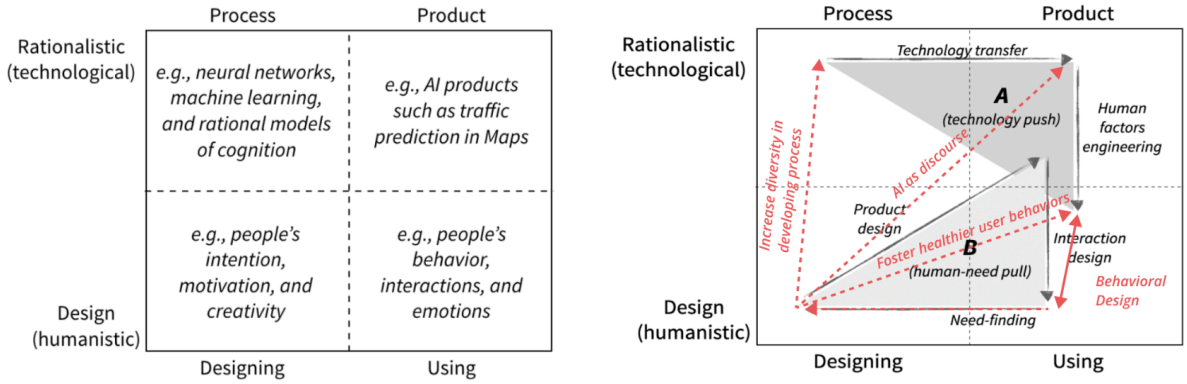
The evolution of AI systems and technologies since Alan Turing's proposal of the Turing Test in 1950 (Turing, 1950; Grudin, 2009) has been marked by two main philosophical perspectives on human-computer interaction: the "rationalistic" and the "humanistic" perspectives (Winograd, 1996; Grudin, 2009). This divide, reflecting the ongoing discourse between the cultures of science and humanities (Snow, 1993), is particularly relevant to my research, as it mirrors the tension between the rational, science-based approach of behavioral science and the more humanistic approach of design.

Behavioral science, grounded in an emphasis on cognitive processing through rigorous experimentation, aligns with the rationalistic perspective, which views AI as computer systems that imitate human abilities and people as "cognitive machines" (Winograd, 2006; Winograd & Flores, 1986). In contrast, design prioritizes lived experience and positions humans as more holistic, aligning with the humanistic perspective that sees AI as a problem-solving tool to enhance human capabilities and conditions (Winograd, 1996; Winograd & Flores, 1986).

This tension between science/technology and "softer" disciplines like design also manifests in notions of "design science" (Simon, 1969), which has influenced both AI and behavioral science. Herbert Simon, a key figure in design science, contributed ideas like "bounded rationality" to behavioral science (Simon, 1972). The tension persists in how behavioral interventions are evaluated, often through traditional scientific measures like randomized controlled trials (RCTs).

Informed by these historical perspectives around AI and Auernhammer's (2022) diagram analysis on the relationship between historical perspectives and HCD research, I find that the gap between the rationalistic and humanistic approaches is similar to the "last mile problem" of AI, where technology capability and service offerings do not fit users' needs and expectations. To bridge this gap, my research proposes the use of behavioral design, which combines the rational science-based approach of behavioral science with the more humanistic approach of design, which prioritizes lived experience and positions humans as more holistic. For example, leverage behavioral design to increase inclusivity in the initial development process of AI technologies by intervening in developers' working behavior; intervening in machine behavior and encourage more ethical outputs, enabling AI to serve as a responsible discourse for larger impact; fostering healthier and responsible behavior interacting with AI. (see red lines in Fig 2 [right]).

The design approach to filling the gap is further supported by scholars who have identified the lack of discourse between user experience (design) and machine learning (rationalistic) fields (Yang et al., 2018) and have proposed a great opportunity for collaboration between AI and Design Research to address concerns around fairness, accountability, and transparency of AI systems (Abdul et al., 2018).



- ***Figure 2 [left]***, Auernhammer (2022) Illustrates the spectrum of the rationalistic view and design perspectives. The rationalistic perspective focuses on thought and people as a formal symbolic representation and focuses on process and product knowledge. The design perspective focuses on knowledge creation about the interactions between people and the enveloping environment, including technologies when designing and using artifacts.
- ***Figure 2 [right]***, "The Last-Mile Gap" between rationalistic view and design perspectives bridging by behavioral design.

2. Empirical Context of Design Research

In this section, I explore the related research and knowledge gap by following topics of the sub-research questions.

2.1 Exploring Designer-AI Collaboration

The incorporation of AI into design has been influenced by various disciplines, including human-computer interaction (HCI), design research, and recent explorations into computational approaches of behavioral science.. HCI researchers have investigated the challenges of working with data and AI as design materials, as well as the role of design in mitigating potential harms of data-driven algorithmic systems (Yang et al., 2020). Design researchers have explored the risks and societal consequences of AI, particularly in perpetuating existing inequities and biases (Birhane, 2021; Lee et al., 2020; Sloane et al., 2020). Researchers exploring computational approaches to behavioral science have developed AI tools and techniques that can be leveraged by designers to create more personalized and adaptive user experiences (Amershi et al., 2019). These contextual characteristics have shaped design discourses and practices by introducing new concepts, vocabularies, and frameworks for understanding and working with AI in design. For example, the concept of "*AI as design material*" (Yang et al., 2020) has emerged to describe the way designers can leverage AI technologies to create new forms of user experiences and interactions. Similarly, the notion of "*AI as autonomous choice architect*" (Mills & Sætra, 2022) has been used to explore the ethical and accountability implications of AI systems that can influence human behavior through personalized choice environments.

The integration of AI into design practices has led to a shift in the way designers approach problem-solving and ideation. Traditionally, design has been a human-centered process, with designers relying on their own creativity, intuition, and understanding of user needs to generate solutions. However, with the advent of AI, designers are now able to leverage vast amounts of data and computational power to inform their decision-making and generate novel design ideas (Yildirim et al., 2022). This has led to a new paradigm of "data-driven design," where designers collaborate with AI systems to analyze user behavior, identify patterns, and create personalized experiences (Yang et al., 2018).

Building on these developments, I believe that to fully welcome and leverage hybrid intelligence in the design process, the research opportunity and direction is to consider AI as a design collaborator. This interpretation of AI as a collaborator in the design process differs from other interpretations that view AI as a mere tool or a replacement for human designers. By framing AI as a collaborator, this research acknowledges the agency and potential contributions of AI systems in the design process while also recognizing the importance of human oversight and accountability. This interpretation aligns with the concept of "hybrid intelligence," which emphasizes the complementary strengths of humans and machines in problem-solving and decision-making (Akata et al., 2020).

In conclusion, hybrid intelligence for designing process presents both opportunities and challenges for this research. The interpretation of AI as a collaborator in the design process acknowledges the agency and potential contributions of AI systems while recognizing the importance of human oversight and accountability, aligning with the concept of "hybrid intelligence" (Akata et al., 2020). By critically examining the historical and current influences of AI on design practices, as well as the ethical and accountability issues raised by the use of AI in design, this research aims to summarize insights and strategies of designer-AI collaboration. To support the major focus on designing for AI's last mile problem, I would refer to this context to inform practices by embracing the potential of hybrid intelligence for design research, being mindful of its limitations and implications, developing a deep understanding of AI-empowered design tools, collaborating closely with AI experts and stakeholders, and adopting a reflective and iterative approach that prioritizes human oversight and accountability to create innovative, personalized, and ethically sound solutions.

2.2 Identifying the “Last Mile Problem” and Desiging Effective Interventions

Transition from designing with AI to designing for AI, it's important to learn how behavioral knowledge aids in identifying the "last mile problem" (Berinato, 2019) and designing effective heterogeneity-respecting behavioral interventions in AI-empowered services.

The "last mile problem" of AI describes the issue of producing data-evidenced insights but failing to communicate them effectively, leading to wasted or misapplied information (Logg, 2019). To realize the full potential of algorithms and address the last mile problem, AI-empowered innovation needs behavioral design. Although algorithms have the potential to greatly improve human judgment and decision making, as they generally outperform the accuracy of experts when directly compared (Logg, 2019), people can only leverage the accuracy of algorithms if they are willing to listen and learn how to use them. A specific example of the "last mile" gap in healthcare is an AI system designed to assist radiologists in detecting lung cancer from chest X-rays. While the AI system may demonstrate high accuracy in a controlled research setting, it faces numerous challenges when deployed in a real-world clinical environment (Cabiza et al., 2020). These challenges partly arise from the AI system's performance being compromised by lower-quality or inconsistently labeled real-world data (hiatus of machine experience), but mostly from radiologists' reluctance to trust the AI system's predictions patients' fear of unfamiliarity with machines and concerns about privacy (hiatus of human trust), , and organizations' inadequate maintenance of the devices (hiatus of organization behavior) (Cabiza et al., 2020). The "last mile problem" is not unique to healthcare; it is prevalent across various AI applications. Despite the rapid evolution of intelligent data analytics, over 85% of AI innovation projects fail to create value for users or deliver viable services (Ermakova et al., 2021; Joshi et al., 2021; Weiner, 2020). Many of these failures stem from a lack of human-centered design, as design research is often not involved until after the decision of what to innovate has already been made (Kross and Guo, 2021; Nahar et al., 2022; Piorkowski et al., 2021). Thus, there is a significant

opportunity for designers to step in and bridge the gap between AI's capabilities and the needs of multiple users.

Mirroring the universalism versus pluralism conflict in design research, modern behavioral science has also faced significant criticism in recent years (Mills et al., 2023). Some of this criticism highlights the need for more contextual behavioral approaches that incorporate heterogeneity (Mills, 2022; Szaszi et al., 2022). For users, this "heterogeneity revolution" (Bryan et al., 2021) is likely to be promoted and accelerated by AI technologies (Rauthmann, 2020), both as a new tool for behavioral science and in conjunction with existing strategies.

Recent studies have probed results using moderation and mediation to identify heterogeneous effects within samples, like evaluating calorie labels or COVID-19 interventions. This deepens understanding of influential factors, enabling tailored interventions for specific environments, individuals, or policy goals (Mills, 2022; Sunstein, 2022). AI can help address heterogeneity analysis challenges (Mills et al., 2023). Deep learning models can holistically examine unique user profiles, integrating more heterogeneity than moderation approaches. Individual-level variables can combine with contextual factors like time or location (Buyalskaya et al., 2023), further accounting for heterogeneity, as many "choice architects" already do (Mills & Sætra, 2022).

Moreover, the heterogeneity revolution invites embracing behavior's complexity as part of complex adaptive systems (Hallsworth, 2023). Complexity perspectives view behavior as part of wider systems, with variables representing intervention points for behavior change (Beer, 1970). Influential "leverage points" have outsized system effects, suggested as valuable targets for impactful behavioral interventions (Abson et al., 2017). AI shows promise for mapping behavioral systems and locating leverage points, potentially enhancing intervention effectiveness (Hallsworth, 2023; Schmidt & Stenger, 2021).

Grounded in the paradigm of the heterogeneity revolution and the concept of behavioral leverage points within complex socio-technical AI systems, I propose that the "designing for AI" aspect of this research aims to identify heterogeneity-respecting behavioral interventions through behavioral design models or tools empowered by big data and analysis algorithms, seeking "leverage point" and designing ethical and effective solutions under hybrid intelligence support. By critically leveraging big data and analysis algorithms to address heterogeneity and complexity, this research seeks to design effective and responsible behavioral interventions in AI products/services, ultimately addressing the "last mile problem" between AI products, practitioners, users, and other related stakeholders from a systematic perspective.

2.3 Evolving Design Methods and Tools

Boundary objects (Star & Griesemer, 1989), which are information or resources used by collaborative teams to foster shared understanding (Lee, 2007), can scaffold cross-disciplinary collaboration among AI practitioners and stakeholders (Yang et al., 2019).

Design and data science collaboration in the context of AI development often faces challenges due to a lack of shared workflow or common language (Yildirim et al., 2022). This gap is characterized by designers envisioning AI ideas that are beyond the limits of existing AI capabilities and cannot be built, while data scientists build AI solutions that users do not want (Yang et al., 2019). Moreover, AI experts can be a scarce resource for design teams (Yang et al., 2018).

Recent HCI research has highlighted the use of boundary objects to facilitate collaboration across different roles and disciplines in industry AI teams (Holstein et al., 2019). One study reported that abstractions of AI capabilities and data visualizations served as boundary objects to facilitate conversations between UX and AI expertise (Yang et al., 2018). Some HCI researchers reflecting on their own design process proposed using wireframes with data annotations as boundary objects (Yang et al., 2019). The most recent study showed that artifacts such as flow diagrams, system maps, and service data blueprints supported participants in establishing a shared understanding during envisioning and detailing data dependencies during prototyping (Yildirim et al., 2022).

Despite these promising findings, there remain significant challenges and opportunities for creating new design methods and tools to support effective human-AI collaboration. One key challenge is developing and assessing boundary objects that can help in AI problem formulation, a critical stage in the AI development process. Furthermore, it is worth exploring whether these boundary objects can be augmented to scaffold discussions around fairness, bias, and privacy, which are crucial ethical considerations in AI development.

Another challenge lies in understanding how collaboration unfolds across multiple roles in AI teams, beyond the focus on design practitioners. Other roles, such as data scientists, business managers, and software developers, should also be considered to create more comprehensive and effective design methods and tools. This presents an opportunity for future research to investigate the types of boundary objects that can help bridge multiple disciplines and stakeholders throughout the AI design and deployment process, and how these objects might facilitate collaboration in various AI development contexts.

Moreover, there is an opportunity to create new design methods and tools that not only facilitate collaboration but also actively promote ethical considerations and responsible AI development practices. By embedding principles of fairness, transparency, and accountability into the design process itself, these new methods and tools could help ensure that AI systems are developed in a socially responsible manner.

In conclusion, while recent research has highlighted the potential of boundary objects to support human-AI collaboration, there remain significant challenges and opportunities for creating new design methods and tools that can effectively bridge the gap between designers, data scientists, and other stakeholders in AI development. By addressing these challenges and seizing these opportunities, researchers and practitioners can work towards creating more effective, ethical, and socially responsible AI systems.

2.4 Envisioning Role and Responsibility of Designers

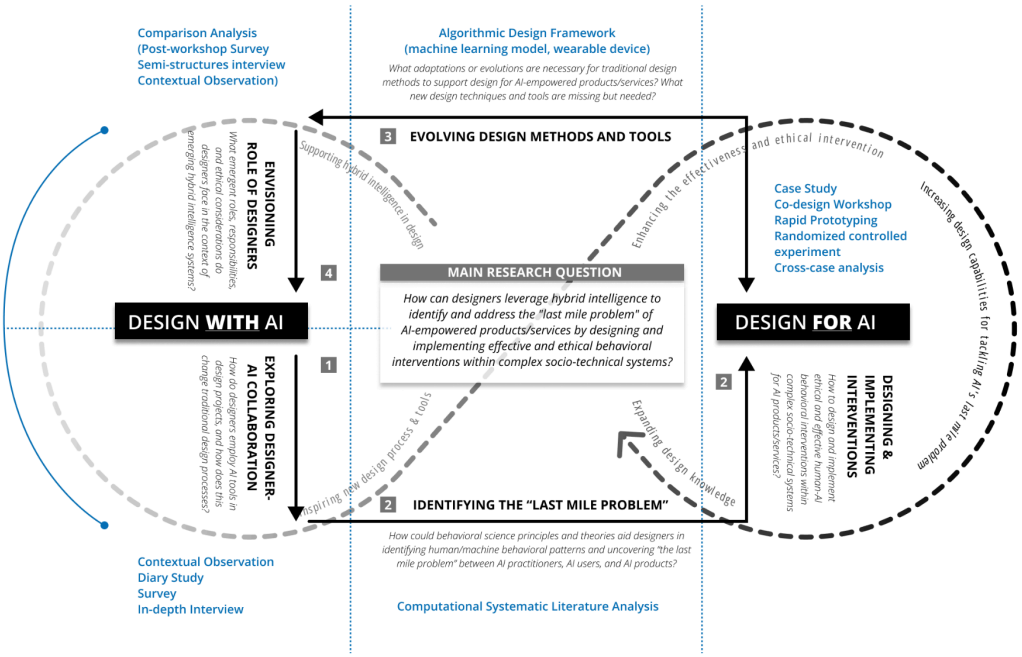
As AI technologies continue to advance and integrate into various aspects of our lives, it is meaningful to define the role and responsibility of designers in shaping the future of human-AI hybrid intelligence. Recent studies have investigated the current collaboration modes of designers within AI design teams, highlighting their contributions in designing human-AI interactions, facilitating alignment, and broadening AI's value space (Yildirim et al., 2022; Windl et al., 2022; Mueller et al., 2019). However, as the boundaries between design of, for, and with AI become increasingly blurred, designers must adapt and learn to use AI while working to create better AI-assisted products and services.

Designers' involvement in AI development could ensure that user needs and values are prioritized throughout the entire lifecycle of AI systems (Yildirim et al., 2022). A recent study found designers who work in AI developing team help with facilitate alignment between disciplines and stakeholders, using knowledge elicitation exercises and boundary objects to set project goals, requirements, and success metrics that prioritize user needs and values. Additionally, designers engage in problem setting, envisioning, and reframing to align on the right design, using concept mapping, co-creation workshops, and systems design methods to explore relationships between technical, cultural, and organizational challenges (Yildirim et al., 2022).

Building on these empirical findings, there are significant opportunities for designers to influence and shape the future of human-AI hybrid intelligence. By actively engaging in designing for AI, leveraging their unique skills, and contributing to the development of new design methods, tools, and frameworks, designers can help grow traditional design knowledge and inform the development of ethical, user-centered, and socially responsible AI systems. Although this section is not the major focus of the research, it can be a meaningful ending step to build on the findings and look forwards to the evolving role and responsibility of designers involved with the new behavioral design approach, preparing them for hybrid intelligence futures.

3. Principles and Methods of Design Research

To address the complex dual-force research question, I plan to employ a mixed-methods approach, combining qualitative and quantitative techniques to gather and analyze data from various sources. This section introduces the methods I plan to use and demonstrates why these approaches are suited to my research by following the sub-research questions and dual-focus diagram below (fig.3). I will introduce my research plan starting with "designing with AI" as a pilot exploration, transitioning between "design with/for AI" as a critical milestone, and then focusing on "designing for AI" for the major research practice. Finally, I will briefly revisit "design with AI" to envision future possibilities.



3.1 Piloting from "Design with AI" - *Investigating current status of designer-AI collaboration through qualitative research*

Working as a Teaching Assistant for the class "Communication Systems: Visualizing Contexts of AI," I have the opportunity to observe 16 design students actively using generative AI tools to build visualizations that address communication design challenges in four AI-deployed domains: healthcare, policy, education, and mobility. Through observing their teamwork and weekly report presentations, I have made several preliminary observations:

1. all students agree that using AI is inevitable and necessary in current and future design processes;
2. current AI tools primarily assist with prototyping and iteration, while only a few support design thinking and modeling;
3. all students recognize that AI tools can be biased and make mistakes, and many treat AI as an asset creator and assistant rather than merely a tool.

These initial findings suggest that hybrid intelligence between designers and AI is already partly occurring within the design process. To effectively research “last mile” design challenges (regardless of whether involving AI or non-AI products), it is crucial to establish a foundational understanding of the current status of designer-AI hybrid intelligence. This understanding will inform the identification of new working processes, tools, strategies, and potential pitfalls to be aware of before addressing design challenges. Therefore, I propose to begin by focusing on "Design with AI" and exploring the first sub-research question: *How can AI be effectively employed to assist designers, and how does this change traditional design processes?* The aim is to gain a basic understanding of : 1) where and how the traditional design process has been altered, 2) what AI aids are effective during the design process and when, and 3) what is lacking and what requires attention when solving AI-related design challenges.

Building upon this experience, I plan to continue conducting **contextual observation** to investigate current state of designer-AI collaboration behaviors. Observing designers as they employ AI in their working process will enable me to identify issues they may not have mentioned in interviews. I aim to gain initial findings and hypotheses around the current roles of designers and AI in the design process, what types of AI assist designers better than others, and how the traditional design process has changed. Data triangulation, combining observations with interviews and other forms of research like surveys or diary studies, is critical to ensure a comprehensive understanding of the hybrid intelligence for design.

The next steps include conducting **diary studies** and **surveys**. I plan to organize the question structure based on my findings and hypotheses from the observation, and ask students to do diary study or survey as part of the final assignment at the end of the course. This step aims to clarify and refine the understanding of the current roles of designers and AI in the design process, why some AI tools assist designers better than others, and whether and how design students perceive AI as changing the traditional design process.

In-depth semi-structured interviews will also be conducted to investigate deeper into understanding human-AI collaboration in the design process. I plan to expand this study from institution to industry, and interview 20 people working in different design roles (researcher, designer, manager) at various levels and organization sizes to understand design processes and collaboration modes in real-life projects. Participants will be asked to articulate challenges, pain points, and best practices for designing human-AI interactions.

By employing these three research methods, the aim is to conduct robust qualitative research to understand the current state of human-AI collaboration in the design process. The insights gained from this study could support the establishment of hybrid intelligent norms, identify opportunity spaces for design methods and tools, and open up the possibility of introducing AI/ML power in research for the next steps – computational literature review and algorithmic-empowered design frameworks.

3.2 Critical Milestones on Transitioning between Design with/for AI

3.2.1 *Exploring behavioral principles for AI from computational literature analysis*

The central focus of my research is the application of a behavioral perspective in tackling the wicked "last mile problem" of AI. Therefore, the second sub-research question, *how can behavioral science principles and theories aid designers in identifying "the last mile problem" of AI-empowered products/services in complex socio-technical systems?*, represents a critical milestone. Progress cannot be made towards the next steps of design practices and experiments without first addressing this question.

To investigate this, I plan to conduct a **computational systematic literature review**. This review aims to 1) examine what behavioral frameworks and theories have been introduced into AI-related studies, 2) identify behavioral knowledge gaps in the current literature, and 3) investigate potential behavioral frameworks for addressing "the last mile problems" in AI research.

A Computational Literature Review (CLR) is a method that identifies and evaluates all relevant literature on a topic using quantitative text analysis methods on extensive databases. CLR demonstrates its strengths in recognizing overlapping trends, structures, and patterns. It automates some of the analysis of research articles by examining impact (citations), structure (co-authorship networks), and content (behavioral-focused) across various disciplines simultaneously (HCI, Behavioral Science, and Design Studies). As such, CLR is the most suitable approach for investigating cross-disciplianry behavioral design perspectives for AI.

I propose that this CLR study can focus on examining how behavioral science and persuasive technology concepts and theories have been applied in AI design and deployment studies. By performing a computational analysis of behavioral research data across different disciplines, the ultimate goal is to identify key behavioral principles or frameworks that can guide designers in problem identification and framing for AI-related product/service challenges.

The proposed CLR study will contribute to the development of a solid foundation for understanding the intersection of behavioral science and AI design. The findings will inform the subsequent stages of this research, including the design and implementation of behavioral interventions for AI's "last mile problem." By establishing a comprehensive understanding of the existing literature and identifying key behavioral principles, this study will ensure that the design practices and experiments in the later stages of the research are grounded in a robust theoretical framework.

3.2.2 *Evolving design methods and tools via algorithmic design frameworks*

Another sub-research question at the intersection of Design with/for AI is: *What adaptations or evolutions are necessary in traditional design methods to effectively integrate AI, and what gaps currently exist in this regard?* I plan to evolve design tools by computationalizing certain paper-based design frameworks into digital versions, **equipped with data analysis ML model**, enabling these design tools to handle more research data and generate more heterogeneous findings, potentially benefiting design strategies.

For instance, the Insight Clustering Matrix could be a suitable design framework for computationalization. Equipped with data analysis algorithms, the computationalized Insight Clustering Matrix could easily absorb a larger volume of data and automatically cluster and generate insight patterns. This would allow designers to collect more comprehensive and detailed research data and approach each design challenge more pluralistically.

Another plan option is to directly computationalize the behavioral model/framework I develop following the proposed CLR study. This algorithmic-empowered behavioral framework can be **equipped with wearable sensors**. In this way, the framework can directly receive users' bio-behavioral data, providing not only more data to learn from but also increasing the diversity of data types. This approach aligns with the heterogeneous revolution ideas from behavioral science, which could help increase prediction accuracy or intervention efficiency. It also incorporates inclusive design and pluriversal design ethics, as the framework and sensors could aid in discovering overlooked cognitive and physiological data and generate various analyses for each user.

The goal of this study is to examine the potential of AI-assisted design models and explore possible ways to adapt traditional design tools for hybrid intelligence futures. It is important to note that this algorithmic design frameworks research does not necessarily have to be conducted after 3.2.1 and 3.1. For example, the Computational Insight Clustering Matrix could be developed concurrently with the CRL study. Moreover, this research is not limited to a single study; I can computationalize more than one framework or iterate on a single behavioral framework several times. The reason I set this study as a milestone between the design with/for AI sections is twofold: firstly, I must complete at least one algorithmic design tool before "designing for the last mile problem" so that I can apply it in real-life practical studies; secondly, after the design practices, I will need to evaluate and iterate several rounds, and I plan to consider the final algorithmic design model as a major contribution. Therefore, it is also a milestone near the end of the research.

By exploring the computationalization of design frameworks and the incorporation of ML-empowered data analysis algorithms and wearable devices, this research aims to not only enhance the efficiency and effectiveness of behavioral design’s capability in addressing AI’s last milep problems, but also contribute to the advancement of design methods and tools in the context of AI integration.

3.3 Major Focus on Designing for AI - *Designing and Implementing behavioral interventions in case studies*

To further explore how behavioral principles can help identify and address the "last mile problem" of AI through behavioral interventions, I plan to conduct multiple **case studies**. Case studies are well-suited for this research as they allow for in-depth, contextual examination of real-world phenomena, enabling the investigation of complex issues and the development of holistic, practical insights. These case studies will involve real-world AI products or services, and will be carried out in collaboration with industry partners or through research assistant work.

I will identify and select 3-5 case studies that represent diverse AI-empowered products or services across different domains, such as healthcare, mobility, and education. The selection criteria will include the presence of a clear "last mile problem" – a clear gap between producing and utilizing insights from algorithms (Logg, 2019) – in an existing AI product or service, accessibility to key stakeholders (e.g., AI developers, users, regulators), and the willingness of the organization to collaborate and implement behavioral interventions.

For each case study, I will conduct a thorough analysis to identify and frame the specific "last mile problem" using the behavioral principles and frameworks derived from the previous computational literature review (Section 3.2.1). This will involve interviews with key stakeholders to understand their perspectives and challenges, observational studies of the AI product or service in use, and analysis of existing data (e.g., user feedback, usage metrics) to identify behavioral patterns and pain points.

Based on the insights gathered from the problem identification and framing phase, I will design and develop tailored behavioral interventions for each case study. This process will leverage the computationalized design frameworks (Section 3.2.2) and involve **co-design workshops** with key stakeholders to generate intervention ideas, **rapid prototyping** and iterative refinement of interventions, and ethical considerations and risk assessment of proposed interventions. Co-design workshops and participatory methods are essential for ensuring that interventions are user-centered, contextually relevant, and ethically sound (Sanders & Stappers, 2008). Rapid prototyping and iterative refinement allow for the early identification and resolution of design challenges, improving the effectiveness and feasibility of the interventions (Neeley et al., 2013).

The designed interventions will be implemented in real-world settings for each case study based on the plurversal principles. To evaluate the effectiveness and unintended consequences of the interventions, I will employ a mixed-methods approach. **Quantitative methods**, such as randomized controlled experiments, wearable devices, pre-post surveys will provide objective measures of intervention effectiveness and impact. **Qualitative methods**, including interviews, focus groups, and observational studies, will offer rich, contextual insights into user experiences, perceptions, and unintended consequences.

After completing the individual case studies, I will conduct a **cross-case analysis** to identify common patterns, challenges, and best practices in designing and implementing behavioral interventions for AI's "last mile problem." Cross-case analysis is a powerful technique for enhancing the generalizability of case study findings and developing robust, theory-based insights (Yin, 2018). This synthesis will contribute to refining the behavioral principles and algorithmic frameworks for AI-empowered products and services.

The case study approach, combined with a mixed-methods research design and participatory methods, provides a robust framework for investigating the practical application of behavioral principles in addressing AI's "last mile problem." By grounding the research in real-world contexts, engaging diverse stakeholders, and employing rigorous data collection and analysis techniques, this section aims to generate actionable method and transferable insights that can inform the development of effective behavioral interventions for AI-empowered products and services.

3.4 Looking Forwards and Echoing Designing with AI - *Envisioning designers role via comparison analysis*

The final stage of this research revisits the beginning hybrid intelligence topic to address the sub-research question: *What emergent roles, responsibilities, and ethical considerations do designers face in the context of emerging hybrid intelligence systems?* This study builds upon the data collected from the case studies in Section 3.3, such as post-workshop surveys, semi-structured interviews, and observations, for analysis and reflection.

I plan to conduct a **comparative analysis** of the results from this study and those from the 3.1 study on the current status of human-AI design collaboration. By comparing the new experimental situation with the current real-life status, I aim to identify similarities and differences and envision the roles of designers and AI in hybrid intelligence futures.

This comparative analysis will provide insights into the potential evolution of designers' roles and responsibilities as AI becomes increasingly integrated into the design process. It will also shed light on the ethical considerations that designers may face towards hybrid inytelligent futures.

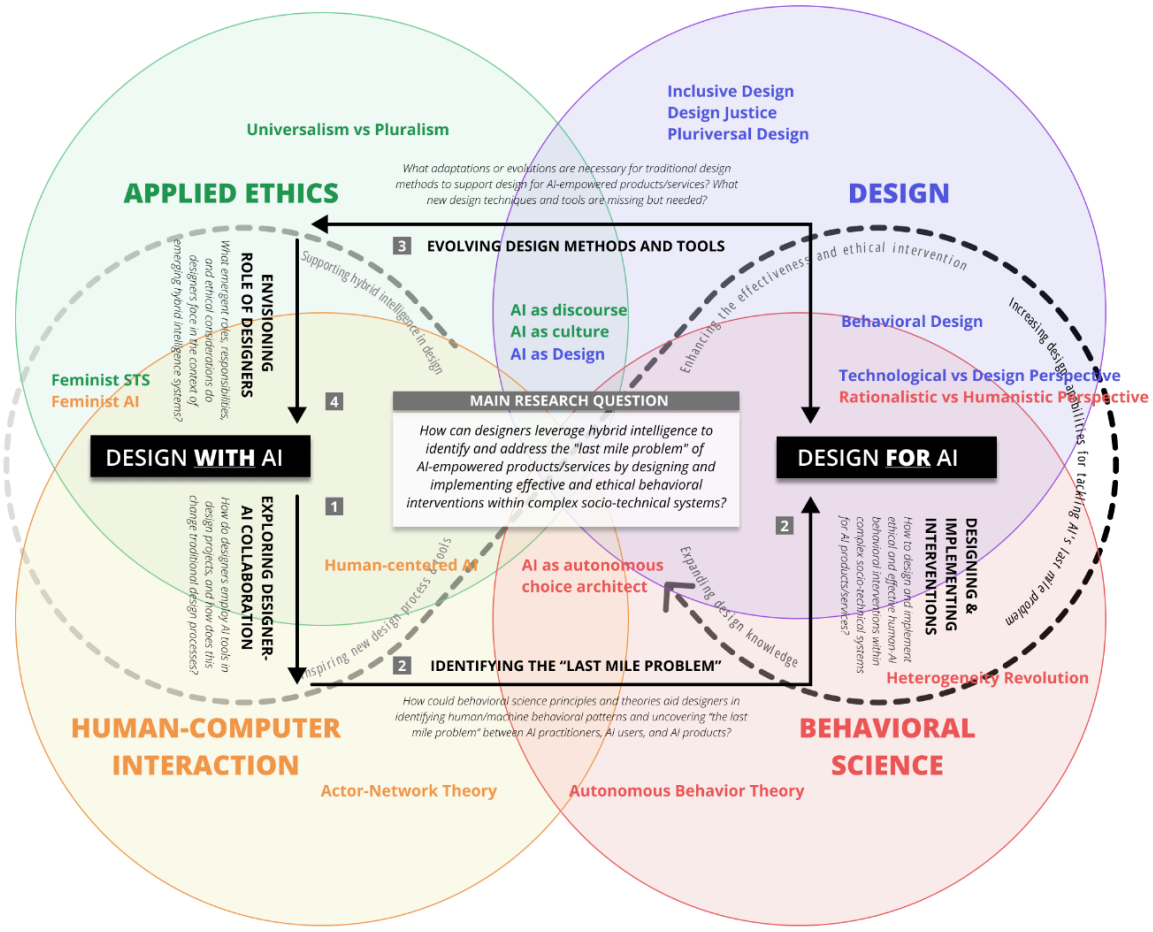
4. Your Research Framework

The theoretical framework below summarizes my research proposal, combining almost all the elements mentioned, including research questions, sub-research topics, research logic and plan, principles, and theories. The aim is to provide a bird's-eye view of my research from a holistic perspective.

Key Visual Elements	**Meaning Explanation**
The "Hourglass - Design with/for AI"	Highlights the dual-force nature of my research, with the major focus centered on "design for AI."
The square arrow lines	Represent the sub-research questions derived from the main question and contribute round the main question.
The numbers beside the sub-question titles	Indicate my research plan, but the order is not strictly fixed.
The overlapping of the square arrow lines with the "hourglass" arrow line	Emphasizes that the research plan and proposed studies do not necessarily follow the numbered sequence but can be circular or repeated.
The curved words on the "hourglass"	Indicate the purpose of each sub-research and illustrate how one sub-research can benefit the others.
The Venn diagram	The venn circles show that my research questions lie at the intersection of HCI, Behavioral Science, Applied Ethics from computer and social sciences, and Design. The small words on the Venn diagram represent the principles and theories supporting my research. The color code indicates the disciplines from which each theory originates.

This theoretical framework serves as a visual representation of the complex and interconnected nature of my research proposal. By presenting the key elements and their relationships in a concise and organized manner, it helps to clarify the overall structure and logic of my research. The framework also highlights the interdisciplinary approach, drawing from various fields to address the research questions effectively.

Furthermore, the framework emphasizes the iterative and adaptive nature of the research plan, allowing for flexibility and refinement as the study progresses. By providing a comprehensive overview of the research proposal, this theoretical framework facilitates a better understanding of the scope, objectives, and potential contributions of my research in the field of design for AI.


Universalism vs Pluralism
APPLIED ETHICS
Inclusive Design
Design Justice
Pluriversal Design
DESIGN
What adaptations or evolutions are necessary for traditional design methods to support design for AI-empowered products/services? What new design techniques and tools are missing but needed?
3 EVOLVING DESIGN METHODS AND TOOLS
ENVISIONING ROLE OF DESIGNERS
What emergent roles, responsibilities, and ethical considerations do designers face in the context of emerging hybrid intelligence systems?
Supporting hybrid intelligence in design
AI as discourse
AI as culture
AI as Design
Enhancing the effectiveness and ethical intervention
Behavioral Design
Increasing design capabilities for tackling AI's last mile problem
Feminist STS
Feminist AI
Technological vs Design Perspective
Rationalistic vs Humanistic Perspective
4
MAIN RESEARCH QUESTION
How can designers leverage hybrid intelligence to identify and address the "last mile problem" of AI-empowered products/services by designing and implementing effective and ethical behavioral interventions within complex socio-technical systems?
DESIGN WITH AI
DESIGN FOR AI
1
2
EXPLORING DESIGNER-AI COLLABORATION
How do designers employ AI tools in design projects, and how does this change traditional design processes?
DESIGNING & IMPLEMENTING INTERVENTIONS
How to design and implement ethical and effective human-AI behavioral interventions within complex socio-technical systems for AI products/services?
Human-centered AI
AI as autonomous choice architect
Inspiring new design process & tools
Expanding design knowledge
2 IDENTIFYING THE "LAST MILE PROBLEM"
Heterogeneity Revolution
How could behavioral science principles and theories aid designers in identifying human/machine behavioral patterns and uncovering "the last mile problem" between AI practitioners, AI users, and AI products?
HUMAN-COMPUTER INTERACTION
BEHAVIORAL SCIENCE
Actor-Network Theory
Autonomous Behavior Theory

5. Reference

- Abascal, J., & Nicolle, C. (2005). Moving towards inclusive design guidelines for socially and ethically aware HCI. *Interacting With Computers*, *17*(5), 484–505. https://doi.org/10.1016/j.intcom.2005.03.002

- Abson, D. J., Fischer, J., Leventon, J., Newig, J., Schomerus, T., Vilsmaier, U., Von Wehrden, H., Abernethy, P., Ives, C. D., Jager, N. W., & Lang, D. J. (2016). Leverage points for sustainability transformation. *Ambio*, *46*(1), 30–39. https://doi.org/10.1007/s13280-016-0800-y

- Acemoğlu, D. (2021). *Harms of AI*. https://doi.org/10.3386/w29247

- Adam, A. (1995). Artificial intelligence and women's knowledge. *Women's Studies International Forum*, *18*(4), 407–415. https://doi.org/10.1016/0277-5395(95)80032-k

- Adam, A. (1997). *Artificial Knowing: gender and the thinking machine*.

- Akata, Z., Balliet, D., De Rijke, M., Dignum, F., Dignum, V., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K. V., Hoos, H. H., Hung, H., Jonker, C. M., Monz, C., Neerincx, M. A., Oliehoek, F. A., Prakken, H., Schlobach, S., Van Der Gaag, L., Van Harmelen, F., . . . Welling, M. (2020). A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. *Computer (Long Beach, Calif. Print)*, *53*(8), 18–28. https://doi.org/10.1109/mc.2020.2996587

- Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S. T., Bennett, P., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for Human-AI Interaction. *In Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems.* https://doi.org/10.1145/3290605.3300233

- Bardzell, S. (2010). Feminist HCI. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1301–1310*. https://doi.org/10.1145/1753326.1753521

- Beer, S. (1970). Managing modern complexity. *Futures (London)*, *2*(3), 245–257. https://doi.org/10.1016/0016-3287(70)90028-5

- Bendixen, K., & Benktzon, M. (2015). Design for All in Scandinavia – A strong concept. *Applied Ergonomics*, *46*, 248–257. https://doi.org/10.1016/j.apergo.2013.03.004

- Berinato, S. (2024, April 2). *Data science and the art of persuasion*. Harvard Business Review. https://hbr.org/2019/01/data-science-and-the-art-of-persuasion

- Bianchin, M., & Heylighen, A. (2017). Fair by design. Addressing the paradox of inclusive design approaches. *the Design Journal (Aldtershot)*, *20*(sup1), S3162–S3170. https://doi.org/10.1080/14606925.2017.1352822

- Birhane, A. (2021). Algorithmic injustice: a relational ethics approach. *Patterns (New York)*, *2*(2), 100205. https://doi.org/10.1016/j.patter.2021.100205

- Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*, *5*(8), 980–989. https://doi.org/10.1038/s41562-021-01143-3

- Buxton, B. (2007). *Sketching user experiences: getting the design right and the right design*. https://cds.cern.ch/record/1190758/files/9780123740373_TOC.pdf

- Buyalskaya, A., Ho, H. S., Milkman, K. L., Li, X., Duckworth, A., & Camerer, C. F. (2023). What can machine learning teach us about habit formation? Evidence from exercise and hygiene.

Proceedings of the National Academy of Sciences of the United States of America, *120*(17). https://doi.org/10.1073/pnas.2216115120

- Cabitza, F., Campagner, A., & Balsano, C. (2020). Bridging the "last mile" gap between AI implementation and operation: "data awareness" that matters. *Annals of Translational Medicine (Print)*, *8*(7), 501. https://doi.org/10.21037/atm.2020.03.63

- Christian, B. (2020). *The alignment problem: Machine Learning and Human Values*. National Geographic Books.

- Collective, C. R. (2014). A Black Feminist statement. *Women's Studies Quarterly*, *42*(3–4), 271–280. https://doi.org/10.1353/wsq.2014.0052

- Costanza-Chock, S. (2018). Design Justice: towards an intersectional feminist framework for design theory and practice. *Proceedings of DRS*. https://doi.org/10.21606/drs.2018.679

- De Sio, F. S., & Van Den Hoven, J. (2018). Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI*, *5*. https://doi.org/10.3389/frobt.2018.00015

- De Vos, J. (2020). The digitalisation of (Inter)Subjectivity. In *Routledge eBooks*. https://doi.org/10.4324/9781315167350

- Dove, G., Halskov, K., Forlizzi, J., & Zimmerman, J. (2017). UX design Innovation. *In Proceedings of the 2017 Chi Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3025453.3025739

- Ermakova, T., Blume, J., Fabian, B., Fomenko, E. V., Berlin, M., & Hauswirth, M. (2021). Beyond the hype: Why do Data-Driven projects fail? *Proceedings of the Annual Hawaii International Conference on System Sciences (1999)*. https://doi.org/10.24251/hicss.2021.619

- Forsythe, D. E. (1993). Engineering Knowledge: The construction of knowledge in artificial intelligence. *Social Studies of Science*, *23*(3), 445–477. https://doi.org/10.1177/030631279302300302

- Gorkovenko, K., Burnett, D. J., Thorp, J., Richards, D., & Murray-Rust, D. (2020). Exploring the future of Data-Driven product design. *In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* https://doi.org/10.1145/3313831.3376560

- Grosz, B. J. (2019). Some reflections on Michael Jordan's article "Artificial Intelligence—The Revolution hasn't happened yet." *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.97b95546

- Grudin, J. (2009). AI and HCI: Two Fields Divided by a Common Focus. *Ai Magazine*, *30*(4), 48–57. https://doi.org/10.1609/aimag.v30i4.2271

- Gunkel, D. J. (2017). Mind the gap: responsible robotics and the problem of responsibility. *Ethics and Information Technology*, *22*(4), 307–320. https://doi.org/10.1007/s10676-017-9428-2

- Hagendorff, T. (2021). Blind spots in AI ethics. *AI And Ethics*, *2*(4), 851–867. https://doi.org/10.1007/s43681-021-00122-8

- Hallsworth, M. (2023). A manifesto for applying behavioural science. *Nature Human Behaviour*, *7*(3), 310–322. https://doi.org/10.1038/s41562-023-01555-3

- Harawy, D. (1988). Situated Knowledges: the science question in feminism and the privilege of partial perspective. *Feminist Studies*, *14*(3), 575. https://doi.org/10.2307/3178066

- Holstein, K., Vaughan, J., Daumé, H., Dudík, M., & Wallach, H. (2019). Improving fairness in machine learning systems: What do industry practitioners need? *arXiv (Cornell University)*, 600. http://arxiv.org/abs/1812.05239

- Hooks, B. (2014). Feminism is for everybody. In *Routledge eBooks*. https://doi.org/10.4324/9781315743189

- *How AI fails us*. (n.d.). Edmond & Lily Safra Center for Ethics. https://ethics.harvard.edu/how-ai-fails-us

- Institutional Ecology, "Translations" and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39 on JSTOR. (n.d.). *www.jstor.org*. https://www.jstor.org/stable/285080

- Joshi, M. P., Su, N., Austin, R. D., & Sundaram, A. K. (2021). Why So Many Data Science Projects Fail to Deliver. *MIT Sloan Management Review*, *62*(3), 85–89. https://www.research.manchester.ac.uk/portal/en/publications/why-so-many-data-science-projects-fail-to-deliver(032a0402-f02f-43d3-bcda-b8efadaf4ee0).html

- Jylkäs, T., Äijälä, M. H., Vuorikari, T. M., & Rajab, V. (2018). AI assistants as non-human actors in service design. *In Proc of DMI Acad Des Manag Conf.*, 1436–1444. https://lacris.ulapland.fi/en/publications/ai-assistants-as-nonhuman-actors-in-service-design(bc93bc22-a028-4e5e-8ff1-1b46976937da).html

- Kayacik, C., Chen, S., Noerly, S., Holbrook, J., Roberts, A., & Eck, D. (2019). Identifying the intersections. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3290607.3299059

- Khadilkar, P., & Cash, P. (2020). Understanding behavioural design: barriers and enablers. *Journal of Engineering Design*, *31*(10), 508–529. https://doi.org/10.1080/09544828.2020.1836611

- Kross, S., & Guo, P. J. (2021). Orienting, framing, bridging, magic, and Counseling: How data scientists navigate the outer loop of client collaborations in industry and academia. *Proceedings of the ACM on Human-computer Interaction*, *5*(CSCW2), 1–28. https://doi.org/10.1145/3476052

- Lam, M., Ma, Z., Li, A., Freitas, I., Wang, D., Landay, J. A., & Bernstein, M. S. (2023). Model Sketching: Centering Concepts in Early-Stage Machine Learning Model Design. *In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3544548.3581290

- Latour, B. (2013). Reassembling the Social. An Introduction to Actor-Network-Theory (translated by Irina Polonskaya). *Èkonomičeskaâ Sociologiâ*, *14*(2), 73–87. https://doi.org/10.17323/1726-3247-2013-2-73-87

- Lazar, J., Feng, J., & Hochheiser, H. (2010). *Research Methods in Human-Computer Interaction*. http://ci.nii.ac.jp/ncid/BB00763455

- Lee, C. P. (2007). Boundary negotiating artifacts: unbinding the routine of boundary objects and embracing chaos in collaborative work. *Computer Supported Cooperative Work*, *16*(3), 307–339. https://doi.org/10.1007/s10606-007-9044-5

- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1), 205395171875668. https://doi.org/10.1177/2053951718756684

- Lee, M. K., Grgić-Hlača, N., Tschantz, M. C., Binns, R., Weller, A., Carney, M. M., & Inkpen, K. (2020). Human-Centered Approaches to Fair and Responsible AI. *In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3334480.3375158

- Liboiron, M. (2021). Pollution is colonialism. In *Duke University Press eBooks*. https://doi.org/10.1215/9781478021445

- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3313831.3376727

- Lugones, M. (2010). Toward a decolonial feminism. *Hypatia (Edwardsville, Ill.)*, *25*(4), 742–759. https://doi.org/10.1111/j.1527-2001.2010.01137.x

- Malsattar, N., Kihara, T., & Giaccardi, E. (2019). Designing and prototyping from the perspective of AI in the wild. *In Proceedings of the 2019 on Designing Interactive Systems Conference*. https://doi.org/10.1145/3322276.3322351

- McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Антропова, H. B., Ashrafian, H., Back, T., Chesus, M., Corrado, G., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F. J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C., King, D., . . . Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature (London)*, *577*(7788), 89–94. https://doi.org/10.1038/s41586-019-1799-6

- Mele, C., Spena, T. R., Kaartemo, V., & Marzullo, M. (n.d.). Smart Nudging: How Cognitive Technologies enable choice architectures for value co-creation. *Journal of Business Research*, *129*, 949–960. https://doi.org/10.1016/j.jbusres.2020.09.004

- Mills, S. (2022). Finding the 'nudge' in hypernudge. *Technology in Society*, *71*, 102117. https://doi.org/10.1016/j.techsoc.2022.102117

- Mills, S., Costa, S., & Sunstein, C. R. (2023). AI, behavioural science, and consumer welfare. *Journal of Consumer Policy*, *46*(3), 387–400. https://doi.org/10.1007/s10603-023-09547-6

- Mills, S., & Sætra, H. S. (2022). The autonomous choice architect. *AI & Society*. https://doi.org/10.1007/s00146-022-01486-z

- Müller, M., Lange, I., Wang, D., Piorkowski, D., Tsay, J., Liao, Q. V., Dugan, C., & Erickson, T. (2019). How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3290605.3300356

- Nahar, N., Zhou, S., Lewis, G. A., & Kästner, C. (2021). Collaboration challenges in building ML-Enabled systems: communication, documentation, engineering, and process. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2110.10234

- Neeley, W. L., Lim, K., Zhu, A., & Yang, M. C. (2013). Building Fast to think Faster: Exploiting rapid prototyping to accelerate ideation during early stage design. *Proceedings of the ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. https://doi.org/10.1115/detc2013-12635

- Nielsen, C. K. E. B. B., Cash, P., & Daalhuizen, J. (2024). The power and potential of Behavioural Design: practice, methodology, and ethics. *Journal of Engineering Design (Print)*, 1–39. https://doi.org/10.1080/09544828.2024.2322897

- Piorkowski, D., Park, S., Wang, A. Y., Wang, D., Müller, M., & Portnoy, F. (2021). How AI developers overcome communication challenges in a multidisciplinary team. *Proceedings of the ACM on Human-computer Interaction*, *5*(CSCW1), 1–25. https://doi.org/10.1145/3449205

- Rauthmann, J. F. (2020). A (More) behavioural science of personality in the age of Multi–Modal sensing, big data, machine learning, and artificial intelligence. *European Journal of Personality*, *34*(5), 593–598. https://doi.org/10.1002/per.2310

- Reider, D., & Partner, P. (2012). Leverage Points—Places to intervene in a system. In *Routledge eBooks* (pp. 152–172). https://doi.org/10.4324/9781849773386-15

- *Review into bias in algorithmic decision-making*. (2024, February 15). GOV.UK. https://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making/main-report-cdei-review-into-bias-in-algorithmic-decision-making

- Saheb, T. (2022). "Ethically contentious aspects of artificial intelligence surveillance: a social science perspective." *AI And Ethics (Print)*, *3*(2), 369–379. https://doi.org/10.1007/s43681-022-00196-y

- Sanders, E., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, *4*(1), 5–18. https://doi.org/10.1080/15710880701875068

- Sayes, E. (2013). Actor–Network Theory and methodology: Just what does it mean to say that nonhumans have agency? *Social Studies of Science*, *44*(1), 134–149. https://doi.org/10.1177/0306312713511867

- Schmidt, R., & Stenger, K. (2021). Behavioral brittleness: the case for strategic behavioral public policy. *Behavioural Public Policy (Print)*, 1–26. https://doi.org/10.1017/bpp.2021.16

- Sciences from below: feminisms, postcolonialities, and modernities. (2009). *Choice (Chicago, Ill.)*, *46*(06), 46–3202. https://doi.org/10.5860/choice.46-3202

- Shneiderman, B. (2020a). Design lessons from AI's two grand goals: human emulation and useful applications. *IEEE Transactions on Technology and Society*, *1*(2), 73–82. https://doi.org/10.1109/tts.2020.2992669

- Shneiderman, B. (2020b). Bridging the gap between ethics and practice. *ACM Transactions on Interactive Intelligent Systems (Print)*, *10*(4), 1–31. https://doi.org/10.1145/3419764

- Sloane, M., Moss, E., Awomolo, O., & Forlano, L. (2022). Participation is not a design fix for machine learning. *Participatory Approaches to Machine Learning*. https://doi.org/10.1145/3551624.3555285

- Snow, C. C. (1959). *The two cultures*. https://www.jstor.org/stable/pdfplus/1578601.pdf

- Speranza, L. (2023, September 8). *How Netflix's choice engine drives its business - by Eric Johnson - behavioral scientist*. Behavioral Scientist. https://behavioralscientist.org/how-the-netflix-choice-engine-tries-to-maximize-happiness-per-dollar-spent_ux_ui/

- Stumpf, S., Strappelli, L., Ahmed, S., Nakao, Y., Naseer, A., Del Gamba, G., & Regoli, D. (2021). Design methods for artificial intelligence fairness and transparency. *IUI Workshops*. http://ceur-ws.org/Vol-2903/IUI21WS-TExSS-13.pdf

- Sunstein, C. R. (2012). Impersonal Default Rules vs. Active Choices vs. Personalized Default Rules: A Triptych. *Social Science Research Network*. https://doi.org/10.2139/ssrn.2171343

- Sunstein, C. R. (2021). Governing by algorithm? No noise and (Potentially) less bias. *Social Science Research Network*. https://doi.org/10.2139/ssrn.3925240

- *Superminds, not substitutes* | *Deloitte*. (2020, July 31). Deloitte Insights. https://www.deloitte.com/cbc/en/our-thinking/insights/topics/talent/technology-and-the-future-of-work/ai-in-the-workplace.html

- Szászi, B., Higney, A., Charlton, A., Gelman, A., Ziano, I., Aczél, B., Goldstein, D. G., Yeager, D. S., & Tipton, E. (2022). No reason to expect large and consistent effects of nudge interventions. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(31). https://doi.org/10.1073/pnas.2200732119

- Toupin, S. (2023). Shaping feminist artificial intelligence. *New Media & Society*, *26*(1), 580–595. https://doi.org/10.1177/14614448221150776

- Turing, A. (2004). Computing Machinery and Intelligence (1950). In *Oxford University Press eBooks*. https://doi.org/10.1093/oso/9780198250791.003.0017

- Turkle, S. (1989). *Artificial intelligence and psychoanalysis: a new alliance* (pp. 241–268). https://dl.acm.org/citation.cfm?id=66747

- Wajcman, J. (1993). Feminism confronts technology. *British Journal of Sociology (Print)*, *44*(2), 369. https://doi.org/10.2307/591252

- Watson, D. (2019). The rhetoric and reality of anthropomorphism in artificial intelligence. *Minds and Machines (Dordrecht)*, *29*(3), 417–440. https://doi.org/10.1007/s11023-019-09506-6

- Weiner, J. (2020). Why AI/Data science projects fail: How to avoid project pitfalls. *Synthesis Lectures on Computation and Analytics (Online)*, *1*(1), i–77. https://doi.org/10.2200/s01070ed1v01y202012can001

- Windl, M., Feger, S. S., Zijlstra, L., Schmidt, A., & Woźniak, P. W. (2022). 'It is not always discovery Time': Four pragmatic approaches in designing AI systems. *CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3491102.3501943

- Winograd, T. (1996). Bringing design to software. In *ACM eBooks*. https://doi.org/10.1145/229868

- Winograd, T., & Flores, F. (1987). *Understanding computers and cognition: A New Foundation for Design*. Addison-Wesley Professional.

- Wright, D., & Meadows, D. H. (2008). *Thinking in systems*. https://cds.cern.ch/record/1608083

- Yang, Q., Cranshaw, J., Amershi, S., Iqbal, S. T., & Teevan, J. (2019). Sketching NLP. *In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3290605.3300415

- Yildirim, N., Pushkarna, M., Goyal, N., Wattenberg, M., & Viégas, F. B. (2023). Investigating how Practitioners use Human-AI guidelines: A case study on the People + AI Guidebook. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2301.12243

- Yin, R. K. (2017). *Case Study Research and Applications: Design and methods*. http://cds.cern.ch/record/2634179

- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The fight for a human future at the new frontier of power*. https://cds.cern.ch/record/2655106