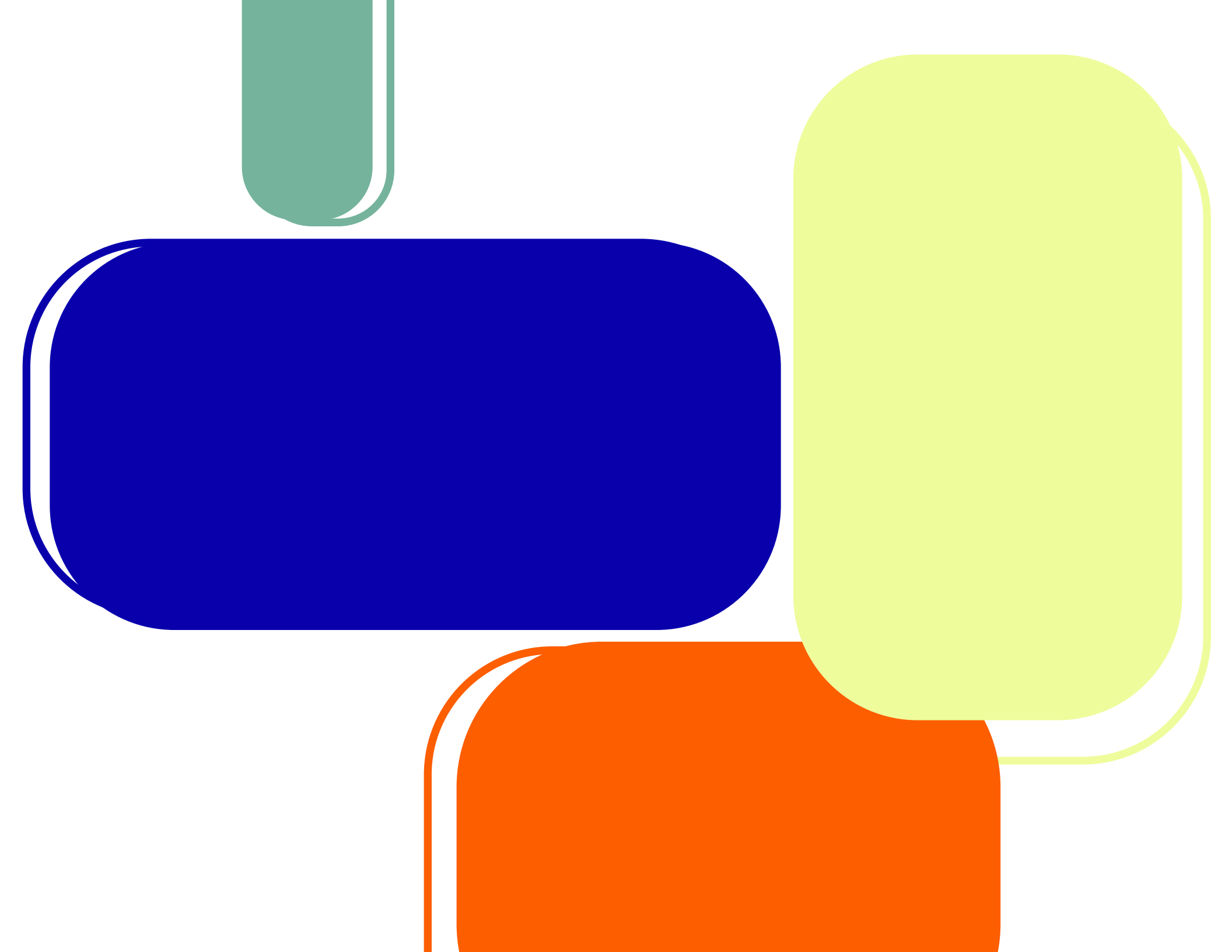
RESPONSIBLE DESIGN OF FUTURE AI TECHNOLOGIES FOSTERING SOCIAL CONNECTIVITY

Noor-ul-Ain Javed

MSc Design for Interaction
Delft University of Technology

August 2024



Responsible Design of Future AI Technologies Fostering Social Connectivity

Master of Science Design for Interaction
Faculty of Industrial Design Engineering
Delft University of Technology

Author
Noor-ul-Ain Javed

Chair
Dr. S. (Sara) Colombo
Faculty of Industrial Design Engineering
Human-Centered Artificial Intelligence,
Sustainable Design Engineering

Mentor
Dr. D. (Dajung) Kim
Faculty of Industrial Design Engineering
Human-Centered Design, Design Aesthetics

August 2024

TU Delft

RESPONSIBLE DESIGN OF FUTURE AI TECHNOLOGIES FOSTERING SOCIAL CONNECTIVITY

Noor-ul-Ain Javed

Masters' Thesis

PREFACE

This research report represents the culmination of several months of challenging myself as both a designer and researcher. Beginning with an in-depth exploration of the intersections among "Social Connectivity," "AI Technology Trends," and "Responsibility," I ultimately present design implications for the responsible development of future AI virtual assistants aimed at enhancing social connectivity.

I began this research with a fractured right ankle (would not recommend) and concluded it fully healed and recovered. The time, physical, and mental effort invested in this work extends far beyond what is captured in this report. I am profoundly grateful for the support of some truly wonderful individuals throughout this journey; without their help, this process would have been unimaginably difficult.

I extend my heartfelt gratitude to my dear supervisors, Dr. Sara Colombo and Dr. Dajung Kim. Their guidance, encouragement, and unwavering confidence in me were instrumental throughout this journey. I also wish to thank Francesca Mauri for her constant positivity and readiness to assist whenever needed. Special appreciation is due to the focus group participants, whose valuable time and active engagement provided crucial insights for this research.

I want to send out love to my dear family. I am forever grateful for my loving parents, my grandma and two sweetest brothers for always having my back and being the biggest supporter of my dreams. A heartfelt thank you to my friends for sharing in both the tears and the smiles. Abhiraj, thank you for always being by my side, giving me the strength and courage I needed. Kash, Kumsal, Dilara, Gul, and David, thank you for the beautiful memories and making me feel at home.

Finally, I would like to express my deepest gratitude to the Justus & Louise van Effen Foundation and TU Delft for deeming me worthy of the prestigious Justus and Louise van Effen Excellence Scholarship. This recognition and financial support not only enabled me to earn this degree but also gave me the opportunity to pursue my dreams.

With all this love and support, I conclude my master's journey and proudly present this graduation project research report. I hope you find the same enjoyment in reading it as I did in creating it.

Noor-ul-Ain Javed

EXECUTIVE SUMMARY

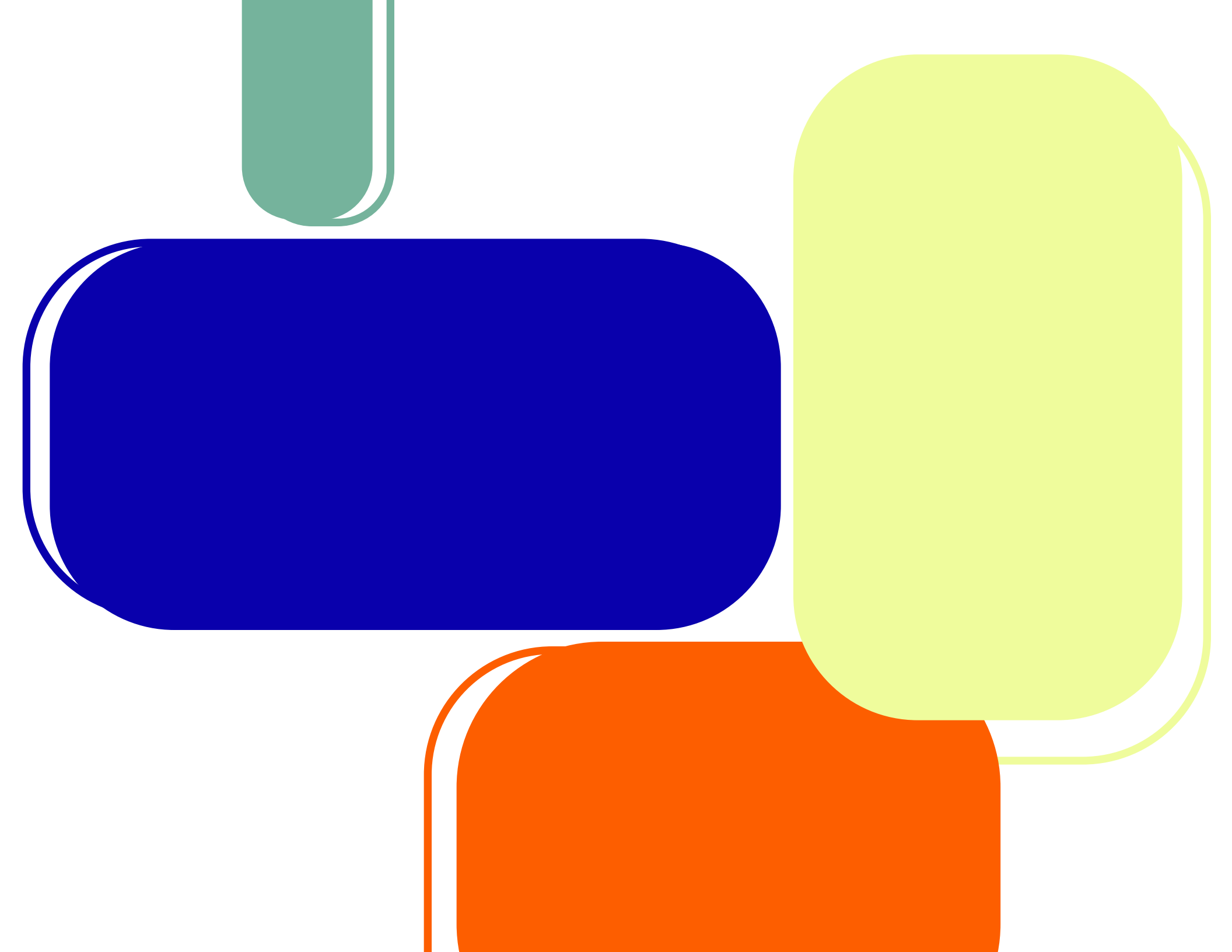
This research examines the growing use of AI technologies in social connectivity through the lens of responsible AI. The project aimed to explore the future development of AI technologies in social contexts and investigate how these technologies can be designed responsibly for young adults, with a focus on their values. The primary objective was to generate design implications that would guide designers in the responsible creation of AI technologies that foster social connectivity.

To achieve this, the research undertook a thorough review of theoretical data to, (1) define social connectivity and its significance, (2) investigate current and emerging trends in AI technologies within social contexts, and (3) examine responsibility and the relevant guidelines for responsible AI. Based on the insights gained, the project focus was refined to concentrate on the responsible development of personal AI virtual assistants that enhance interpersonal social connectivity amongst young adults. The in-depth theoretical analysis culminated in the development of the "Framework for Designing Responsible AI Virtual Assistants for Social Connectivity," which connected the explored topics and contextual values to the trends observed in AI virtual assistants and their potential social effects.

Building on these insights, the values of young adults pertinent to AI in the social domain were identified and elicited through design experiments. The research employed a combination of Speculative Critical Design and Value Sensitive Design approaches to effectively design, plan, and execute these experiments.

Ultimately, by identifying, eliciting, and defining relevant values, the research generated design implications in the form of nine value pyramids to guide responsible AI development. The research discovered nine values that young adults uphold regarding AI-fostered interpersonal connectivity: Human Agency and Oversight, Privacy and Data Governance, Transparency, Societal and Environmental Wellbeing, Personal Effort, Human-Human Connectivity, Genuity and Authenticity, Personalization, and Emotional Connectivity. These values were crucial in generating "Implications for Design," where each value was contextually defined and expanded into value principles and design norms, following Van De Poel's (2013) Hierarchy of Values. This study ultimately generated design implications intended to guide designers in the responsible creation of future AI virtual assistants that foster interpersonal social connectivity.

TABLE OF CONTENTS



1. RESEARCH INTRODUCTION

This chapter lays the foundation for the research project by outlining its context, scope, and objectives. Focusing on AI-mediated social connectivity and its potential effects on human-to-human interactions, this research aims to provide guidance on the responsible design of future AI technologies fostering social connectivity. The subsequent sections of this chapter, elaborating upon these aspects are as follows:

1.1 Introduction

"The pace of change has never been this fast before…and it will never be this slow again."

Justin Traudeau.

Recent trends and advancements in AI have permeated various aspects of life, including digital social connectivity. In our increasingly interconnected world, technological developments have enabled people from across continents to stay connected 24/7. With the integration of cutting-edge AI tools and features, such as content generation, automatic replies, emotion and speech recognition, content recommendation and personalization amongst many others, into digital platforms, communication methods are becoming more engaging, personalized, and efficient. Over the years, several leading companies have been developing and promoting these technologies to enhance connectivity and form meaningful bonds amongst people. For instance, Meta highlighted its latest features across its widely popular social media platforms, including WhatsApp, Facebook, Instagram, and Messenger, by stating: "***Technologies that bring the world closer: We are building new ways to help you explore your interests and connect with people you care about"*** (Meta, 2024). Given the rapid advancements of social digital platforms and how widely accessible they are to people, it is crucial to critically reflect on their impact as well.

While recognizing the usefulness and convenience that AI technologies bring to social connectivity, it is equally important to be aware of how they are, or may be, transforming the world in the future. In 1943, Winston Churchill notably stated, ***"We shape our buildings, and thereafter, our buildings shape us."*** Batya Friedman and David G. Hendry agreed with this but emphasized that this is true for any technology or tool (Friedman & Hendry, 2019). Acknowledging that we first design our tools and technologies, which in turn shape our lives, it is crucial not to overlook the potential consequences and effects of such developments. Emphasizing the ethical and responsible design of these innovations is imperative, requiring a critical assessment of their possible impacts on both individuals and society as a whole.

Building on the emphasis of the co-evolution of technology and social structures (Friedman & Hendry, 2019), it is essential to understand how AI-driven social connectivity may transform societal behaviors and norms in the future. As emerging technologies increasingly shape our lives, considering their potential impacts and consequences becomes imperative (Risnes et al., 2024). Gaining a comprehensive understanding of current trends in AI technologies designed to enhance social connectivity is crucial for anticipating future developments. Equally important is understanding people's needs, values, and concerns regarding these technologies to ensure their ethical and responsible development. With this knowledge, designers may be guided to create technologies that align with ethical considerations. As Sam Altman, CEO of OpenAI, articulated, ***"We want to maximize the good and minimize the bad, and for AGI to be an amplifier of humanity"*** (OpenAI, 2023). In light of this vision, it is essential to address ethical challenges, mitigating potential

negative impacts and enhance desired outcomes with a responsible design of future AI technologies fostering social connectivity.

1.2 Research Scope & Assignment

Research Scope

This research aims to investigate the increasing application of AI technologies in the domain of social connectivity from a responsible AI perspective, with the goal of generating design implications to guide designers toward the responsible development of these technologies. To achieve this, this research will: *(1) define social connectivity and clarify its relevance to this research; (2) explore current and emerging trends in the application of AI in the social domain; and (3) analyze the aspect of responsibility and the responsible AI guidelines relevant to this research.*

Building on the insights gained from the previous steps, this research seeks to *identify* and *elicit stakeholder values* pertinent to the application of AI in the social domain through design experiments. To effectively design, plan, and execute these experiments effectively, the research will employ a combination of *Speculative Critical Design* and *Value Sensitive Design* approaches. Ultimately, by *identifying, eliciting,* and *defining relevant values*, the research aims to *generate design implications* that can guide designers in responsibly developing AI technologies for social connectivity.

Research Assignment

Generate design implications by identifying and eliciting the values of young adults through design speculations, to guide designers in the responsible development of future AI technologies that foster social connectivity.

1.3 Research Questions

Building on the research scope and assignment outlined in Section 1.2, the following research questions have been formulated to guide the investigation.

(a) What are the trends in AI technologies fostering social connectivity?

(b) What values do young people uphold regarding interpersonal connectivity fostered by AI technologies?

(c) How can design implications be generated to guide a responsible design of future AI technologies fostering social connectivity?

These questions are designed to first develop a deeper understanding of the research context through a comprehensive literature review. They will then guide the design, planning, and execution of experiments to elicit relevant values, and finally, use these insights to formulate design implications. The subsequent chapters aim to address these questions, thereby advancing towards the achievement of the research goal.

1.4 Inspiration from Fictional Media

"Dreams are powerful. They are repositories of our desire. They animate the entertainment industry and drive consumption. They can blind people to reality and provide cover for political horror. But they can also inspire us to imagine that things could be radically different than they are today, and then believe we can progress toward that imaginary world."

(Dunne & Raby, 2013)

Inspiration for this research has been drawn from various sources, including contemporary communication methods, technological advancements, human behavior and psychology, as well as fictional media. This section sets the stage for this research by introducing some narratives from movies and TV series that explore human relationships with fictional technologies and their socio-cultural impacts.

The fictional world is rich with stories exploring the complex and nuanced relationships between humans and technology. These narratives portray human-technology interactions in diverse ways, from human experiences in virtual and augmented realities to forming bonds with agents and robots in the real world. Similarly, numerous examples of technology's role in fostering interpersonal relationships serve not only as engaging and entertaining narratives but also as sources of inspiration and insight. Many elements of these fictional stories appear increasingly plausible as technological advancements continue to progress.

The animated movie *"Ron's Gone Wrong,"* while seemingly a playful children's film, delves into the significant impact of technology dependence, particularly on personal robots, for forming friendships. It prompts reflection on the true value and nature of these bonds and raises questions about the true intentions of companies promoting such technologies. Similarly, *"Robot and Frank"* explores how reliance on social robots can yield unexpected outcomes, depending on individual intentions, and it challenges viewers to consider values like privacy, data collection and storage, transparency, and trust within interpersonal connections.

Finally, the TV show *"Black Mirror,"* renowned for its critical take on speculative futures, especially amongst design researchers (Dunne & Raby, 2013), provides compelling examples of how technology can have dystopian effects on society. The episode *"Nosedive"* vividly illustrates how social technologies can disrupt valuable human relationships. It

examines the harmful consequences of social media, emphasizing how these platforms create pressure to conform to societal norms and foster superficiality in online interactions. This episode serves as a stark warning about the implications of social media, highlighting how the pursuit of online validation over authenticity can lead to suppressed emotions, the maintenance of unrealistic identities, and the loss of valuable connections with friends and family.

These fictional examples illustrate the co-evolution of technology and social structures discussed in Section 1.1, highlighting the unforeseen consequences that different technologies can have on interpersonal relationships, social connectivity, norms, behaviors, and societal structures. This research draws inspiration from these narratives, focusing on advancements in AI technologies designed to enhance social connectivity. By emphasizing imaginative and critical thinking and grounding it to reality, the research aims to reflect on current AI developments in social connectivity, stressing the importance of minimizing potential harms while actively supporting and upholding human values. Consequently, these fictional examples motivate the need for this research and encourage a more reflective, ethical, and responsible approach to future technological development.

Fig 1. Fictional media inspirations illustrating the co-evolution of technology and social structures

2. THEORETICAL FOUNDATION

As outlined in Section 1.2, this research aims to investigate the growing application of AI technologies in the social domain and their potential impact on human connectivity. By adopting a responsible and ethical AI perspective, this research seeks to develop design implications that can guide the responsible development of future AI technologies aimed at enhancing social connectivity. To accomplish this, the chapter unfolds a comprehensive literature review, dissecting interconnected areas of study into three distinct topics that provide the theoretical foundation for this work.

First, this chapter explores the concept of social connectivity, defining it and clarifying its relevance to the research. Second, it examines current and emerging AI trends within the social domain. Finally, it highlights the importance of responsible and ethical AI, analyzing relevant guidelines to inform this research. The subsequent sections of this chapter are as follows:

2.1 Social Connectivity

"We are wired to be social."

Matthew Dylan Lieberman (Lieberman, 2013).

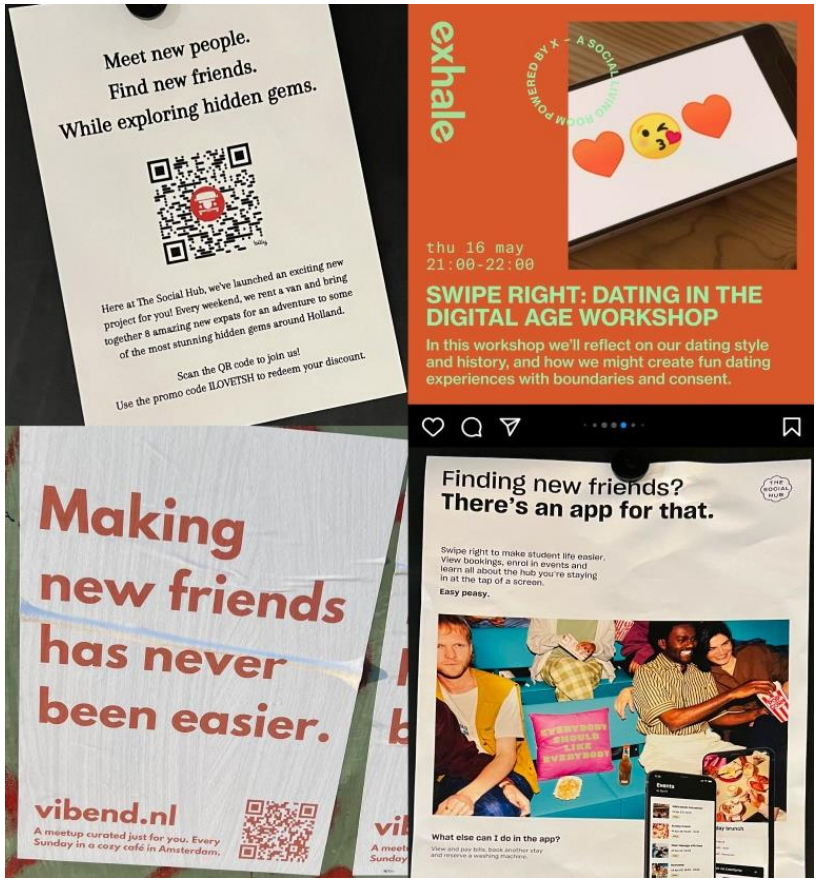


Fig 2. Posters collected from public spaces promoting digital methods for connecting with people and making friends

Social connectivity is a crucial aspect of our lives, and over the years, numerous technological advancements have emerged to support it. Digital social platforms, such as Instagram, Facebook, WhatsApp, Snapchat and many more, not only help us maintain existing connections but also enable the formation of new ones. Several other applications and digital platforms are being developed and promoted to help people form and maintain connections. *Figure 2* presents images of various posters collected from cafes, streets, hotels, and other public spaces, which promote and highlight some digital methods for connecting with people and making friends. The promotion and encouragement of building and maintaining connections through various digital means underscores the significance of social connectivity in our lives. Recent advancements in AI technologies and their integration across various aspects of life have also impacted the domain of social connectivity. These developments, which are discussed in detail later in this chapter in *Section 2.2*, are significantly transforming how we interact and connect with others.

Social connectivity is a broad term that encompasses various types of connections with different people. Therefore, it is essential to establish a clear understanding of "social connectivity" for this research. When discussing AI technologies that foster social connectivity, it is important to specify the particular aspects being addressed. To provide clarity and help refine the research context, objectives, and outcomes, the following sections will define the specific type of social connectivity that this research will focus on.

2.1.1 Social Connectivity and Its Significance

Although there is no consensus among the researchers on what social connectivity really means (O'Rourke & Sidani, 2017), there is mutual understanding on it being a basic human need (O'Rourke & Sidani, 2017; Haslam et al., 2015). According to Maslow's Hierarchy of Needs, "Love and Belonging" occupies the third level of the pyramid, highlighting its crucial role in human life. This group of needs includes friendship, intimacy, trust, acceptance, receiving and giving affection, and love (Pichère & Cadiat, 2015). Similarly, researchers have defined social connectivity as the experience of feeling close to other people. It encompasses caring for others, feeling cared for, loved, and valued, as well as experiencing a sense of belonging to a community or group. These feelings are fundamental to the formation and maintenance of interpersonal relationships (O'Rourke & Sidani, 2017; Haslam et al., 2015; Eisenberger & Cole, 2012).

In his book "*Why Our Brains Are Wired to Connect*", Matthew D. Lieberman, a neuroscientist and Professor at UCLA's Department of Psychology, Psychiatry, and Biobehavioral Sciences, underscores the intrinsic social nature of humans. Lieberman argues that people are naturally inclined to be curious about others' lives and are driven by a fundamental need to stay connected with family and friends. These connections and the sense of belonging to a group or community provide the values that shape our individual identities (Lieberman, 2013).

An important aspect of social connectivity is its "subjective" nature, which means it relies significantly on an individual's personal perception of their connections. This interplay between actual social interactions and how they are ultimately perceived to influence an individual's understanding of social connectivity contributes to the complexity of its definition (Dias et al., 2018). Additionally, social connectivity is influenced by an individual's prior experiences, preferences, and personality traits (Cacioppo, 2006).

Thus, it is established that social connectivity is a fundamental human need, inherently subjective and personal. This understanding has driven a multitude of technological developments in this domain, particularly with the advent of AI. The continuously emerging AI features and functionalities that facilitate social connectivity are not only meeting the essential human desire to connect but also providing users with the ability to customize and personalize their interactions. These advancements empower individuals to tailor their social experiences according to their unique needs and definitions of connectivity.

2.1.2 Pathways to Social Connectivity

By establishing an understanding of the significance of social connectivity, this research investigates the various ways in which it is experienced. These diverse experiences inform the design, development, and implementation of AI technologies

within digital social platforms, aiming to meet users' needs and provide a personalized experience.

The three distinct layers in which social connectivity is experienced, as identified in the "*Journal of Positive Psychology*", are Intimate Connectedness, Relational Connectedness, and Collective Connectedness (Hawkley, 2005; Stavrova & Luhmann, 2015). Each layer is defined as follows:

Fig 3. Pathways to Social Connectivity: Intimate Connectedness, Relational Connectedness, and Collective Connectedness (Hawkley, 2005; Stavrova & Luhmann, 2015)

1. Intimate Social Connectedness

This layer of social connectivity refers to the perceived closeness to a nurturing companion who affirms our values as individuals. Such companions may include a spouse, romantic partner, or anyone with whom we share a deep, mutual bond of trust and affection. These inner-circle relationships are vital for mutual assistance, encompassing both emotional and practical support during crises. Research indicates that people typically allocate approximately 40% of their available social time to their five most intimate social connections (Sutcliffe et al. 2012; Dunbar, 2014; Stavrova & Luhmann 2015).

2. Relational Social Connectedness

This layer of social connectivity pertains to the perceived presence of friendships or family connections that offer support and mutual aid. These interpersonal relationships, typically encompassing 15 to 50 individuals within our middle social circle, are less intimate than those in our innermost circle but still provide a sense of closeness and substantial support (Dunbar, 2014; Cacioppo, 2015).

3. Collective Social Connectedness

This layer of social connectivity refers to the perceived meaningful connection with a larger group of individuals, such as teams, volunteer groups, schools, and organizations. Engaging with a network or community through common interests or a shared sense of purpose fosters a sense of belonging to something greater than one's own self (Cacioppo, 2015). This dimension of social connectedness typically corresponds to the outermost social circle, encompassing anywhere from 150 to 1500 individuals (Dunbar, 2014).

Understanding the diverse pathways through which social connectivity is experienced allows for a deeper insight into people's needs and expectations across their personal, professional, and community relationships. This comprehension is crucial for leveraging advancing AI technologies in digital social platforms, as it helps address the specific needs associated with each layer of social connectivity. In order to further refine the research context and explore relevant AI trends, this research aims to specify a particular layer of social connectivity in the following section.

2.1.3 Research Definition of Social Connectivity

Building on an understanding of social connectivity—its significance in human life and the various ways it can be experienced—this research aims to define its relevance within this specific context. By examining AI technology trends, current social media usage, and the roles these platforms play in people's lives, the research seeks to reflect on these developments to guide the responsible design of AI technologies for social connectivity. To achieve this, the research focuses on understanding the values of young adults concerning social connectivity, specifically emphasizing the second layer of social connectivity: *"Relational Connectedness."*

This emphasis is driven by the values associated with social connections that involve emotional and personal bonds, such as close friendships and relationships with family members. To sustain these connections, people frequently utilize social media platforms and are benefitting from AI features embedded within them. These platforms facilitate social connectivity in many ways, such as maintaining relationships with distant friends across time zones, staying in touch with family members, coordinating activities with close friends, or connecting with individuals with shared interests. As new AI features are developed and integrated into these widely used platforms, this research aims to understand what is important to individuals with such social connections. Ultimately, the refined goal is to guide the responsible design of future AI technologies that enhance these vital interpersonal connections.

Specifically concerning *"Interpersonal Connectivity"*, this research aims to explore relevant AI technology trends in this domain, as detailed in the following section.

2.2 AI Technology Trends

Understanding the trends and advancements in AI technologies for social connectivity is crucial for this research. To guide a responsible design of future AI technologies, it is essential to comprehend their ongoing developments and anticipated evolution. While looking into the general trends, an emphasis is laid on the progressive AI features and functionalities that are being embedded especially within social platforms.

The powerful capabilities of AI features and technologies implemented in social platforms are significantly transforming how we interact and connect digitally. Several large and small enterprises are spearheading the integration and implementation of AI technologies, transforming methods of social connectivity. Leading companies in this domain, often referred to as tech giants, such as Alphabet (Google's parent company), Apple, Microsoft, Meta, Amazon, and the emerging player OpenAI, are consistently setting new benchmarks. These companies are continually researching, testing, launching, and refining increasingly powerful AI features and tools integrated within their products. By doing so, they are revolutionizing our methods of connection. AI developments and functionalities are central to their innovations, consistently helping them outpace competitors and enhance experiences for their users. To understand current and anticipate future trends in AI for social connectivity, it is essential not only to analyze general developments in AI technologies but also to examine the latest features and newly launched tools created by these leading companies and made available to the public. This section provides a deeper understanding of AI trends relevant to the research context, ultimately refining the specific trends that this research will focus on.

2.2.1 General AI Trends

In recent years, there have been groundbreaking advancements in various subfields of AI. After an extensive review of the literature, this research identified advancements in *Natural Language Processing (NLP)* as among the most notable developments. These developments have driven the creation of numerous AI models and introduced various functionalities and capabilities to the world of social connectivity. Additionally, such developments in NLP have significantly informed the growth of different AI models comprising *Conversational AI (CAI)* and *Generative AI (GAI)*, which are being actively integrated into social platforms to foster social connectivity. This section of the research report delves into these advancements to better understand the specific AI trends within the social domain.

2.2.1.1 Advancements in Natural Language Processing (NLP)

Google defines *Natural Language Processing (NLP)* as ***"a field of AI that allows computers to understand and process human language."*** NLP further branches out into *Large Language Models (LLMs)* and *Small Language Models (SLMs)*, which have seen significant advancements in recent years aimed at enhancing their overall performance. Amongst the popular examples of powerful LLMs stands Llama 3, which was introduced by Meta as ***"the most capable openly available LLM to date"*** (MetaAI, 2024). With its launch, Llama 3 was made available for a broad range of use cases due to the nature of parameters it was pre-trained on, making it one of the best open-source models of its class (MetaAI, 2024). It offers new capabilities, including improved reasoning and was put into the hands of the community, allowing people to build new applications, tools, products and much more.

Alongside the development of Large Language Models (LLMs), Microsoft aimed to further enhance AI accessibility by focusing on Small Language Models (SLMs) as well. Although smaller than LLMs, but still quite large with several billion parameters, SLMs are small enough to run offline on a phone (Ho, 2024). According to Sebastian Bubeck, leader of the Machine Learning Foundations group at Microsoft Research, ***"Small language models can make AI more accessible due to their size and affordability. Simultaneously, we are discovering new ways to make them as powerful as large language models."*** Microsoft researchers have already developed and released two SLMs – Orca and Phi – challenging the notion of scale requirement for good performance and assuring that in recent years, it is expected to see even more improved language models (Ho, 2024).
Such developments in Large Language Models (LLMs) and Small Language Models (SLMs) within the AI domain of Natural Language Processing are enabling new functionalities, such as Conversational AI (CAI) and Generative AI (GAI), that are transforming how we connect. The following sections will explore the integration of these AI advancements into digital platforms and social media, highlighting the new features and functionalities they offer for widespread use.

2.2.1.2 Conversational and Generative AI (CAI & GAI) Models

This research highlights advancements in the AI branch of Natural Language Processing (NLP), which have led to the emergence of new Conversational and Generative AI (CAI and GAI) models. This section analyzes some of the trending models, currently being developed and launched by leading companies; to provide a deeper understanding of the features and functionalities they offer within the social domain.

This research notes the development of several AI models capable of processing images, videos, and audios, in addition to textual inputs and outputs, outlining a common trend of AI designed to be multimodal. Kevin Scott, the principal engineer for Microsoft's Office of the Chief Technology Officer emphasized upon the significance of developing multimodal AI by stating, ***"Multimodality has the power to create more human-like experiences that can better take advantage of the range of senses we use as humans, such as sight, speech and hearing"*** **(Ho, 2024)**

By launching pre-trained models like ChatGPT and GPT-4, OpenAI introduced its API, which possesses the abilities of comprehending and generating *human-like* text (DataCamp, n.d.). With a broad domain expertise and general knowledge, GPT-4 was designed to follow complicated instructions in natural language and is capable of accurately solving difficult problems (DataCamp, n.d.; OpenAI, n.d.). It is also collaborative and creative, capable of learning a user's writing style as well as editing, generating, and iterating large amounts of text (DataCamp, n.d.). It also accepts visual input and can generate analysis, classifications, and captions for the input in the form of images (DataCamp, n.d.; OpenAI, n.d.).

Similarly, Open AI's *DALL·E* was designed to create realistic, original images and artwork from natural language text inputs, allowing it to combine concepts, styles, and attributes in its creations. In addition to these models, OpenAI developed *Whisper*, a versatile speech recognition model capable of

identifying, transcribing, and translating multiple languages (; OpenAI, n.d.). Furthermore, the development of *Sora*, OpenAI's text-to-video model, also represents a significant advancement in AI capabilities. This model can generate videos up to one minute long while adhering to given prompts and maintaining visual quality. OpenAI aims to enhance *Sora's* performance to address problems requiring real-world interaction, training AI to understand and simulate physical motion in dynamic environments (OpenAI, 2024).

Similarly, alongside other AI models, Meta's *Voicebox* introduces advanced capabilities such as audio sampling, noise reduction, editing (e.g., removing unwanted noise), and audio stylization (Meta, 2023). This sophisticated AI model not only excels in generating high-quality audio clips but also in editing pre-recorded audio while preserving the original style and content. For example, if a speech is disrupted by background noise, *Voicebox* can seamlessly regenerate the affected segment without the need for re-recording. Furthermore, *Voicebox* supports multiple languages and is currently capable of producing speech in six different languages (Meta, 2023). Beyond speech generation, *Voicebox* also features in-context text-to-speech synthesis, allowing it to match the audio style of an existing clip and generate text-to-speech outputs with an audio sample as short as two seconds (Meta, 2023).

The development of such AI models and their intended functionalities underscores three key trends central to this research. Through an analysis of various models, including those mentioned above, this research identifies that AI technologies are increasingly being designed to be: *Multimodal, Multipurpose, and Humanlike*. These specific AI technology trends are described as follows:

1. Multimodal AI

Companies are increasingly emphasizing the importance of multimodality in AI. AI systems are being designed to handle multiple data types, including text, audio, images, and videos, allowing for more versatile and comprehensive processing capabilities.

2. Multipurpose AI

The versatility of AI technologies has been significantly advanced through the development of various models. These AI models are being designed to perform multiple functions and provide users with a range of features, thereby enhancing their utility and adaptability across different applications.

3. Humanlike AI

AI models are increasingly being designed to exhibit human-like qualities, setting a new standard for AI performance. This trend emphasizes that the more natural and human-like an AI is, the more effectively it can fulfill its intended functions (Ho, 2024). Current developments focus on endowing AI models with characteristics such as distinct personalities, natural-sounding responses, and the ability to express opinions and interests. These enhancements aim to make interactions with AI more engaging and enjoyable, while also supporting users in decision-making, offering recommendations, answering queries, and providing various forms of assistance (Meta, n.d.; Meta, 2023).

By analyzing recent AI models and ongoing advancements, this research gains insight into current developments in the field. Building on this understanding, it aims to explore the specific applications of these trends within the social domain. The following section explores how emerging AI features are enhancing social connectivity and facilitating interactions among people.

2.2.1.3 Transformation of Personal AI Virtual Assistants

As a result of analyzing general AI trends and advancements in Natural Language Processing (NLP), which have led to the development of various Conversational and Generative AI (CAI and GAI) models, this research identifies three primary categories of AI trends*: (1) Multimodal AI, (2) Multipurpose AI,* and *(3) Humanlike AI*, as explained previously in Section 3.2.1.2. In reviewing the practical applications of these AI models within the social domain, this research identifies trends in the development of AI-powered personalized virtual assistants. A prevalent and emerging application of recent AI advancements is the development of conversational agents that interact with users in an intuitive and natural manner (DataCamp, n.d.). These AI-powered virtual assistants are increasingly integrated into social platforms, providing users with innovative ways to enhance their connections.

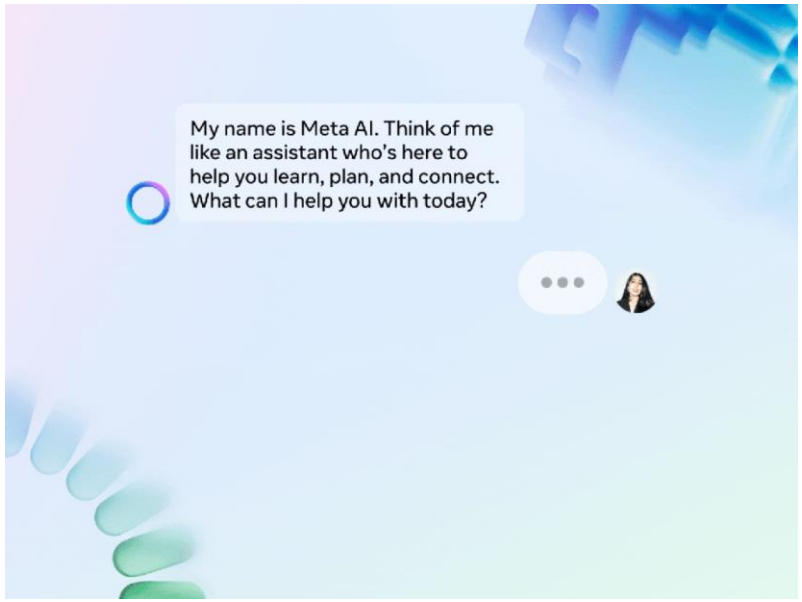
Companies are exploring the creation of multifunctional AI chatbots with unique features by leveraging advancements in trending AI models that offer multimodality, multipurpose functionality, and human-like characteristics. Leading technology companies (as noted in Section 2.2) are currently making substantial investments in the development of intelligent virtual assistants (DataCamp, n.d.). Considering these developments, several analysts are foreseeing a *'post-app era'* with virtual personal assistants paving the way. In this post-app era, people will tend to use only a few prevalent applications (e.g. messaging platforms) empowered by digital assistants that will integrate a multitude of functions (Research and Markets, 2018; Gaggioli, 2018). These personal AI virtual assistants are foreseen to be embedded in everyday environments and contexts that will make them more accessible to users (Research and Markets, 2018; Gaggioli, 2018). Open AI is facilitating this innovation with the development of its powerful Application Programming Interfaces (APIs) (DataCamp, n.d.; OpenAI, n.d.). According to OpenAI, its models hold great potential and are a prime candidate for creating intelligent virtual assistants and chatbots. Similarly, in promoting multimodality in AI, Microsoft demonstrated its implementation through the recent developments in its chatbot, Microsoft Copilot, which is now capable of processing not only text and search data but also audio and images (Ho, 2024).

Focusing on the development of AI virtual assistants aimed at enhancing social connectivity, Meta is set to spearhead a new era of personal virtual assistants within the social domain. Built with its powerful LLM called Llama 3, Meta recently launched its virtual assistant called Meta AI. Offering numerous functionalities, Meta AI was introduced to the world as ***"the most intelligent assistant you can use for free."*** This assistant can now be used on Meta's social applications such as Facebook, Instagram, WhatsApp, and

Messenger for various purposes such as creating, connecting, learning, or getting things done (Meta, 2024).

The features and functionalities of Meta AI are highly relevant to this research, as this virtual assistant exemplifies both the latest AI trends and their specific applications within the social domain. To gain a deeper understanding of the new functionalities provided by Meta AI, this research examines it as a case study, focusing specifically on AI trends in social connectivity.

2.2.2 Personal AI Virtual Assistants Fostering Social Connectivity: Meta AI – A Case Study



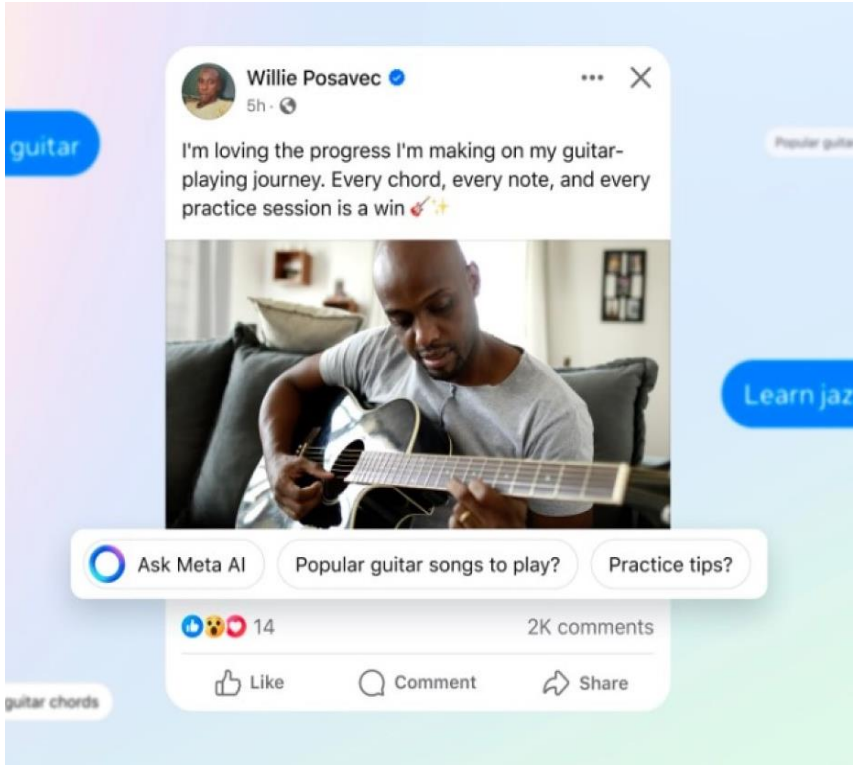
By seamlessly incorporating Meta AI across its digital social platforms, Meta has established new paradigms for fostering social connections. In this realm, Meta, being a multinational technology conglomerate and ranking among some of the largest information technology companies, is already setting new examples of latest AI technology collaborations within existing social platforms. With the launch of its new assistant, Meta AI, Meta has enabled a plethora of new AI assisted functionalities within its range of applications. Exploring the functionalities of Meta AI provides insights into emerging AI trends within social applications. This analysis not only sheds light on the current state of AI technologies fostering social connectivity but also offers a glimpse into the future direction of these advancements.

While Meta AI provides a wide range of functions and features, this research specifically examines its role in fostering *Relational Connectedness* (Section 2.1.3) with interpersonal connections, such as those with friends and family. Through an in-depth analysis of recent updates and articles published by Meta, along with a review of relevant literature, this research identifies several key functions of Meta AI that exemplify emerging AI trends within the social domain. The relevant characteristics of Meta AI's functionality, as they pertain to the context of this research, are outlined below.

1) Accessibility:

Meta AI offers users access to real-time information from across the web without requiring them to switch between apps. This feature is integrated into the *"search"* function across platforms like WhatsApp, Facebook, Instagram, and Messenger. Additionally, Meta AI is embedded within social

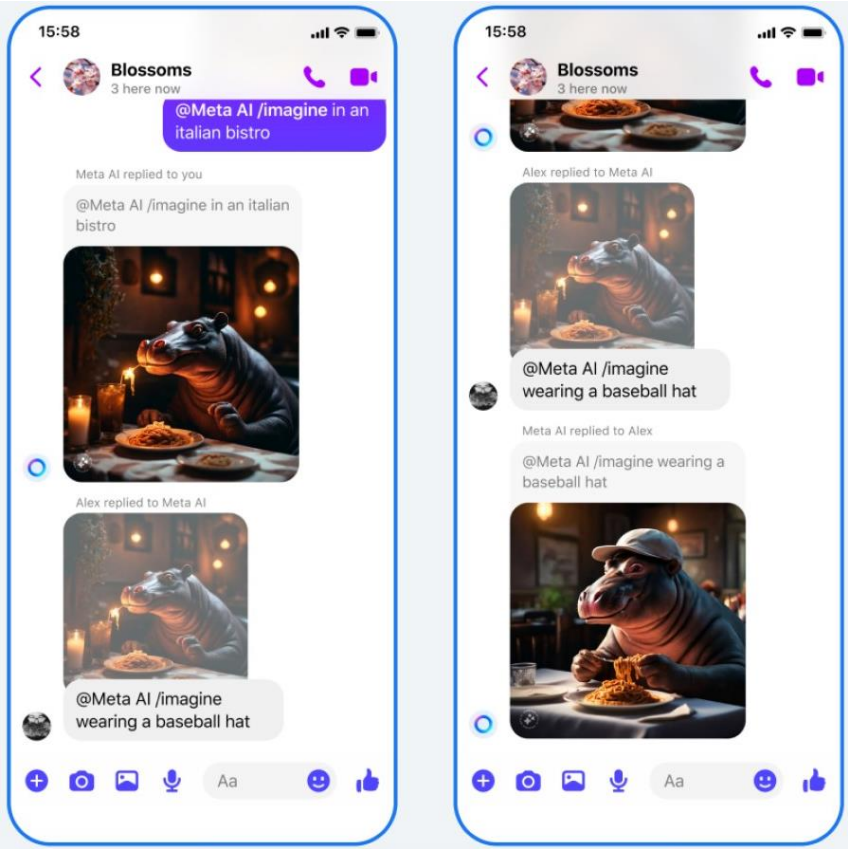
media feeds, enabling users to directly request more information on posts they find interesting. This seamless in-app integration allows users to ***"make Meta AI work for them"*** in various contexts (Meta, 2024). For example, when planning a night out with friends, users can ask Meta AI to recommend restaurants that meet specific criteria, such as offering vegan options or having sunset views. Similarly, Meta AI can assist in organizing activities by searching for events like concerts that match users' interests.



2) Content Generation:

Using Meta AI's *"imagine"* feature, users can now create images from text in real time. This functionality enables users to ask Meta AI to generate images in various styles, animate those images, create multiple iterations, or even turn them into GIFs for sharing with friends. The image generation feature also supports image transformations and allows for collaborative creation of AI-generated images with friends, introducing new avenues for entertainment, creativity, and self-expression (Meta, 2024). By introducing these features, Meta emphasizes the importance of creativity and imagination in interactions with friends, aligning with the company's vision of enhancing social experiences through creative collaboration and expression. Meta introduced these features by stating: ***"Now you can kick images back and forth, having a laugh as you try to one-up each other with increasingly wild ideas."*** (Meta, 2023)

Furthermore, Meta has also equipped Meta AI with the ability to share specific reels in chats upon users' requests (Meta, 2023). In addition to visual content generation, Meta AI allows users to generate text for various purposes, such as crafting birthday wishes, drafting introductions for personal profiles, or editing feed posts. This feature not only produces the requested text based on the user's prompt but also provides alternative options, allows for modifications, and enables users to select the tone of the message, enhancing personalization and flexibility (Meta, 2023).

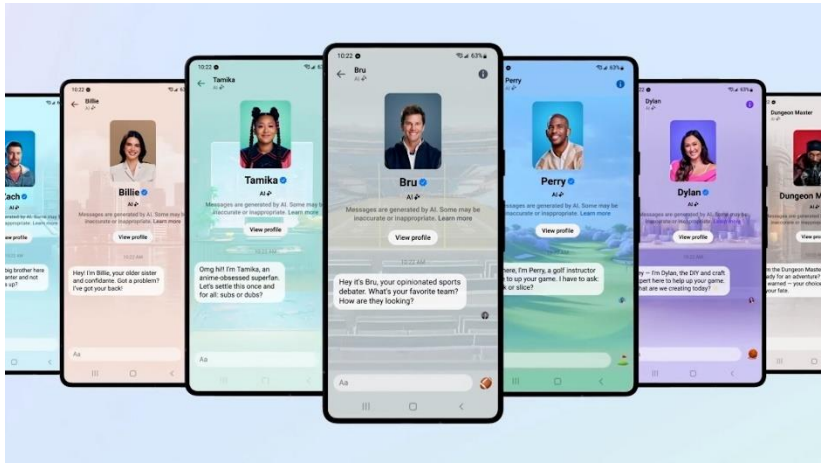


3) Personalized & Human-like Assistance

Meta AI has been developed to enable users to interact with it as if it were a person. It is designed with distinct personalities, interests, and opinions, making interactions more engaging and enjoyable for users (Meta, n.d.; Meta, 2023). Meta has also created 28 additional AIs for users to interact with, each featuring unique backstories. These AIs have been developed in collaboration with celebrities, influencers, and cultural icons, who contribute to bringing these personalities to life.

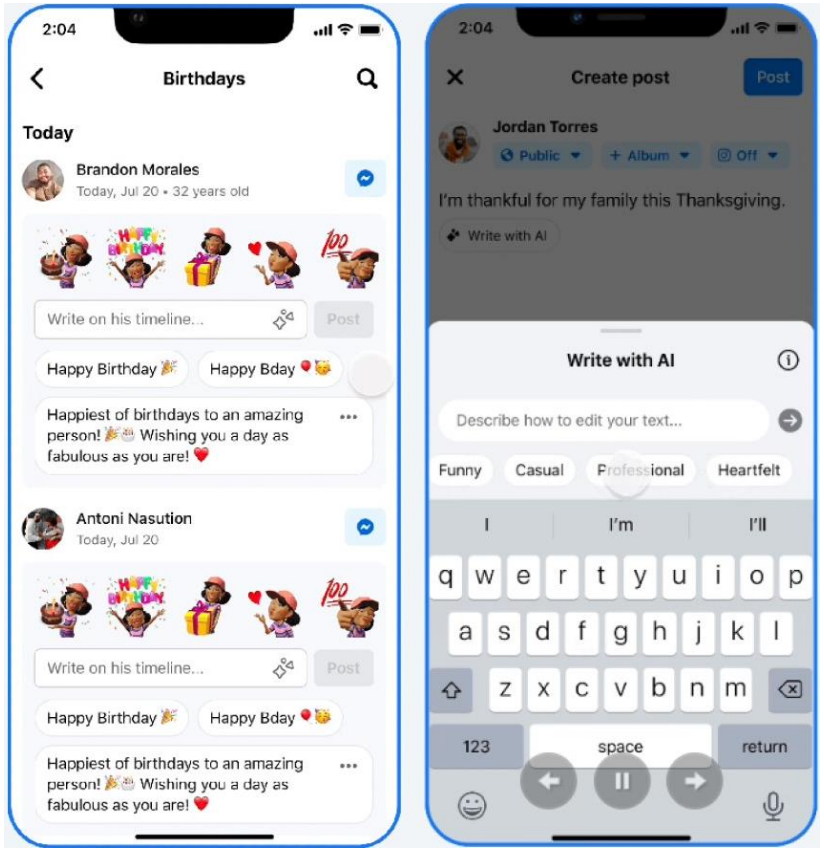
Amongst its social applications, alongside Meta AI, Meta has 28 other AIs that people can chat with. Each of these AIs have been designed to give a feeling of familiarity when talking to them and have their unique backstories. To build this even further, Meta partnered with several influencers, celebrities, and cultural icons to embody and play some of these AI. Each of them has their profiles on Facebook and Instagram where people can explore what they are all about. Some such examples are Kendall Jenner as Billie, No-BS, ride-or-die companion, MrBeast as Zach, The big brother who will roast you - because he cares, Roy Choi as Max, Seasoned sous chef for culinary tips and tricks, Sam Kerr as Sally, Free-spirited friend who will tell you when to take a deep breath and many more. These AI personalities are showcased with detailed profiles, including descriptions such as "MMA Expert," "Creative Writing Partner," or "Travel Expert," allowing users to easily choose the AI that best suits their interests or needs. [(Meta, 2024; Meta, 2023).

Meta is also advancing personalization through the development of a *"sandbox"*, allowing users to create their own custom AI models (Meta, 2023). Furthermore, Meta is experimenting with incorporating long-term memory into its AI systems. This feature will enable AIs to retain conversation histories, allowing users to resume interactions where they left off and facilitating more meaningful and continuous dialogue (Meta, 2023).



4) Enhanced Communication

Meta AI is enhancing communication by highlighting important conversations and ensuring they remain visible, particularly in busy group chats. It also assists users in planning and organizing activities by providing personalized recommendations and creative ideas directly within their chats (Meta, 2024; Meta, 2023). For new conversations, Meta AI can suggest engaging topics to keep discussions interesting and dynamic, while overall enhancing communication between users (Meta, 2024).



2.2.3 Key Insights

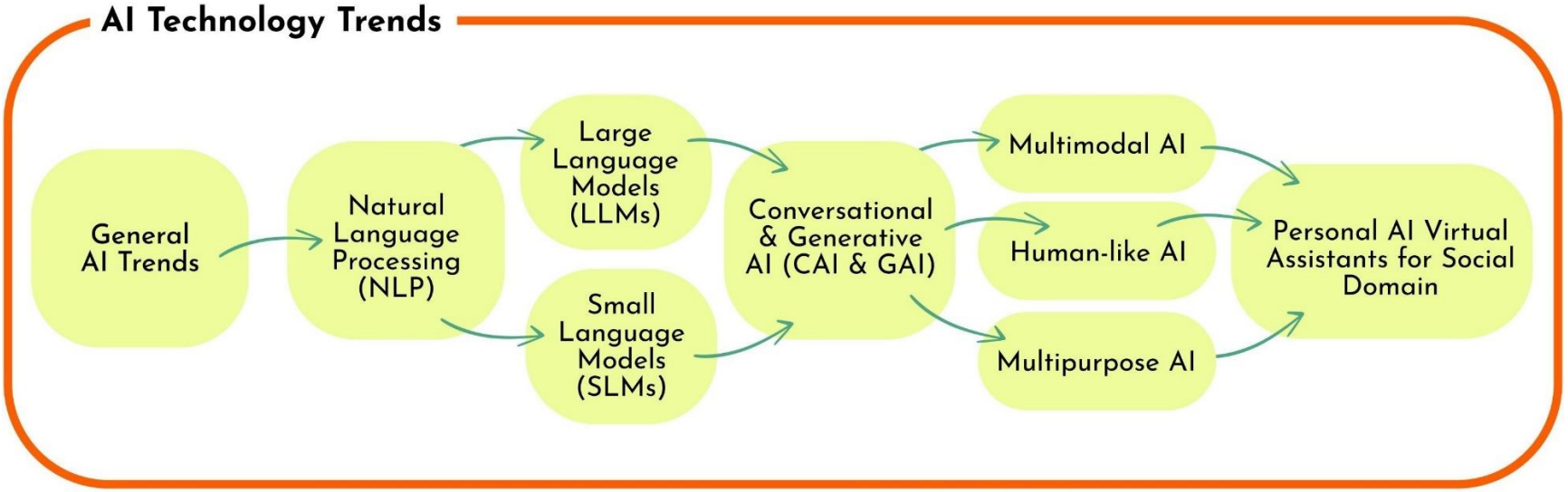


Fig. 4 Summarized and simplified representation of AI technology trends observed in Section 2.2

Through an in-depth analysis of current AI technology trends, this research highlights the transformation of personal AI virtual assistants and their significant impact on the social domain. The AI trends driving the development of AI virtual assistants, as analyzed in the previous sections, are summarized and simplified in Figure 4. Driven by advancements in various AI models, AI virtual assistants are revolutionizing communication methods, making their development crucial to this research. Thus, this research narrows its focus to *the role of future AI virtual assistants in enhancing interpersonal social connectivity*.

This section concludes that trends in various AI models are driving their development toward being multimodal, multipurpose, and human-like. These trends are being further adapted and integrated into the creation of intelligent virtual assistants, which are increasingly utilized for interpersonal social connectivity. After a comprehensive analysis of the features and functions currently offered by AI virtual assistants in the social domain, this research refines its focus within the research context and draws its attention to the following conclusions.

Personal AI virtual assistants in social domain are designed to be:

1. Multipurpose, offering various functionalities
2. Multimodal, processing various forms of data
3. Humanlike, having characters, personalities, opinions, long term memory
4. Self-learning, autonomously improving performance without human intervention
5. Guiding, influencing and supporting decision-making process and making decisions
6. Generating content
7. Training on personal data
8. Simulating social interaction between people

2.3 Responsible AI

"What is better and for whom (or what)? How to evaluate what's better? And how to deal with unforeseen consequences or unwanted side-effects of design interventions?"

(Eggink et al., 2020)

By refining the concept of *"Social Connectivity"* within the context of this research and examining the related "*AI Technology Trends"* in that domain, this study advances towards understanding Responsible AI. Through an analysis of Responsible AI, the research aims to establish a clear definition of what Responsible AI entails within this specific context.

With the rapid expansion of AI and its integration across various sectors, the urgency to design AI technologies in a responsible and ethical manner has never been greater. This necessitates not only individual awareness but also coordinated efforts at the governmental level to formulate policies that regulate the development and application of AI technologies effectively, ethically and responsibly.

2.3.1 Three Levels of Responsibility

To understand the current efforts aimed at guiding the responsible design of AI technologies, this research, after analyzing relevant literature, dissects the concept of *"Responsible AI"* into three distinct levels. *First*, it examines responsibility at the individual level, focusing on the role of designers in creating AI ethically and responsibly. *Second,* it considers responsibility at the organizational level, highlighting the role of companies in developing principles that ensure responsible AI design. *Third*, it explores responsibility at the governmental level, emphasizing the role of policymakers in establishing policies and frameworks that guide the ethical development and application of AI.

1) Responsibility of Designers:

The first layer underscores the critical role individual designers play in considering the ethical implications and impacts of their work. Designers today are not only applying their skills to address social issues but are also increasingly driven ***"to do good for the society"*** (Tromp et al., 2011). Designers hold a significant responsibility towards society, as their creations inevitably affect the lives of many. Given that design is inherently future-oriented, this responsibility extends to future generations as well. Introducing new services and products into the world requires designers to be acutely aware of the potential consequences of their activities. This mindfulness ensures that their innovations contribute positively to society and align with ethical standards, safeguarding the well-being of both current and future users (Eggink et al., 2020).

The field of design fosters awareness by encouraging reflection on ethical considerations. Designers can employ various approaches, each with distinct goals and ideologies, yet all sharing a common emphasis on responsibility as a core design principle (Eggink et al., 2020). Critical design serves as an example of this by showing radical alternatives for common practices. It challenges assumptions and explores the implications of design choices (Malpass, 2010). Similarly, for interaction technology, value-sensitive design integrates ethical values into the design process, where the answer to ***"what is right?"*** should be understood from an ethical standpoint (Friedmann, 1996). Other methodologies and approaches include but are not limited to; design futuring approaches such as speculative design, which imagines future scenarios to provoke reflection and debate; responsible design, which prioritizes ethical responsibility and societal impact; and participatory design, which incorporates stakeholders in the development process for a consideration of all consequences of design (Ehn, 2008). By employing such approaches, designers can contribute to the creation of AI technologies that are ethically sound and socially beneficial.

2) Responsibility of Companies:

"AGI (artificial general intelligence) would also come with serious risk of misuse, drastic accidents, and societal disruption. Because the upside of AGI is so great, we do not believe it is possible or desirable for society to stop its development forever; instead, society and the developers of AGI have to figure out how to get it right."

Sam Altman, CEO of Open AI (OpenAI, 2023).

The second level emphasizing responsible AI, as defined in this project, focuses on the steps taken by technology companies in being ethical and responsible. This research emphasizes that companies engaged in the research, development, and implementation of new technologies, while striving to make them accessible to the wider public, also bear the responsibility of ensuring that their products and services are trustworthy and ethical. Although it is impossible to precisely predict the future consequences of AI developments, companies are striving to maximize benefits and minimize adverse effects, aiming for AI to serve as an amplifier of humanity. Recognizing the importance of managing significant risks, technology companies are dedicating substantial efforts to making AI more responsible (OpenAI, 2023).

This research examines the various approaches companies have taken towards responsible AI. For instance, Microsoft articulated its comprehensive list of general requirements in the *"Microsoft Responsible AI Standard, v2 General Requirements"* (Microsoft, 2022). OpenAI outlined its principles through the *"OpenAI Charter"* (OpenAI, n.d.) and *"Product Safety Standards"* (OpenAI, n.d.). Similarly, Meta made its *"Meta Llama Responsible Use Guide"*(Meta, n.d.) publicly available and established its core values through *"Meta's five pillars of responsible AI"* (Meta, n.d.). In light of recent advancements in Generative AI, companies have also been particularly focused on implementing newly designed features responsibly (Meta, 2023).

Such efforts suggest that numerous companies are actively addressing ethical concerns related to responsible AI. While the specific goals of each company may vary—ranging from minimizing harm and building trust to learning and iterating or reducing risks of misuse (OpenAI, n.d.; OpenAI, 2023)—their efforts converge on some common *value considerations*. The general principles for responsible AI, as summarized through the literature analysis discussed above, include Accountability, *Transparency, Fairness, Reliability, Privacy, Safety, Security, Inclusiveness, Interpretability, and Trust, among others* (Microsoft, 2022; OpenAI, 2024.; Meta, n.d.; Meta, 2023; Google, 2023.).

3) Responsibility of Policymakers:

The third level of responsibility in AI technology development, as defined by this research, underscores the efforts of policymakers in addressing relevant ethical concerns. This layer extends to the governmental level, where policies and regulations are crafted to safeguard user interests and ensure ethical AI practices.

With a focus on the European Union market, this research dives deeper into the efforts made by the European Commission in drafting the *"Ethics Guidelines for Trustworthy AI"* (European, 2019). This comprehensive document not only outlines seven key requirements that AI systems must meet to be considered trustworthy but also introduces a specific assessment list called *"Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment"* (European, 2020), designed to verify the implementation of each requirement. These efforts by the EU Commission have garnered significant recognition, especially within Europe, and have been frequently referenced. Since their release, these guidelines have served as a comprehensive tool for companies and individuals to assess the ethical implications of AI products. The seven key criteria for AI to be considered trustworthy, as outlined in the *"Ethics Guidelines for Trustworthy AI,"* are:

1. Human Agency and Oversight
2. Technical Robustness and Safety
3. Privacy and Data Governance
4. Transparency
5. Diversity, Non-discrimination and Fairness
6. Environmental and Societal Well-being
7. Accountability

2.3.2 Understanding Responsible AI Within the Scope of This Research

By developing a general understanding of *Responsible AI* and analyzing ways in which it can be achieved and implemented, the research proceeds to define what *Responsible AI* specifically means within the context of this study.

By comprehending the role of individual designers, this research explored design approaches and methodologies for designers, allowing them to involve critical thinking, stakeholder participation, and reflective practices. These approaches provide valuable insights that inform the transition to the design phase of this research (Chapter 4), offering guidance for the responsible design of AI technologies. By examining the roles of companies and policymakers, this research identified several criteria, standards, and overarching approaches for implementing responsible AI on a larger scale. Several common values were noted throughout the different guidelines, charters, principles or requirements by both companies as well as policy makers, as analyzed above in Section 2.3.1. To explore the application of these values, this research further examines the *'Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment,'* hereafter referred to as *ALTAI*.

ALTAI, discussed in Section 2.3.1 under *'Responsibility of Policy Makers,'* outlines the implementation of key requirements from the European Commission's *'Ethics Guidelines for Trustworthy AI.'* Its elaborate and actionable framework makes it an effective tool for evaluating responsible AI practices. Among the seven values presented, this research highlights the relevance and significance of four key values within its specific context. These four values were selected based on observations of AI trends in social connectivity (as outlined through this research) and their potential impact on interpersonal relationships, as outlined in ALTAI. The selected values are:

1. Human Agency and Oversight
2. Privacy and Data Governance
3. Transparency
4. Societal and Environmental Wellbeing

Value Selection Criteria

ALTAI provides a framework of questions designed to assess the responsibility of AI systems, along with explanations of the values' significance and their potential impacts. This research identified and selected questions from ALTAI that are directly applicable to the observed AI trends in virtual assistants. Additionally, the research focused on values associated with impacts specific to social connectivity and human interactions. By aligning the selected values with these impacts and the relevant questions from ALTAI, this research ensures that the values chosen are highly pertinent to the context of social connectivity and human relationships. An example of this selection process is illustrated below in Figure XX.

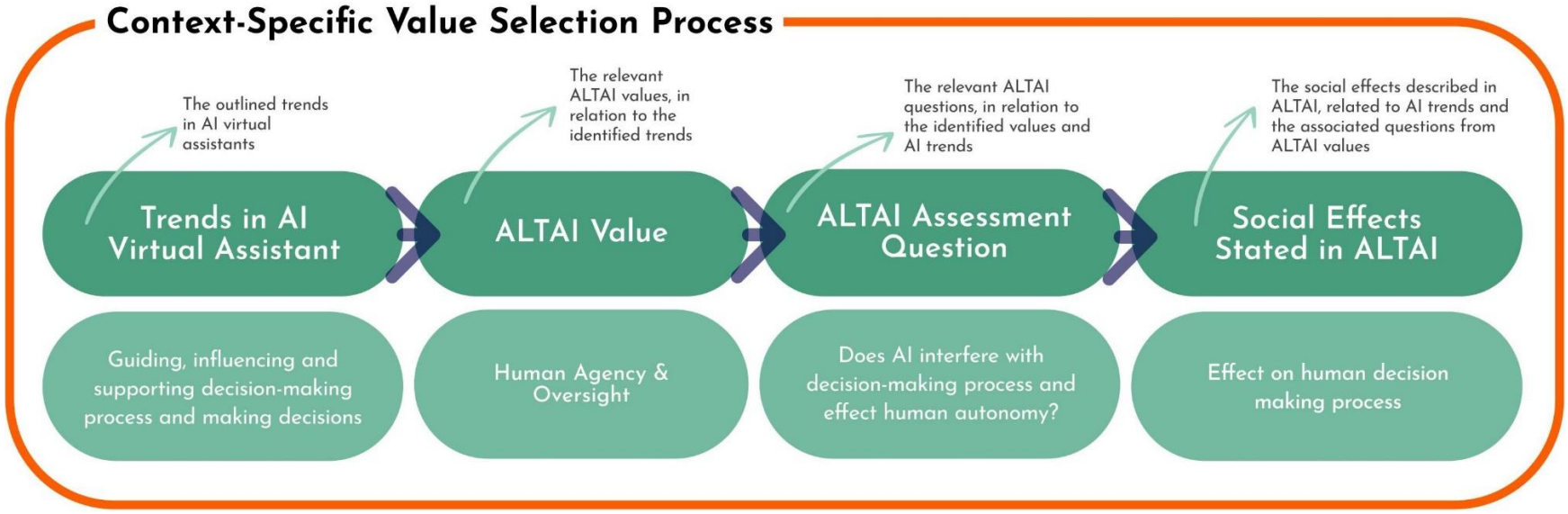


Fig. 5 Process for selecting values and the relevant social effects from ALTAI, based on the trends noted in AI virtual assistants

This research acknowledges the limitations of concentrating on only four specific values and understands that valuable insights may also be derived from the remaining values. However, due to their greater relevance to the research context, the focus is placed on the four selected values to effectively advance towards achieving the research goal.

The selection of ALTAI values is crucial for identifying key human values related to AI virtual assistants that foster social connectivity. Ultimately, these values guide the design phase of this research and inform the generation of design implications that will guide the responsible creation of future AI virtual assistants aimed at enhancing social connectivity, as discussed in the forthcoming chapters.

2.4 Key Takeaways

This chapter establishes the theoretical foundation of this research by conducting a comprehensive review of relevant literature and refining the research scope. The research focuses on the growing application of AI technologies in the social domain and their potential impact on human connectivity. Hence, the theoretical framework is divided into three key topic areas: first, defining *"Social Connectivity"* in the context of this research; second, exploring current and emerging *"AI Technology Trends"* within the social domain;

and third, analyzing *"Responsible AI"* and establishing relevant guidelines for this study.

Finally, this chapter synthesizes the research analyses and refines the research scope from each of the key topic areas as follows:

1) Social Connectivity

Focus on *Relational Connectedness* involving emotional and personal bonds, such as close friendships and relationships with family members. To sustain these connections, people are frequently utilizing social media platforms and are benefitting from AI features embedded within them. Within this domain of social connectivity, this research focuses on the effects on *interpersonal connections*.

2) AI Technology Trends

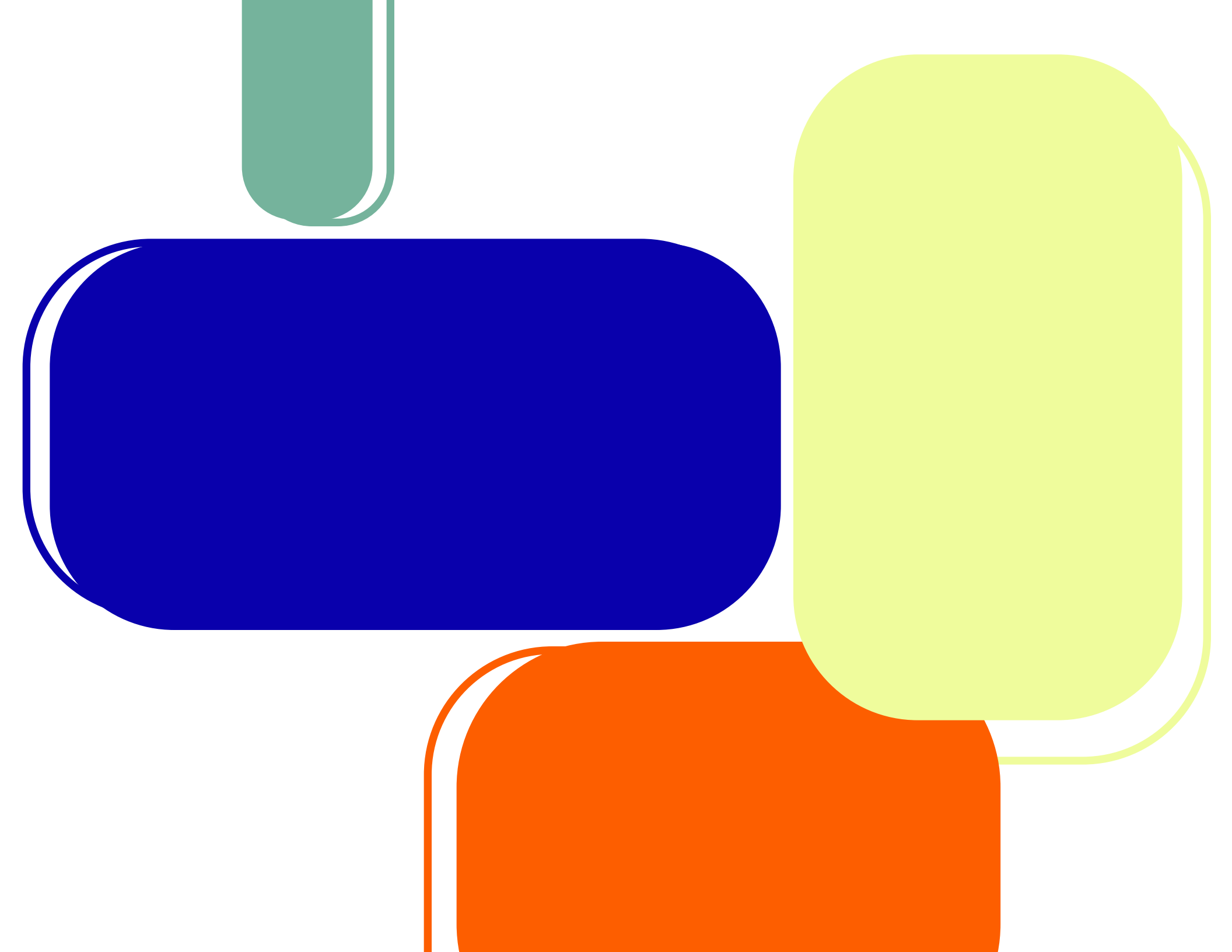
Focus on the role of *future AI virtual assistants* in enhancing interpersonal social connectivity. The outlined trends in AI virtual assistants relevant to the research context are:

1. Multipurpose, offering various functionalities
2. Multimodal, processing various forms of data
3. Humanlike, having characters, personalities, opinions, long term memory
4. Self-learning, autonomously improving performance without human intervention
5. Guiding, influencing and supporting decision-making process and making decisions
6. Generating content
7. Training on personal data
8. Simulating social interaction between people

3) Responsible AI

Focus on selected four ALTAI values, in relation to trends in AI virtual assistant and their relevant effects on interpersonal connections. The selected values are:

1. Human Agency and Oversight
2. Privacy and Data Governance
3. Transparency
4. Societal and Environmental Wellbeing



3. FRAMEWORK FOR DESIGNING RESPONSIBLE AI VIRTUAL ASSISTANTS FOR SOCIAL CONNECTIVITY

This chapter concludes the theoretical research conducted in *Chapter 2*, synthesizing the key insights vital to the research context and organizing them into a comprehensive framework. This framework, titled *"Framework for Designing Responsible AI Virtual Assistants for Social Connectivity,"* underscores the significance of interpersonal social connectivity, the emerging trends in AI virtual assistants, and the relevant values that guide their responsible design. In addition to summarizing and integrating the research insights from the previous chapter within the research context, this framework also directs the design phase of the study and aids in formulating design implications. The subsequent sections will provide a detailed explanation of the framework and its significance for this research:

3.1 Introducing the Framework

The *Framework for Designing Responsible AI virtual assistants for social connectivity*, hereafter referred to as *framework,* is illustrated on the following page, in figure 6.

This framework incorporates the three key topic areas, *Social Connectivity, AI Technology Trends,* and *Responsible AI*, examined in *Chapter 2*, titled *Theoretical Foundation*. The insights obtained from analyzing these areas refined the scope of this research and facilitated the development of the framework.

The framework comprises three main headings, each of which is detailed as follows:

1) Values

The four selected values from ALTAI, along with their relevant questions as per the selection process explained in Section 2.3.2 and illustrated in figure 5.

2) Trends in AI Virtual Assistants

The eight major trends observed in the development of AI virtual assistants for social connectivity, as detailed in Section 2.2.3.

3) Effects on Interpersonal Connectivity

The social effects of AI outlined in ALTAI, in relation to the associated values and the research's focus on interpersonal connectivity, as detailed in Section 2.1.3 and Section 2.3.2. Their selection process is explained in figure 5.

Based on the four selected values, the framework is structured into four distinct segments. Each segment features one of the ALTAI values along with its list of relevant questions (also extracted from ALTAI), as per the criteria established earlier in the report (Section 2.3.2 and figure 5). Adjacent to these values and questions are the trends observed in AI virtual assistants that correspond to the concerns outlined under each ALTAI value. Additionally, the framework connects these values and trends to the potential effects on interpersonal connectivity, as detailed in ALTAI. This includes addressing how the observed trends in AI virtual assistants for social connectivity align with and impact the identified concerns. The details of each segment within the framework are elaborated in the following sections.

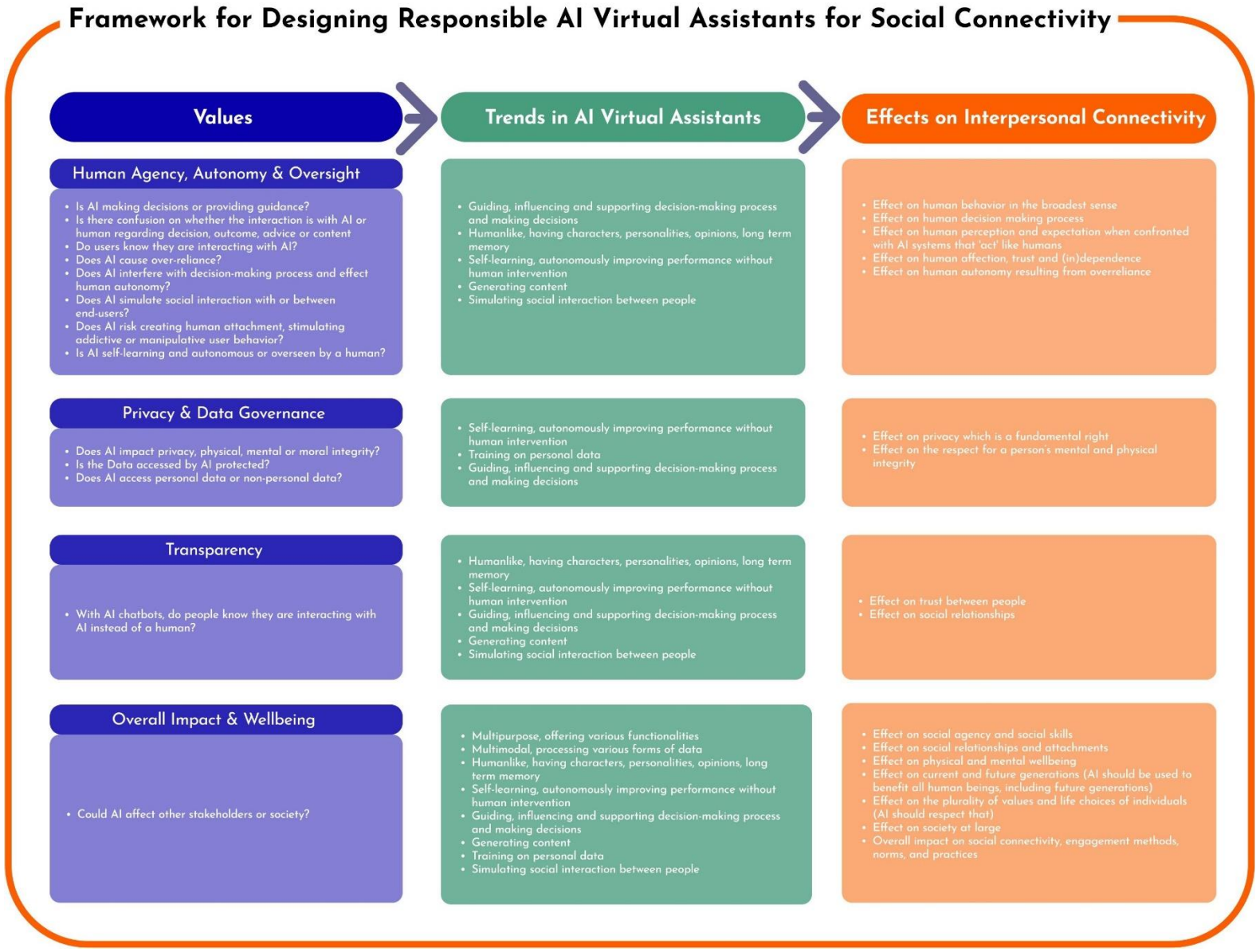


Fig. 6 Framework for designing responsible AI virtual assistants for social connectivity

3.1.1 Framework Segment 1 – Human Agency, Autonomy & Oversight

Fig. 7 Framework segment 1: Human agency, autonomy and oversight

The first segment of the framework (Figure 7) highlights the importance of human agency, autonomy, and oversight in the responsible design of future AI virtual assistants aimed at enhancing interpersonal connectivity. This segment begins by presenting key questions related to human agency, autonomy, and oversight. It then correlates these concerns with relevant trends observed in AI virtual assistants. Finally, it links these

trends and concerns to their potential impacts on interpersonal social connectivity.

For instance, the question *"Is AI self-learning and autonomous or overseen by a human?",* addresses the trend of AI virtual assistants being designed to be *"self-learning, autonomously improving performance without human intervention",* which relates to the effect on interpersonal connectivity, stated as, *"Effect on human affection, trust and (in)dependence."*

3.1.2 Framework Segment 2 – Privacy & Data Governance

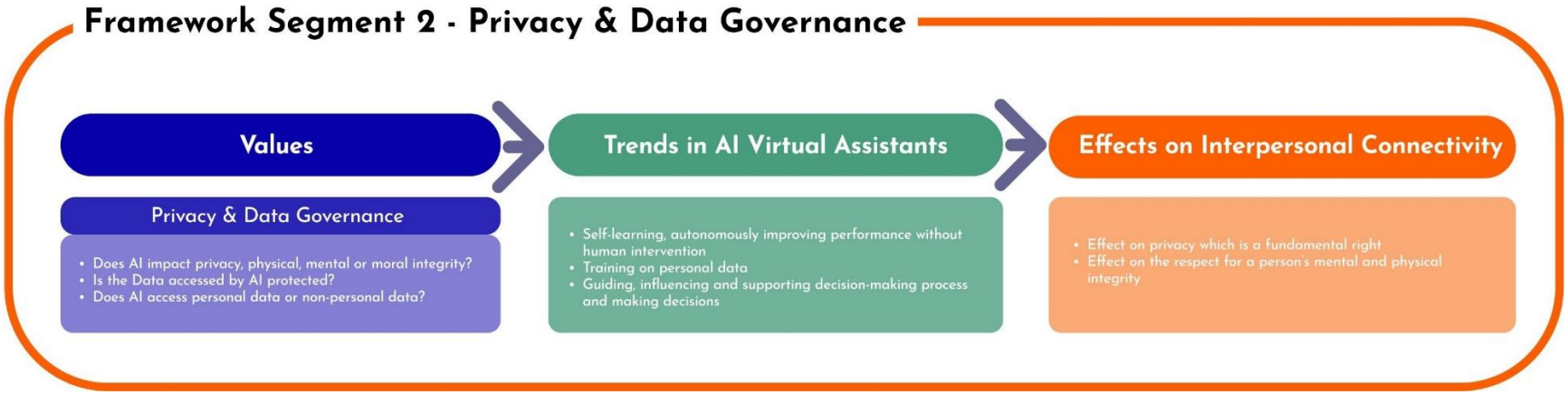


Fig. 8 Framework segment 2: Privacy and data governance

The second segment of the framework (Figure 8) reflects on the importance of privacy and data governance when it comes to designing responsible personal AI virtual assistants for social domain. This segment begins by presenting key questions related to privacy and data governance. It then correlates these concerns with relevant trends observed in AI virtual assistants. Finally, it links these trends and concerns to their potential impacts on interpersonal social connectivity.

For instance, the question *"Does AI access personal data or non-personal data?",* relates to the trend of AI virtual assistants *"training on personal data",* which is finally linked to the *"Effect on privacy which is a fundamental right."*

3.1.3 Framework Segment 3 – Transparency

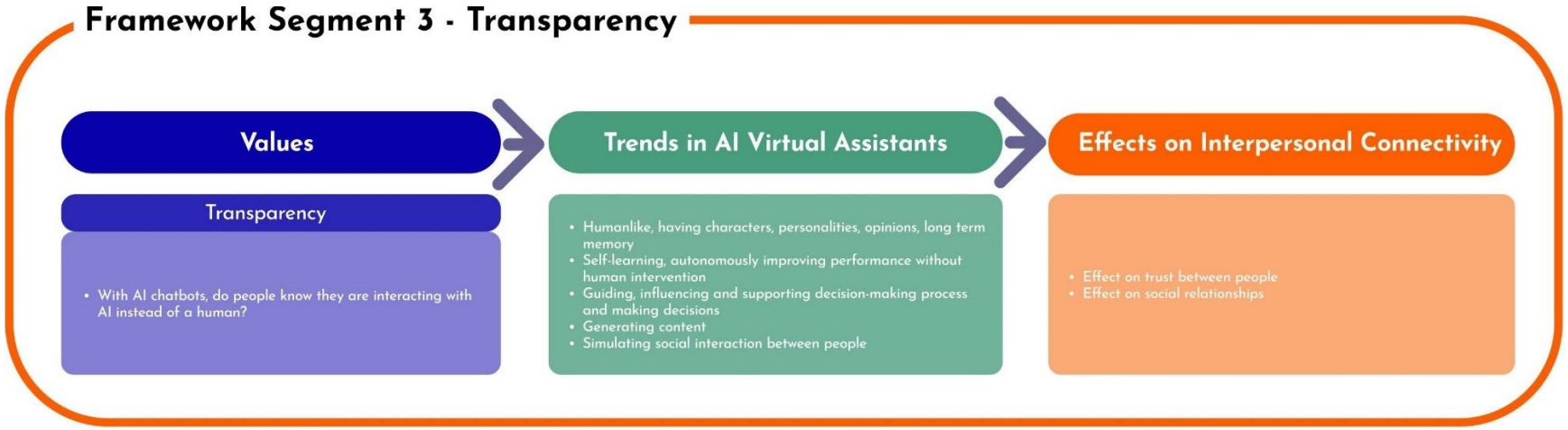


Fig. 9 Framework segment 3: Transparency

The third segment of the framework (Figure 9) addresses ethical concerns related to transparency in communication facilitated by personal AI virtual assistants and their potential impact on interpersonal connections. Like other segments, this segment also presents questions related to transparency. It then correlates these concerns with relevant trends observed in AI virtual assistants. Finally, it links these trends and concerns to their potential impacts on interpersonal social connectivity.

For instance, the question *"With AI chatbots, do people know they are interacting with AI instead of a human?",* relates to the trend of AI virtual assistants being designed to be *"humanlike, having characters, personalities, opinions, long term memory"*. The AI trends are then linked to the effects on interpersonal connectivity, such as, *"Effect on trust between people."*

3.1.4 Framework Segment 4 – Overall Impact & Wellbeing

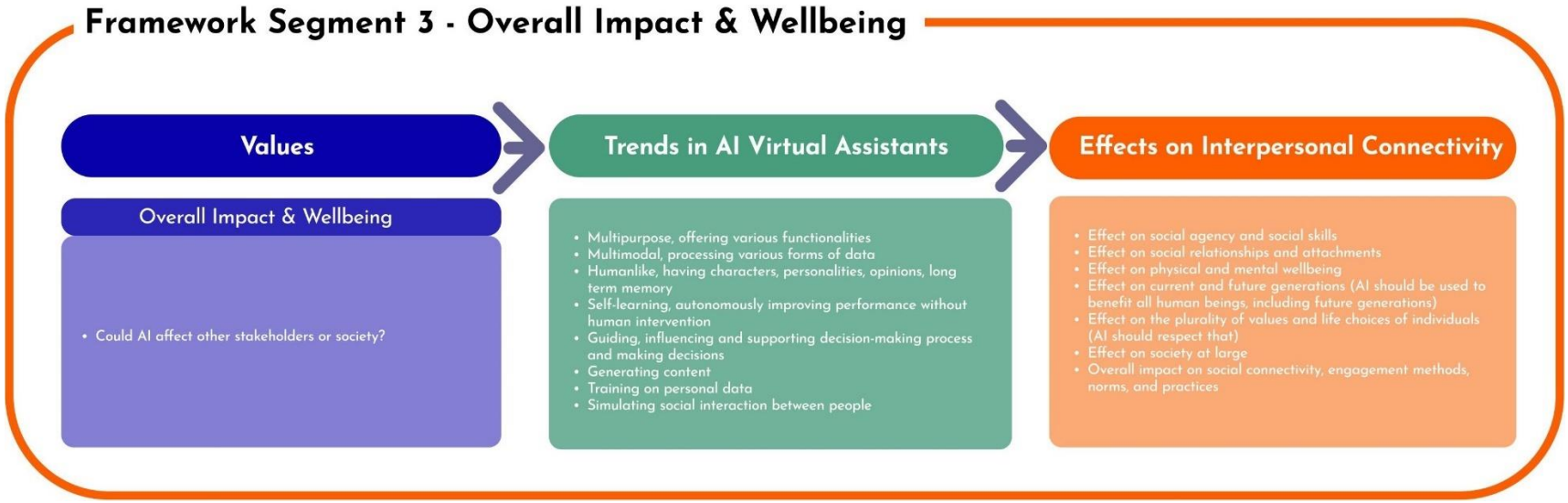


Fig. 10 Framework segment 4: Overall impact and wellbeing

The final segment of the framework (Figure 10) underscores the broad societal impact of the trending AI virtual assistants fostering interpersonal connectivity.

In this segment, the question *"Could AI affect other stakeholders or society?"* is quite broad. This question incorporates all the trends noted in AI virtual assistants, as outlined in this research. The effects listed encompass both individual and societal impacts. For instance, the AI trend *"simulating social interaction between people"* can be related to the *"effect on social relationships and attachments."*

3.2 Significance of the Framework

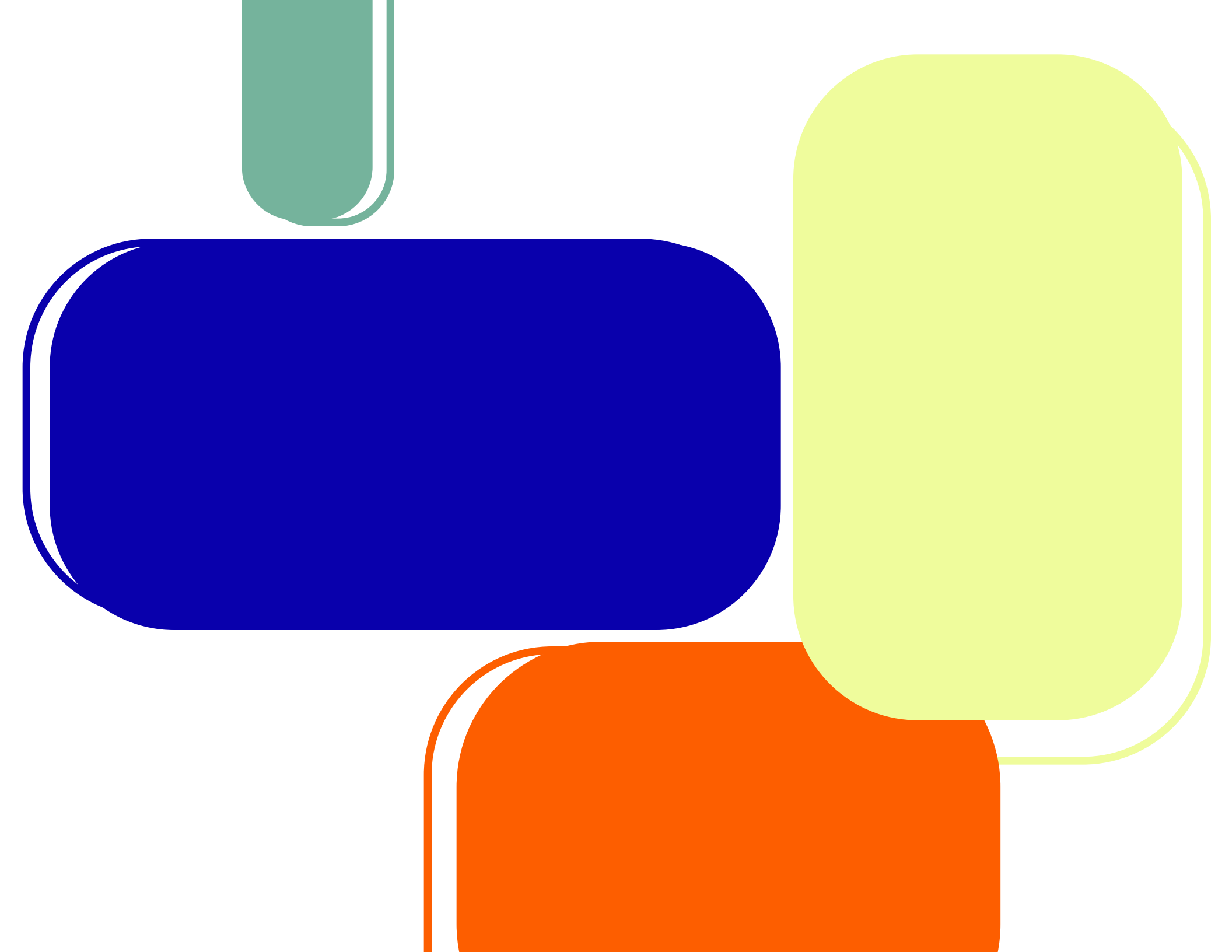
The framework, titled *"Framework for Designing Responsible AI Virtual Assistants for Social Connectivity,"* as presented and detailed in the previous sections, is of significant importance to this research. It serves as a guiding structure that steers the research towards its ultimate goal of generating design implications for the responsible design of future AI virtual assistants aimed at fostering social connectivity. The role and significance of this framework within the context of this research are outlined below:

1. It synthesizes and concludes the theoretical research and narrows the scope of the project

2. It establishes a direct connection between the three interrelated topic areas—*Social Connectivity, AI Technology Trends, and Responsible AI*—that are essential to this research, each of which was analyzed individually in Chapter 2

3. It identifies values, relates them to trends in AI virtual assistants and reflects on their effects on interpersonal social connectivity

4. It bridges the theoretical research and design phases of this project by distilling key insights from the literature and identifying the essential elements to be incorporated in the design phase. Ultimately, its role in the design phase is crucial in guiding this project towards achieving its goal of generating design implications for a responsible design of future AI virtual assistants fostering interpersonal connectivity.



4. RESEARCH METHODOLOGY

Focusing on the development of future AI virtual assistants designed for interpersonal social connectivity, this research aims to elicit the values of young adults concerning these advancements, alongside the previously identified values (Section 2.3.2). By analyzing current trends in AI technology, this research aims to anticipate the direction of these developments, as well as their potential outcomes and impacts on human relationships. Through the adoption of Design Futuring to explore potential future scenarios, the research aims to facilitate discussions among young adults, enabling them to articulate what aspects they consider most important in these emerging technologies for social domain.

To advance the research towards designing futuristic scenarios that stimulate discussion and elicit values through dialogue, this chapter explores Speculative Critical Design (SCD) and Value Sensitive Design (VSD) as key approaches. It delves into a comprehensive understanding of these two approaches, analyzing their distinct features, similarities, and combined potential. The chapter concludes by integrating relevant insights from both approaches and outlining the final approach adopted for this research.

4.1 Speculative Critical Design (SCD)

"What we are interested in, though, is the idea of possible futures and using them as tools to better understand the present and to discuss the kind of future people want, and, of course, one's people do not want."

(Dunne & Raby, 2014)

Futures Studies, also referred to as *Design Futuring*, involves designing with a focus on the long-term impacts by exploring potential future scenarios. This approach incorporates *Discursive Design* practices (Dunne & Raby, 2013), such as *Speculative Critical Design (SCD)*. Speculative Critical Design (SCD) strives to be imaginative, encourages reflection on current realities, probes possible futures, and plays a crucial role in envisioning the world of tomorrow (Dunne & Raby, 2013; Mitrović, 2016; Mitrović et al., 2021).

The book *'Beyond Speculative Design: Past – Present – Future'*, emphasizes on the role of *SCD* in:

"...challenging the status quo of design practice and education: from questioning modernist functionalism, decoupling it from the market, and bringing back its discursive role; to raising discussion about political and social issues relating to technological development, and finally in envisioning alternative futures." **(Mitrović et al., 2021).**

By fostering "social dreaming" about desired futures, Speculative Critical Design (SCD) extends critical practice into realms of imagination and diverse visions of potential scenarios. It promotes critical thinking, raises awareness, and inspires action by encouraging designers to connect the dots and view the world from various perspectives (Mitrović et al., 2021). SCD utilizes design as a medium to provoke thought and dialogue, engaging in discursive activities that stimulate discussion and debate (Johannessen et al., 2019). Thus, by crafting speculative scenarios, SCD challenges existing assumptions and stimulates reflection and dialogue about the future (Dunne & Raby, 2013; Mitrović et al., 2021).

"Thinking differently about tomorrow is one thing, finding a way to act on it is another. We need to be actively engaged in that conversation, collaborate, compromise and help society build those futures."

Phil Balagtas, Beyond Speculative Design: Past – Present – Future (Mitrović et al., 2021).

For this research, Speculative Critical Design (SCD) provides an approach to envision potential futures of AI virtual assistants within the social domain, grounded in current trend analysis. By developing speculative scenarios, this research aims to facilitate discussions among young adults, encouraging them to express their perspectives on the possible social impacts of such AI advancements. Through these dialogues, the project seeks to understand young adults' perceptions of AI virtual assistants for social connectivity, address ethical

concerns and their effects on human relationships, and ultimately elicit relevant values to guide a responsible design of such AI technologies.

4.1.2 Application of Speculative Critical Design (SCD)

SCD is considered more of an approach than a strict methodology and provides designers with the flexibility to adopt tools and methods best suited to the context (Johannessen et al. 2019). Drawing from the overall application process as described by Johannessen et al. (2019) and Mitrović (2015), this research adapts and outlines the steps most relevant to its context as follows:

1) Defining the Context for Debate

The first step involves defining the design space and the context for social debate through critical design research (Johannessen et al., 2019; Mitrović, 2015). Emphasis is placed on identifying trends and emerging changes in the present through groundbreaking innovations and technological advances that have the potential to shape our future. This *'Signal Scanning'* phase serves as a source of inspiration to imagine possible future scenarios while ensuring they remain grounded in current realities and ongoing developments (Colosi, 2023).

2) Designing Scenarios

The exploration of possible futures within Speculative Critical Design often takes the form of scenarios (Dunne & Raby, 2013). Following the initial step of gaining a deeper understanding of and defining the research context, the identified signals are synthesized into specific speculative scenarios that outline possible future worlds (Colosi, 2023). This scenario development phase begins by posing 'what if' questions (Johannessen et al., 2019; Dunne & Raby, 2013). These scenarios, designed to be provocative, open-ended, fictional, and intentionally simplified, aim to create spaces for discussion and debate. The narratives of these scenarios are crafted using various methods to be provocative enough to engage the audience on emotional, intellectual, and psychological levels (Johannessen et al., 2019). Their fictional nature enables viewers to suspend disbelief and allow their imaginations to explore, momentarily setting aside current realities to contemplate possible future developments (Dunne & Raby, 2013).

The development of possible futures serves as a medium to stimulate imaginative thought and speculation. The scenarios not only facilitate this but also incorporate critique, as people begin to reflect on the present through the lens of these imaginative projections. The developed scenarios should appear believable, scientifically plausible, and connected to current realities. This grounding in today's world is essential. A believable series of events leading to the new situation, often referred to as the 'perceptual bridge,' is necessary even if the scenarios are entirely fictional (Johannessen et al., 2019; Dunne & Raby, 2013).

To strengthen the scenarios and support the 'suspension of disbelief,' they are meticulously curated according to the prior research and context. This involves addressing questions such as 'when is the scenario situated,' 'what rules apply in the

scenario,' 'should the scenarios be utopian, dystopian, neutral, or ambiguous,' 'who should be the protagonist of the scenarios,' and 'what kinds of products or services are people using,' among others (Mitrović, 2015; Colosi, 2023). This allows viewers to relate the scenario to their own world, using it as a tool for critical reflection.

3) Delivering the Scenarios

Critiquing norms and stimulating social debate are fundamental goals and desired outcomes of Speculative Critical Design (SCD). This is achieved by materializing the designed scenarios from earlier phases through narratives, objects, or a combination of both, in any medium that appropriately represents the concept. These scenarios can be delivered through provocative prototypes, also known as 'provotypes,' or diegetic prototypes (Casnati et al., 2023). This stage focuses on presenting these narratives to the audience in an engaging and convincing manner, employing storytelling and aesthetic techniques found in literature and film (Luria & Candy, 2022). The presentation of scenarios serves as a provocative tool to stimulate discussion between the designer and the audience. This debate forms part of a larger design research process. In this project, it involves value elicitation through speculation.

These steps are essential to this project as they 1) leverage the literature analysis and insights from previous chapters and grounding material, integrating them with the signal scanning phase; 2) inform the key aspects to consider when designing speculative scenarios; and 3) provide direction on how to materialize and present these scenarios effectively to provoke meaningful and impactful discussions.

4.2 Value Sensitive Design (VSD)

"Value Sensitive Design guides the shape of being with technology."

(Friedman & Hendry, 2019).

The working definition of *value* in terms of human values, as defined in the book *'Value Sensitive Design: Shaping Technology with Moral Imagination'* is, ***"what is important to people in their lives, with a focus on ethics and morality"*** (Friedman et al., 2017). Centered around these human values, *Value Sensitive Design (VSD)* offers a theoretically grounded approach to designing technology that considers human values throughout the design process (Friedman et al., 2017). In their book, Batya Friedman and David G. Hendry further elaborate on the interactional stance on technology and human values adopted by VSD: *"Unlike approaches based primarily on technological determinism or on social determinism, interactional theories such as value sensitive design posit that human beings acting as individuals, organizations, or societies shape the tools and technologies they design and implement; in turn, those tools and technologies shape human experience and society* (Friedman & Hendry, 2019)."

This research underscores that the relationship between technology and human values is inherently interactional. It highlights the co-evolution of social structures and technology,

emphasizes the importance of analyzing stakeholder values and a commitment to continual progress rather than the pursuit of perfection. This research employs *'Value Scenarios'* as a method to elicit values (Friedman & Hendry, 2019). Value scenarios highlight specific features of a technology within its immediate context of use, in the form of a narrative, facilitating stakeholder engagement in research investigations. By presenting the context through scenarios, this method helps surface both overarching and individual societal and humanistic considerations related to the technology and its use.

Through Value Sensitive Design (VSD), this research seeks to identify and elicit the values of young people concerning interpersonal connections facilitated by AI virtual assistants. The insights gained from these values will inform the development of design implications, guiding designers in the responsible and ethical creation of future AI virtual assistants that promote interpersonal social connectivity.

4.2.1 Application of Value Sensitive Design (VSD)

Value Sensitive Design (VSD) offers a range of tools and methods that address diverse objectives in the ethical and responsible development of technology (Friedman et al., 2017; Friedman & Hendry, 2019). This project utilizes *Value Scenarios* to elicit values within the research context. The following steps outline its effective application within this research.

1) Defining and Framing Technical Aspects

VSD emphasizes the importance of *placing technology within a situated context*, identifying these elements as central to the project. This approach brings relevant value issues to the forefront. By *explicitly supporting project values*, researchers and designers convey their perspectives and thoughts, backed by comprehensive background research. At this stage, it is also crucial to determine which aspects should be addressed through socio-cultural interventions, focusing on the technology under consideration and its value implications. Once that is defined, a set of initial values is identified to form a starting point for the design process (Umbrello & Van De Poel, 2021)

2) Designing and Delivering Value Scenarios

The designer's stance, identified values, technology implications along with the technology and its context are put together in the form of value scenarios, depending on the intended goal and context. The value scenarios, in any form that supports the context, e.g. text, sketches, videos, story boards etc. are presented to the stakeholders, while defining aspects such as the ethical tone, timescale, mundanity, levels of detail etc. (Richmond et al., 2023). Furthermore, Value Sensitive Design (VSD) recommends conducting worst-case analyses during the technical design process to anticipate and address potential significant harms (Friedman & Hendry, 2019).

In the research paper *"Parenting from the Pocket: Value Tensions and Technical Directions for Secure,"* the authors utilize short stories in textual format to illustrate their findings

(Czeskis et al., 2010). They explain this process by stating that *"value scenarios comprise fictional vignettes that emphasize social and value implications of a 'hypothesized' technology. A few rich, nuanced scenarios can help to focus attention on indirect as well as direct stakeholders, nefarious and unusual uses, value tensions, and longer-term societal implications that might otherwise go unnoticed."* Similarly, in the research paper *"Exploring Value Dilemmas of Brain Monitoring Technology through Speculative Design Scenarios,"* (Risnes et al., 2024) the authors develop scenarios as brief, simple videos to introduce various ways of using a brain monitoring device. Likewise, the research *"Broadening Privacy and Surveillance: Eliciting Interconnected Values with a Scenarios Workbook on Smart Home Cameras"* (Richmond et al., 2023) explores the development of multiple scenarios through simple sketches compiled in a workbook.

The scenarios are delivered in multiple ways. For instance, primary stakeholders are invited to participate in a semi-structured interview (Czeskis et al., 2010; Richmond et al., 2023) or a focus group (Risnes et al., 2024). These approaches, while involving stakeholders, balance addressing essential questions with uncovering unexpected and new insights.

3) Value Elicitation

Based on the insights gained from the stakeholders, designers can identify the corresponding values, informed by their prior work and research. Once the key values emerge from the dialogue, different methods or combinations of methods may be adopted for value elicitation or representation (Friedman & Hendry, 2019).

4) Formulating Design Norms

Based on the identified and elicited values, design norms are formed, as introduced in the notion of values hierarchy (Van de Poel, 2013). These norms further guide the responsible development of technologies by ensuring alignment with stakeholder values.

These steps are essential to this project as they 1) define technical aspects such as the specific details about the technology and facilitate initial value identification; 2) inform the aspects to consider when designing scenarios and guide with scenario delivery; 3) guide with the process of value elicitation; and 4) aid in forming design norms that align with the elicited and identified values.

4.3 Adopted Research Methodology

Both approaches explored in the previous sections, SCD and VSD, emphasize the interactional relationship between humans and technology and how this relationship shapes our way of living. This project employs a combination of insights from both approaches and adapts a methodology that aligns with its context and goals. An overview of the methodology employed in this project, which bridges the previous chapters with the forthcoming ones, is outlined below:

1) Framing the Technical Space

The first step entails establishing the core elements that are critical to the research context and essential for defining the technical space. A comprehensive understanding of the relevant topics is achieved through thorough literature reviews. This phase also defines the boundaries of the research context, distinguishing relevant elements from those that are not, and provides direction by offering a more specific context to guide the subsequent design phase.

2) Value Identification

A set of initial values, relevant to the research context is developed which provides a starting point for the design process. These values are the ones that should be respected according to the principles defined for AI (Friedman & Hendry, 2019). They are also interpreted and conceptualized according to the research context (Umbrello & Van De Poel, 2021).

3) Developing Design Scenarios

In this phase, scenario ideas are developed within the design space defined earlier. This process begins with the generation of *"What If"* statements to stimulate thought and create narratives, setting up the scenario space for the development of fictional scenarios. These *"What If"* statements serve as a foundation for exploring ideas and finalizing the concept for the design scenarios. The final scenarios depict a futuristic, fictional world that incorporates hypothesized technology and characters, designed to be both provocative and believable.

4) Delivering Design Scenarios to Stakeholders

Combining tools from *Speculative Critical Design (SCD)* and *Value Sensitive Design (VSD)*, the developed narratives materialized into scenarios visuals presented to direct stakeholders. Stakeholders are invited to participate in semi-structured focus groups where these scenarios are showcased to provoke discussion. This setup aims to elicit useful insights and surface values relevant to the research. The focus groups are structured to encourage participants to freely share their thoughts, facilitating a deeper understanding of their underlying values.

5) Eliciting Stakeholder Values

Finally, the two approaches collectively facilitate value elicitation through participant discussions. Insights from these discussions are extrapolated and subjected to thematic analysis (Delahunt, 2017). This process aids in reflecting upon initially identified values and gaining a context specific understanding of values at stake (Umbrello & Van De Poel, 2021). It also elicits new stakeholder values within the context of the research and the scenarios presented, guiding the project toward its end goal of forming design implications.

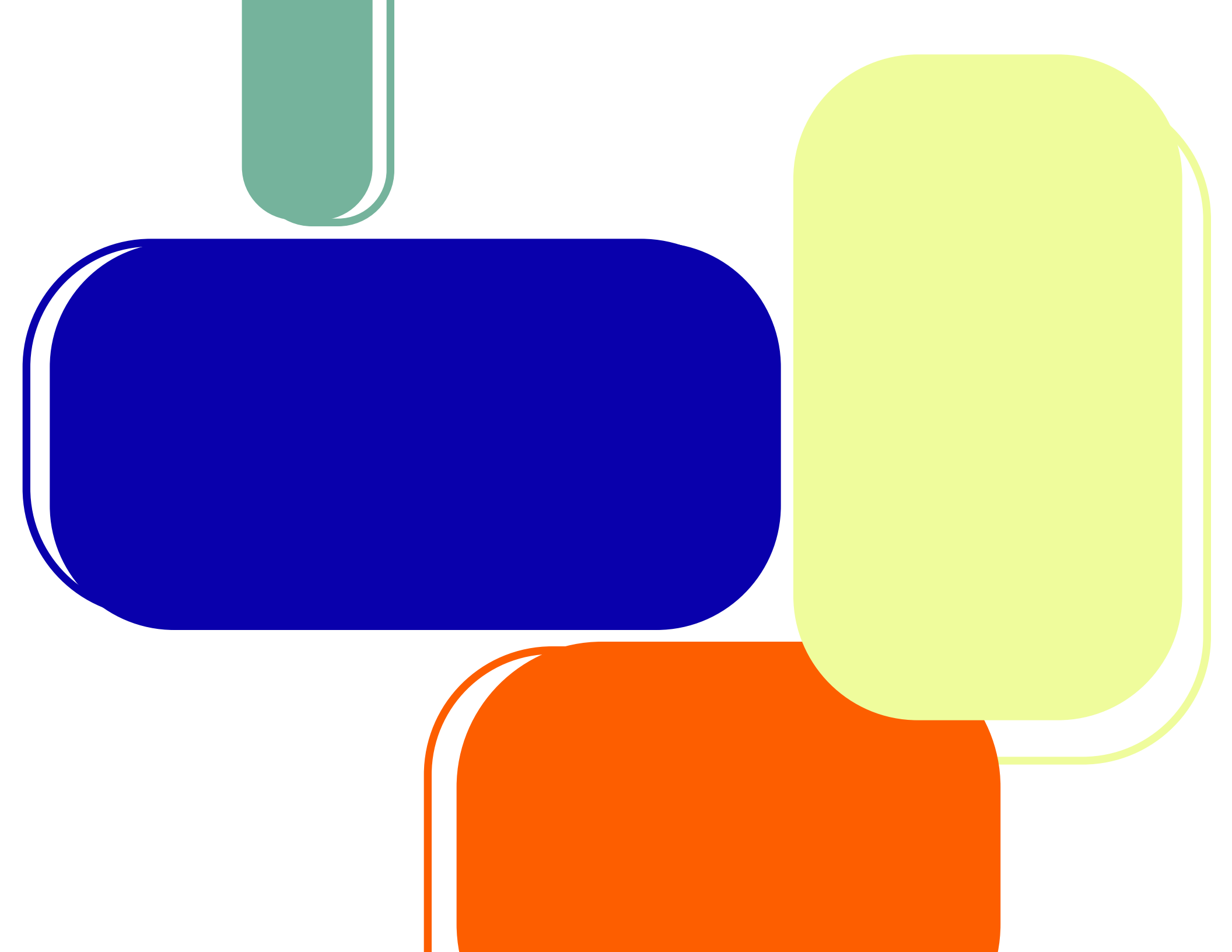
6) Generating Implications for Design

To generate design implications, value hierarchy is used as a tool (Van de Poel, 2013). The identified and elicited contextual values are translated in design norms to guide a responsible design of future AI technologies (Umbrello & Van De Poel, 2021).

4.4 Key Takeaways

This chapter explores SCD and VSD as research approaches to:

1. Establish a connection between the theoretical research and contextual insights previously gathered and the forthcoming design phase.
2. Establish a connection of the previously identified responsible AI values to the forthcoming design phase.
3. Guide the development of design scenarios by incorporating insights from the two aforementioned points.
4. Provide a starting point for the design phase
5. Guide the design scenarios to be futuristic and speculative yet realistic and grounded to current trends.
6. Guide the development of the scenarios in a way that provokes critical thinking and discussion among young adults.
7. Guide the thematic analysis of the data gathered through scenario discussions to gain full awareness of what is at stake, gain a context specific understanding of the identified values and elicit new contextual values
8. Guide the formation of design norms from the context specific values, using values hierarchy, in order to generate design implications to guide future designers with a responsible design of AI virtual assistant fostered interpersonal social connectivity.



5. DESIGN SCENARIOS

This chapter builds upon the insights gained through the previous chapters to develop and deliver design scenarios. The design scenarios are intended to provoke discussions among young adults about the technology within the given context, allowing them to articulate their values by expressing their thoughts. Using the methodology outlined for this research in Chapter 4, the design scenarios are developed as guided by the framework presented in Chapter 3. According to the contextual insights in the framework regarding the identified values, AI virtual assistant trends and their social effects, this chapter explains how the scenarios were designed, developed, finalized, delivered and discussed. The chapter contains the following sections:

5.1 Scenario Development

The scenarios were developed by first establishing key parameters to ensure their effectiveness, including the time period in which the scenarios are set, the ethical tone to be conveyed, the desired level of detail, the everyday nature (mundanity) of the scenarios, and the identities of the protagonists. By defining these elements in alignment with the framework presented in Chapter 4, several *"what if"* statements were generated to guide the creation of the scenarios. These statements served as the foundation for developing a narrative around a fictional technology featured in the scenarios. Once the narrative was finalized, the scenarios were translated into visual representations. These visual scenarios were then presented to stakeholders in a semi-structured focus group. The scenario development process is outlined in detail in the following sections.

5.1.1 Setting the Scene

1) Timescale:

The timescale for the scenarios was set to 5-10 years from the present, approximately between 2030 and 2035. This relatively near-future setting was chosen to envision the evolution of current technological trends and speculate on their development over the coming years. By situating the scenarios closer to the present, the aim was to make them more comprehensible, believable, and relatable for the participants, thereby enhancing their realism and engagement.

2) Ethical Tone:

The ethical tone of the scenarios was intentionally set to be ambiguous. This approach was selected to foster open discussion and allow participants the freedom to express their views without being influenced by predefined narratives. By avoiding explicit framing as utopian, dystopian, or neutral, the ambiguous tone encourages participants to interpret the scenarios from their own perspectives, thereby promoting a more diverse range of insights and reflections.

3) Level of Detail:

Similarly, the level of detail was carefully crafted to offer a broad and general understanding of a fictional technology and its associated functionalities. This approach was intended to allow participants to imagine themselves interacting with the technology in various contexts and for a range of purposes. Rather than exploring the technical intricacies in depth, the emphasis was placed on communicating the technology's capabilities and potential applications.

4) Level of Mundanity:

The scenarios were designed to illustrate new technological possibilities in the future while capturing the unique aspects of the fictional world. These scenarios aimed to be idiosyncratic yet firmly rooted in everyday life and routine activities, emphasizing the extraordinary within the context of ordinary existence.

5) Protagonist:

The protagonists in the scenarios were crafted to be a blend of the participants themselves and fictional characters. This dual approach was chosen to encourage participants to engage with the scenarios from both their own perspective and the viewpoint of the characters, thereby deepening their reflection and involvement.

5.1.2 Generating "What Ifs"

Based on prior research and the *"grounding material"*, several *"what if"* statements were crafted to initiate the formation of narratives. The *framework* outlined in Chapter 4 guided the development of *"what if"* statements by: *first,* integrating current AI trends in features and functionalities; *second,* identifying the potential social effects of these advancements; and *third,* highlighting the relevant values at stake. These insights from the framework were paired with imaginative and creative thinking and several *"what if"* statements were generated accordingly. Thus, each statement extended beyond current AI trends to explore potential future developments, raising relevant concerns about their impact on values and social connections as outlined in the framework.

Designed to trigger thought and provide direction, these statements were kept simple and structured as open-ended prompts. Some examples include:

- What if you no longer have to worry about making new friends because your personal AI virtual assistant can do it for you?
- What if all communication occurs only through personal AI virtual assistants?
- What if your personal AI virtual assistant can keep you connected with your loved ones by sharing your live updates with them throughout the day?
- What if your personal AI virtual assistant can represent you digitally and maintain your social connections?
- What if your AI virtual assistant is a digital version of yourself and assists you in maintaining your connections for you?
- What if your AI virtual assistant is a digital version of yourself, responsible for maintaining your connections with friends and family, and over time, these individuals develop a stronger attachment to the digital version than to the real you?
- What if, during a conflict with a friend, you can let your AI virtual assistant take over and resolve the situation for you?

Fig. 11 Generating "what if" statements

These *"what if"* statements served as catalysts for generating ideas and guiding further development. Through careful thought and analysis, they were subsequently refined and filtered to better align with the research objectives. The refined and shortlisted statements are as follows:

- What if AI virtual assistants were highly personalized, capable of mediating conflicts on your behalf, and managing difficult, awkward, or uncomfortable conversations with your friends or family?
- What if AI virtual assistants could provide highly personalized support by learning from your communication style and automating the management of your friendships, ensuring that you remain constantly connected with your friends?

Ultimately, the final what if statement selected to proceed with was *"What if AI virtual assistants could provide highly personalized support by learning from your communication style and automating the management of your friendships, ensuring that you remain constantly connected with your friends?"*

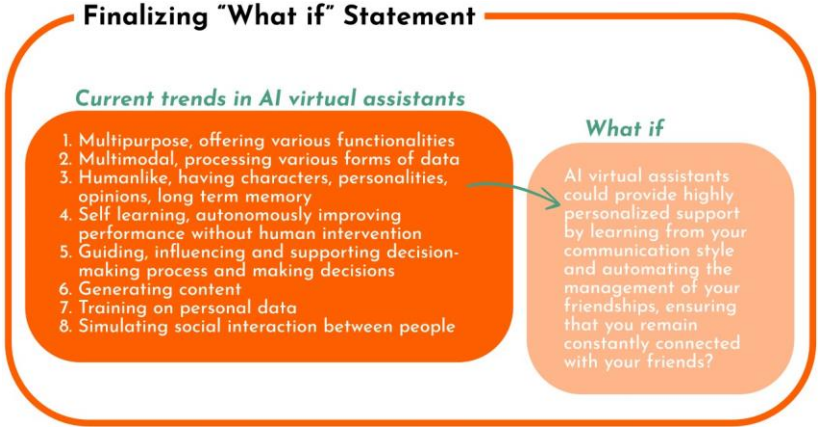


Fig. 11 Finalizing *"what if"* statement according to the AI trends defined in the framework

The *framework* guided the selection of this statement by:

1. Providing a comprehensive list of current AI trends in virtual assistants, facilitating imaginative and creative thinking about their potential future developments and the features and functions they might offer (figure 10).
2. Identifying responsible AI values at stake which facilitated reflection on how future developments in AI

virtual assistants might impact these values within the given context.

3. Highlighting the relevant social effects on interpersonal connectivity, facilitating the imagination of scenarios in which these effects could emerge

The final *"what if"* statement played a crucial role in refining the concept of fictional technology, guiding the development of detailed textual narratives that culminated in the final scenarios. Imagination around highly personalized virtual assistants of the future was inspired by current trends in AI virtual assistants, particularly the concept of digital assistants evolving to become digital versions of their users. This idea was further influenced by Meta AI's character-based virtual assistants, which mimic the facial identities, personalities, expressions, and communication styles of real people, such as celebrities. This sparked the notion that future virtual assistants could be developed for individual users, mimicking their facial identities, digitally representing them by learning from their behaviors and communication styles, adopting elements of their personality and preferences, and assisting with various tasks while fostering social connections with their friends.

Building on this concept of a fictional *"Digital Twin Assistant"* technology (fig. 12), several narratives were constructed (fig. 13), eventually leading to the selection of final narratives that explain the technology and its contexts of use. The following sections provide a detailed exploration of these finalized scenarios.

Fig. 12. Brainstorming ideas on the development of a fictional “Digital Twin Assistant” technology

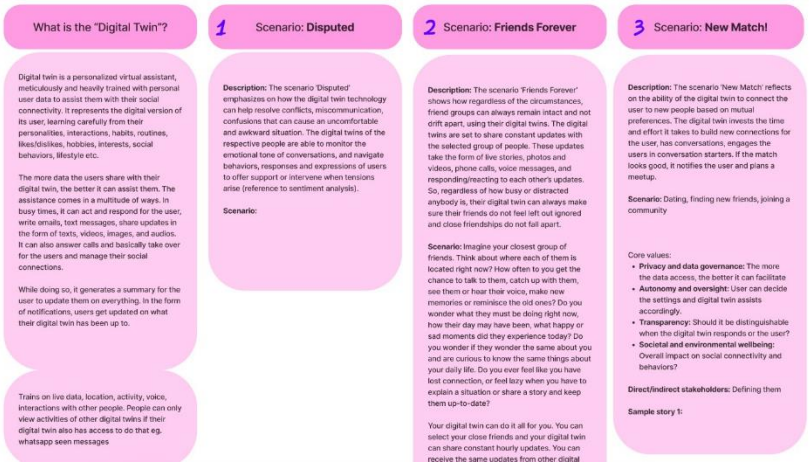


Fig. 13 Exploring textual narratives based on the "Digital Twin Assistant" technology

5.1.3 Finalizing the Scenario

The final scenarios developed introduce a fictional personal AI virtual assistant called the *"Digital Twin Assistant (DTA)"* and illustrate its role in fostering interpersonal connections, such as friendships. The three finalized narratives, titled *"DTA 2.0," "The Birthday Wish,"* and *"The Hiking Trip,"* first aid in explaining how the DTA functions and then provide two specific examples of its use. These three scenarios are described in detail below.

1) DTA 2.0 by Connect

This scenario was developed to introduce the *Digital Twin Assistant (DTA)* technology, developed by a fictional company called *"Connect"*, highlighting its various characteristics, functions, and capabilities. The scenario is narrated as follows:

"Years ago, when we launched our very first collection of character-based AI virtual assistants, it sparked a revolution. In recent years, with the groundbreaking release of our Digital Twin Assistant (DTA) technology, we have come a long way in transforming the presence of AI virtual assistants in people's lives. The DTA was not just a digital version of you but also the most intelligent, efficient, and personalized digital assistant.

With the launch of update 2.1, the Digital Twin Assistant now manages your social connectivity like never before. It represents you across social platforms and handles your connections optimally, helping you maintain existing relationships and forge new connections. Given access, the Digital Twin Assistant can learn from your data, such as phone calls, messages, photo gallery, social media engagement, and browsing history. Additionally, it analyzes your expressions, behaviors, and engagement methods in real time to understand your socializing and engagement styles. By accessing more personal data, the DTA trains more effectively, offering highly personalized assistance.

It can now connect with your friends, understand your personality traits, preferences, interests, and hobbies, and assist you accordingly. Whether it's joining a video call with friends on your behalf, replying to your messages, planning your trip to the mountains, teaching your fried chicken recipe to your friends or finding the right words to offer condolences, your DTA has got you covered.

With such advanced personalization, the DTA ensures all your social connections are well-maintained and nurtured. We're thrilled to share these advancements with you today and can't wait for you to experience the new capabilities of the Digital Twin Assistant.

Install the update now and connect like never before!"

This scenario narration integrates key values identified from the *framework* (Chapter 4), including accessibility, the use and management of personal data, and human-like abilities such as decision-making, taking actions, and expressing emotions and sentiments. It also addresses relevant concerns related to transparency and considers the broader social effects of these capabilities.

2) The Birthday Wish

This scenario explores how the *Digital Twin Assistant (DTA)*, introduced in the narrative above, can help people maintain connections with their long-distance close friends. The scenario is narrated below:

"Anna met Lisa when she was still in school, and from that moment, Lisa became one of her closest friends. As they grew older, life took them on different paths. Anna moved to another city for graduate studies, and later, Lisa moved abroad for a new job.

Despite the physical distance, busy schedules, and time differences, they both made an effort to stay connected. As their lives grew more hectic, Anna started using her Digital Twin Assistant (DTA) to help maintain their connection.

Anna gave her DTA access to their chats, calls, photos, videos, and social media interactions. With this information, the DTA learned about their deep bond and automatically generated a personalized, heartfelt birthday wish for Lisa every year. Anna can review and approve the birthday wish before the DTA sends it on her behalf, ensuring Lisa feels loved and remembered on her special day."

This scenario encompasses essential aspects of interpersonal connections, such as emotional expression and trust, and examines how the DTA could be employed to sustain friendships. It also addresses critical concerns related to transparency, the DTA's autonomy and decision-making, human control and involvement, and the broader impact of such technology on interpersonal relationships and society.

3) The Hiking Trip

This scenario examines how the Digital Twin Assistant (DTA) could facilitate digital communication, such as participating in a video call on behalf of the user and interacting with others accordingly. The narrative depicts a group of friends attempting to plan a trip during a video call, with both actual humans and the DTAs of the absent participants. The scenario is narrated below:

"Two weeks ago, Zayn, Sabrina, Moira, and Fin scheduled a group call to plan their long-anticipated hiking trip to the mountains. At the last minute, personal and work emergencies prevented Zayn and Fin from joining, so they

had their Digital Twin Assistants (DTAs) take their place in the group call.

As the call began, the usual chit-chat quickly shifted to trip planning. Everyone shared their thoughts on which trail they preferred. Zayn's and Fin's DTAs represented their thoughts, ideas, and opinions to their best abilities.

Zayn's DTA constantly disagreed with the others about choosing the easier trail, as Zayn enjoys steeper and more challenging hikes. Moira reminded the group that Zayn had mentioned wanting a more relaxed hike this year, though Zayn's DTA insisted on his usual preference for intense trails.

Meanwhile, Fin's DTA advocated for the riverside trail, appreciating its scenic beauty. Sabrina and Moira also voiced their preferences and disagreements, contributing to a lively discussion.

Despite differing opinions, the group managed to agree on certain points. For the remaining decisions, Sabrina and Moira agreed to wait for Zayn and Fin to get back to them after receiving the call summary from their respective DTAs."

This scenario integrates critical elements, including a dynamic setting where humans interact with another individual's DTA. The DTA facilitates users by digitally representing them in the interaction and later updating them on the proceedings. It also embodies and communicates the user's thoughts and ideas to others. This scenario further addresses concerns related to human autonomy, control, and oversight of the DTA's actions, the implications of the DTA recording the conversation, and issues of personal data privacy and transparency, particularly regarding the awareness of human participants that they are engaging with a DTA.

5.1.4 Materializing the Scenarios

Once finalized, the scenarios were brought to life. The details of their materialization are shared below:

1) **DTA 2.0 by Connect**

The DTA concept was translated into a video. This video represents the DTA being showcased to the general public through a fictional company's launch event. The name of the company is *"Connect,"* which features the tagline *"like never before."*

To create this video, firstly, a script was prepared in textual form and then converted to audio using ElevenLabs speech synthesis tool. Once the speech was generated, the audio file was uploaded to Canva video tool. The visuals used in the form of images and video clips were gathered from an online tool called Pexels. Some clips were also taken from Meta AI's video of character based virtual assistants: https://about.fb.com/wp-content/uploads/2023/09/Social-Profiles-for-AI_Header-1.mp4?_=1. These clips were compiled, along with the speech audio file, to create the final video called *"DTA 2.0 by Connect"*. The final video can be viewed here:

https://www.youtube.com/watch?v=XSajuf35-24

2) The Birthday Wish

This scenario was materialized by creating visuals of a DTA generated birthday wish, accompanied by a short story explaining the context. The visual showcases a digital screen presenting a slideshow of memorable photos alongside a personalized birthday wish generated by the DTA. The birthday wish, imbued with a sentimental touch, expresses love, reflects on core memories and experiences, and extends best wishes. The aesthetic of the generated birthday wish mirrors the personal aesthetic preferences of the two friends, Anna and Lisa. While the wish concludes with the greeting *"Lots of Love, Anna,"* a *"DTA Generated"* watermark is visible at the bottom right corner, indicating that the birthday wish was generated by Anna's DTA for Lisa. The visual was compiled on Figma, using visuals from Pexel. The text for the birthday wish was generated using Chat GPT 4. The final visual for *"The Birthday Wish"* scenario can be viewed on the following page in figure 14.

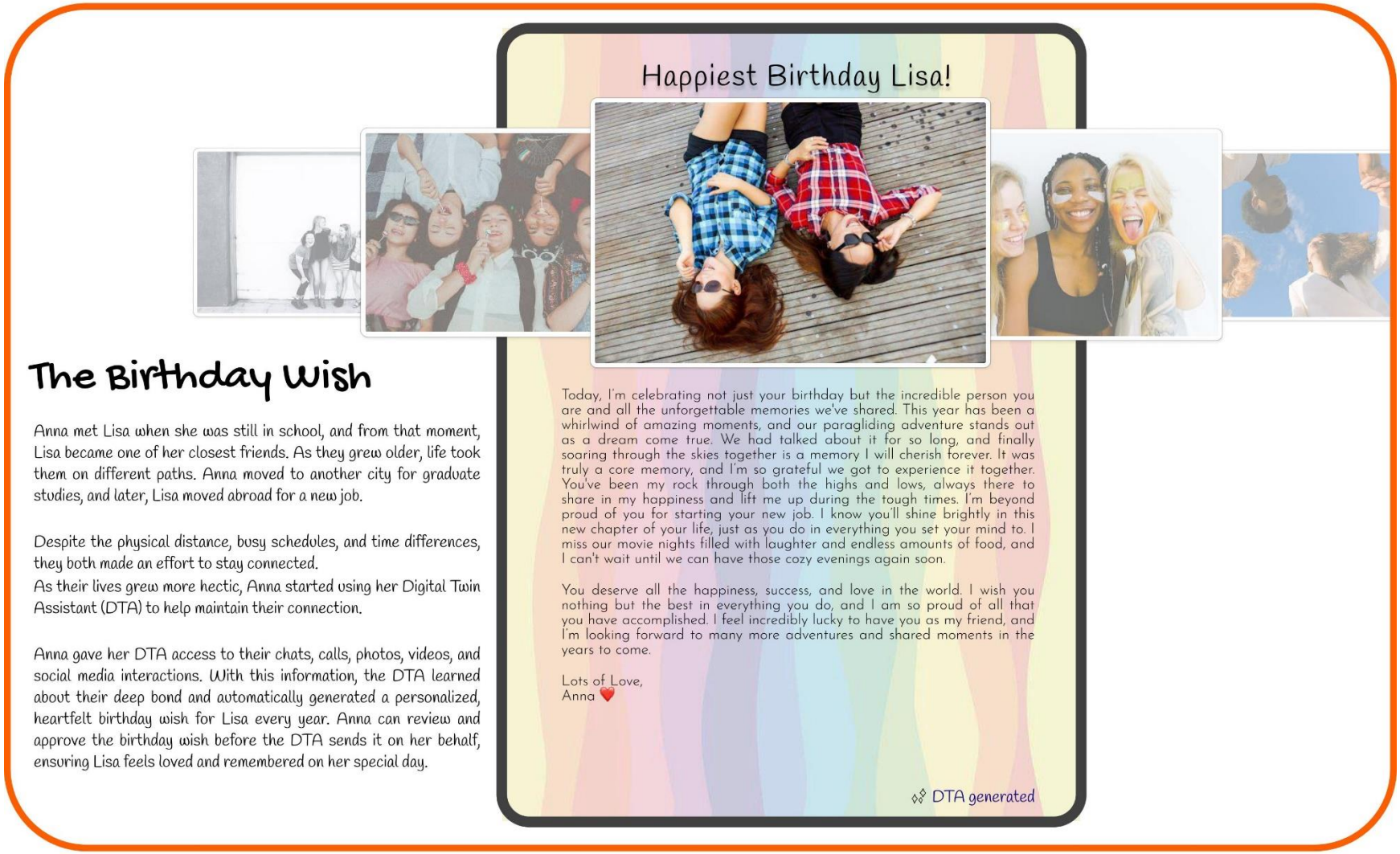


Fig. 14 Visual created for the *"The Birthday Wish"* scenario

3) The Hiking Trip

This scenario is represented through a visual showcasing a digital screen with an ongoing video call. The call displays four small windows, each featuring one of the four friends. Moira and Sabrina appear as real people, while Zayn and Fin are depicted as AI-generated images with plain backgrounds, indicating that their DTAs mimic their facial identities when representing them. Each window is labeled with the participant's name, such as *"Moira,"* and for the DTAs, the label includes *"DTA,"* for example, *"Zayn's DTA."* The DTAs are shown with expressions, smiling alongside Moira and Sabrina. The visual of the digital screen presenting this video call setting is accompanied by a short story explaining the context of the scenario. The visual was compiled on Figma and the AI image generation was done through Playground AI. The final visual for "The Hiking Trip" scenario can be viewed on the following page in figure 15.

The Hiking Trip

Two weeks ago, Zayn, Sabrina, Moira, and Fin scheduled a group call to plan their long-anticipated hiking trip to the mountains. At the last minute, personal and work emergencies prevented Zayn and Fin from joining, so they had their Digital Twin Assistants (DTAs) take their place in the group call.

As the call began, the usual chit-chat quickly shifted to trip planning. Everyone shared their thoughts on which trail they preferred. Zayn's and Fin's DTAs represented their thoughts, ideas, and opinions to their best abilities.

Zayn's DTA constantly disagreed with the others about choosing the easier trail, as Zayn enjoys steeper and more challenging hikes. Moira reminded the group that Zayn had mentioned wanting a more relaxed hike this year, though Zayn's DTA insisted on his usual preference for intense trails.

Meanwhile, Fin's DTA advocated for the riverside trail, appreciating its scenic beauty. Sabrina and Moira also voiced their preferences and disagreements, contributing to a lively discussion.

Despite differing opinions, the group managed to agree on certain points. For the remaining decisions, Sabrina and Moira agreed to wait for Zayn and Fin to get back to them after receiving the call summary from their respective DTAs.

Fig. 15 Visual created for the *"The Hiking Trip"* scenario

5.2 Scenario Delivery

Once the scenarios were developed, they were presented and discussed with participants in semi-structured focus groups. The process of delivering these scenarios to the participants and facilitating focus group discussions is explained below.

Participant Recruitment

To recruit participants for this research, a recruitment poster and a Google Form were utilized. The poster provided a brief overview of the project, while the Google Form collected information on participants' age, academic background, availability for participation, and their general awareness of AI technology trends. These materials were subsequently shared in personal and professional group chats consisting of young adults, as defined in *Chapter 1* of this research. Through this process, a total of *14 participants* were recruited. The participants, *aged between 22 and 29*, represented a *mix of genders* and came from both *design and non-design backgrounds*, with a diverse range of familiarity with current AI trends. In total, *three focus groups* were conducted, with each focus group *1.5 hours* long.

5.2.1 Focus Group Setup

The focus group setup, in which the researcher took the role of facilitator, began with participants reviewing an informed consent form (Appendix B) and a detailed focus group schedule (Appendix C). Each participant was provided with a workbook containing seven distinct worksheets designed to guide the discussion. The worksheet titles were informed by Value Sensitive Design (Appendix D), emphasizing key aspects necessary for understanding and eliciting participant values. Additionally, insights from the framework (Chapter 3), such as the effects on interpersonal social connectivity, were incorporated. These titles structured the discussion to align with the research objectives while allowing participants the freedom to openly express their thoughts. The discussions facilitated by these titles were instrumental in uncovering participant values in relation to the presented context.

The actual worksheets presented to the participants can be found in the (Appendix D). The worksheet titles are listed below:

Worksheet 1: Initial thoughts

Worksheet 2: This makes me feel

Worksheet 3: I like

Worksheet 4: I dislike

Worksheet 5: I fear/worry

Worksheet 6: I dream

Worksheet 7: Overall impact - Relationships with others, Communication and social interactions, Life choices, Social skills and behaviors, Physical and mental wellbeing, Overall society

5.2.2 Focus Group Proceedings

Following a brief introduction and background of the research project, participants viewed the *"DTA 2.0 by Connect"* video on a screen. After watching the video, they were asked to record their initial thoughts on Worksheet 1. This activity served as a brain-dump exercise, which was followed by a group discussion. The discussion provided a platform for participants to openly share their thoughts and immediate reactions to the video.

Then the participants were shown *"The Birthday Wish"* scenario on the screen. After a few moments of immersion in the scenario, where the participants considered themselves interchangeably as both the characters in the scenario, they were asked to complete *Worksheets 2-6*. *Worksheet 2* provided participants with an opportunity to express their feelings about the scenario. Meanwhile, *Worksheets 3, 4,* and *5* facilitated their articulation of aspects they liked, disliked, and feared or worried about. This exercise was followed by a group discussion where participants openly shared their ideas and thoughts. The same process was repeated for the final scenario, *"The Hiking Trip."*

The session facilitated free and open expression, encouraging a natural flow of conversation with the aid of worksheets and semi-structured questions. The activity concluded with participants completing the final worksheet, *Worksheet 7*, where they shared their overall views on all three scenarios combined. This worksheet incorporated sub-titles such *Relationships with others*, *Communication and social interactions, Life choices, Social skills and behaviors, Physical and mental wellbeing,* and *Overall society*. After noting down their answers on the worksheets, participants engaged in discussion for the final time to express their thoughts and conclude the session.

The session was recorded using a recording device and some pictures were clicked using a camera. All the materials including worksheet and focus group schedule print outs, required stationery and snacks were provided to the participants. The focus group set-up is shown in figures 16 and 17, while figures 18 represents how participants engaged with the scenarios presented during the focus groups.

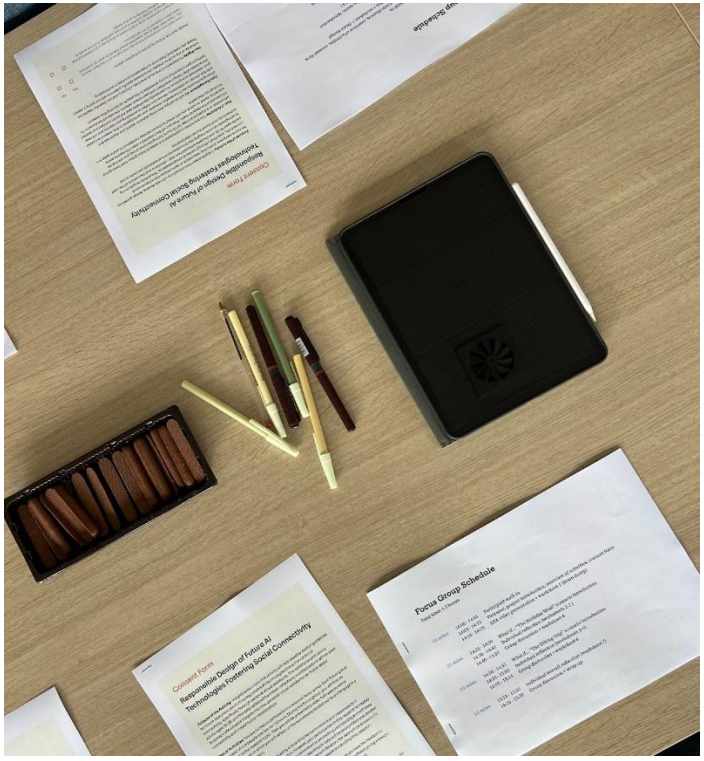


Fig. 16 Focus group set-up

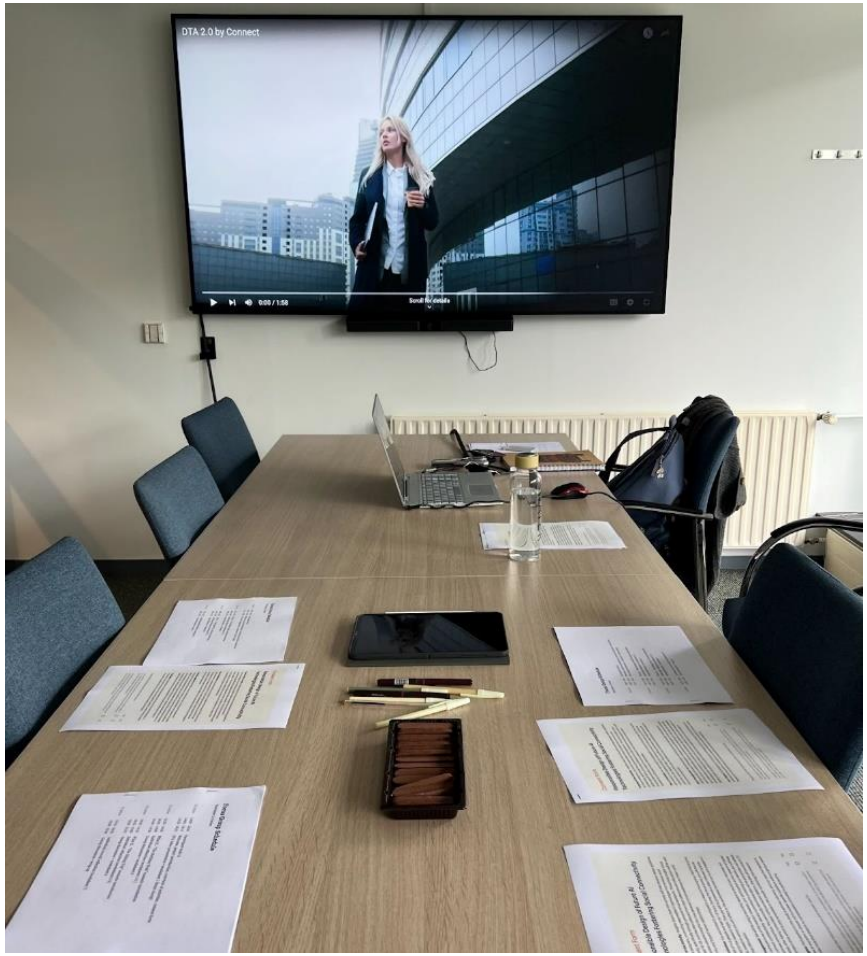


Fig. 17 Focus group set-up

Fig. 18 Participants engaging with the presented scenarios

6. SCENARIO FINDINGS

The scenario-based discussions within the focus groups generated a substantial amount of qualitative data and provided numerous valuable insights for this research. This chapter details the process undertaken to synthesize and organize the data to surface participants' values. Through data analysis, its categorization into value themes, and the identification of contextual values crucial to this project, the chapter ultimately presents a set of values specific to the research context of AI virtual assistants fostering interpersonal connectivity. The contents of this chapter are as follows:

The data collection through the scenario-based focus groups resulted in *400 pages of transcripts*, *140 completed worksheets*, and *4.5 hours of audio recordings*. This extensive dataset was first synthesized and analyzed based on the primary themes of discussion, as guided by the worksheets. The analyzed data was subsequently categorized into value themes as guided by thematic analysis principles (Delahunt, 2017) and Value Sensitive Design (Sections 2.2.2 and 2.3). These themes served as the basis for validating and elaborating on previously identified values, as well as for eliciting new contextual values. The process of value validation, elaboration, and elicitation, and its significance to this research, is detailed in the following sections and represented in figure 19.

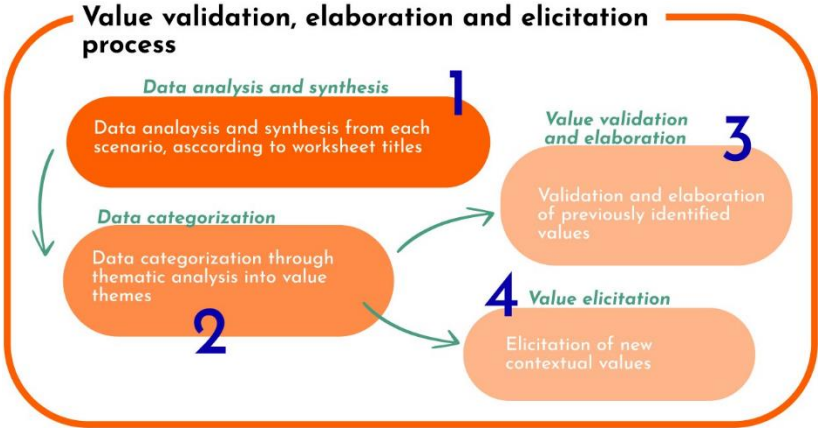


Fig. 19 Process of value validation, elaboration, and elicitation

6.1 Data Analysis

The scenario-based discussions within the focus groups yielded a substantial amount of qualitative data and provided numerous valuable insights for this research. The data collected from participants' worksheet responses and verbal discussions was systematically organized and analyzed in the following sequence:

1. The analysis of the data collected from the first video scenario, as recorded in Worksheet 1: Initial Thoughts, revealed insights across multiple themes. Following the principles of thematic analysis, the data was examined to identify recurring themes, which included: autonomy, control and decision-making, lack of trust in the DTA, context of use, lack of genuine connection, and access to personal data. These themes reflected participants' immediate thoughts and impressions upon viewing the video.

2. The analysis of the data collected from the remaining two scenarios was conducted in alignment with the titles of Worksheets 2-6.

3. The analysis of the data on the overall impact of the three presented scenarios was organized according to the sub-titles of Worksheet 7.

The analysis is reported as follows.

1) DTA by Connect Video

Worksheet 1: Initial Thoughts

The video elicited mixed emotions of excitement and fear. Participants' initial reactions included both enthusiasm for the technology's potential uses and significant ethical concerns, emotions, and criticism regarding its implications. While some participants were excited about the new opportunities and forms of assistance the technology could provide—comparing it to science fiction or fictional assistants like Iron Man's J.A.R.V.I.S. and describing it as *"interesting and very useful"*—the majority expressed apprehension, using terms such as *"scary," "horror," "creepy,"* and *"unsettling"* to describe their feelings. The initial excitement was clearly overshadowed by deeper concerns about the ethical consequences, leading to substantial criticism, as noted by one participant:

"It seems like something superficially amazing and super cool but very scary underneath."

These mixed feelings of excitement and fear were often linked to underlying concerns about the functionality of the DTA. To better understand these reactions, they were categorized according to the recurring themes of discussion as follows:

1. **Autonomy, Control and Decision Making**

Participants conveyed a strong desire to maintain control, often using phrases such as being in the *"driving seat"* rather than the *"back seat."* They found the prospect of the DTA autonomously making decisions based on its interpretation of their preferences to be *unsettling*, describing the potential loss of autonomy and control as *"scary."* This concern was further highlighted as participants questioned their role, the purpose of their existence, and whether they would ever be able to regain control and autonomy, as expressed in their statements:

"If this has to happen, then what is the purpose of our being? What am I going to do?"

"It would be okay if it is taking some decisions and asking for approval in the end because then I will be making the end decision."

"I feel that I take the back seat in this new era of technology. Do I really have autonomy once I give it access? Will I be able to get that control back?"

Participants emphasized the importance of maintaining control and autonomy, asserting that decisions made by the DTA should not be entirely autonomous, but rather subject to their oversight and approval.

2. **Lack of Trust on the DTA**

Participants conveyed significant difficulty and reluctance in trusting the DTA to accurately represent their emotions and decision-making processes. They expressed concerns that the DTA might not possess a sufficient understanding of their personal characteristics, leading to fears that it could fail to act in accordance with their true preferences and intentions. This was explained by a participant as:

"It seems like another version of me just pretending to be me. In real life, it's not really me, it may not know me well enough and maybe the decisions it makes are not the ones I would make."

Participants further articulated that, particularly in the context of emotional conversations, actions, or decisions, they would not trust DTA's ability to respond appropriately. As one participant noted:

"I wouldn't trust it to encapsulate my emotions, like, emotions are driven by time. So, it wouldn't know what emotions I'm feeling at the moment."

Consequently, participants indicated that while they might be comfortable allowing the DTA to facilitate certain social connections, they would not trust it with managing personal, heartfelt relationships.

3. Context of Use

Participants deviated from the context of use explained to them initially to express a preference for using the DTA to manage non-emotional tasks that they do not enjoy personally. They identified a range of mundane tasks—such as setting alarms, drafting and responding to emails, handling professional and formal connections, sending short messages, or managing reminders—that they would prefer the DTA to handle according to their individual preferences and personal choices. In that context, a participant expressed:

"It is quite exciting actually. I can just say okay, handle these three conversations 'diplomatically' or write a reply to them. This way, I will have more time to do the other stuff that I like. I really don't have to focus on drafting the perfect e-mail or things like that, I can just give my DTA instructions, and it can do it for me."

Another participant expressed enthusiasm for the DTA's personalized functionality, highlighting its ability to learn from individual responses and actions. They noted that this capability could make the DTA a highly valuable tool. As one of them stated:

"It is going to have an insight of my personality so it kind of knows me. I have a personality of my own and there are things that I would like it to train on. Like...do not ever say those words or use these kinds of phrases because that represents me a bit more and having that consistency across all channels...that is very exciting. If that can be achieved, I will definitely be a customer."

Hence, in evaluating the features and functionality of the DTA, participants articulated their concerns and fears, while also highlighting specific contexts of use that elicited excitement.

4. Lack of Genuine Connection

Participants indicated that while the DTA facilitates social connectivity, they perceive this connectivity as *"pseudo"* or *"fake."* Participants noted that despite the DTA's ability to learn and be personalized, it cannot replicate the genuine connection that humans achieve. They expressed concerns that this artificial connection could lead to trust issues, affecting their ability to trust both others and their DTAs. As expressed by a participant:

"On a platform where I am looking for a real connection, I wouldn't use it...or to write personal heartfelt messages to others."

Participants highlighted a clear boundary between genuine and artificial interactions, particularly in emotionally sensitive and intimate contexts. They expressed concerns about the appropriateness of using a DTA to handle tasks such as writing condolences to loved ones. Participants feared that such interactions would be perceived as disrespectful and would not be well received if friends or family knew that the responses were generated by a DTA.

5. Access to Personal Data

Participants voiced concerns that achieving a high level of personalization, where the DTA could accurately represent them across digital platforms, would require them to disclose extensive personal information. Additionally, participants questioned the extent of knowledge the DTA would acquire once granted access to various aspects of their lives. This unease was met by broader concerns about current technological practices, where users frequently encounter requests for personal data access from various platforms, often without a clear understanding of what they are consenting to or the implications of such access. As expressed by a participant:

"We're getting to the point where automation is taking over. I feel like even at this point, we get such a bombardment of approvals of things like cookies...simple things. At some point you might say okay yes whatever and then after three months, you realize you said yes to something horrible."

The concerns regarding how the DTA trains and operates on personal data also prompted the participants to reflect on the practices of current social media platforms, which similarly have access to extensive personal data that users often do not fully understand or are unaware of.

2) The Birthday Wish

Like the *"DTA by Connect"* video, this scenario elicited strong reactions from participants and provoked a range of thoughts and reflections. The analysis is categorized according to the Worksheet titles as follows:

Worksheet 2: This makes me feel

Participants expressed mixed emotions as they expressed their feelings about the scenario. They also expressed curiosity about how the scenario might unfold further and questioned whether the birthday wish was the extent of the communication or if it served as a conversation starter between the two characters. Both these situations made them feel different about the scenario. As expressed by a participant:

"If this is like only a conversation starter kind of a thing, I would be okay with it but then...if this is the only communication, then if I were Lisa, I would feel offended."

While the participants acknowledged the ease, efficiency, and convenience the DTA provided for Anna, they strongly emphasized being in Lisa's position, if this was the only communication from Anna's side, would make them feel "*sad,*" *"frustrated," "disappointed," "indifferent," "uncomfortable,"* and *"offended."* They further noted that the sentimental value

of the birthday wish is diminished because it is generated by the DTA rather than being personally crafted by Anna. They compared it to the value of a handwritten note, emphasizing that they would prefer receiving a simple *"Happy Birthday"* text message instead, as it feels more like coming directly from their friend rather than a heartfelt message from a digital assistant. They expressed that as Lisa, they would *feel less valued* and *cherished* as it seems like very little effort from Anna's side. They emphasized the importance of Anna's involvement in crafting the birthday wish, noting that even minimal effort, such as providing prompts to the DTA or asking for suggestions, adds meaningful value to the message. Despite the assistance from the DTA, this personal input was seen as crucial in maintaining the authenticity of the gesture. As expressed by a participant:

"As Anna, I would be okay if I am involved in the process, like using ChatGPT, I can get support from my DTA. It is like buying a Christmas card from the store with a nice message on it, but then also adding a personal touch by writing something yourself."

"If I'm investing my time into expressing my emotions and taking the DTA's help just like using a thesaurus, that is better than being completely AI generated, but not the same as crafting the whole thing myself."

While participants felt more *"at ease"* with the idea of Anna being involved in crafting the birthday wish, they remained skeptical about the presence of a watermark. As Anna, they expressed a desire to remove the watermark; however, as Lisa, they would like to know if the wish was DTA generated, rather than being *"fooled"*. However, participants agreed that messages which are less intimate and more generic would likely be viewed more favorably, even if they were DTA-generated or marked with a watermark. This discussion prompted participants to consider whether they would still feel as comfortable using ChatGPT if it also came with a watermark.

Additionally, participants expressed feelings of *"discomfort"* concerning the extent of personal data the DTA must access to generate something so personal and intimate. This concern extended to all stakeholders involved, as it was noted that when Anna grants her DTA access to her personal data, it inadvertently gains access to Lisa's data as well, without Lisa's consent. The idea of the DTA having access to such personal information left participants feeling *"scared."* This concern also prompted participants to reflect on current social media practices, particularly the extensive access that existing platforms already have to personal data, such as photo galleries and live locations of not just them but also of others through them.

Worksheet 3: I like

Participants acknowledged that AI possesses superior memory capabilities compared to humans. In this context, they appreciated the potential of their DTA to function as a digital memory book, capable of maintaining records of significant interactions with close friends that they might otherwise forget. A participant stated:

"As a human it's very hard to remember everything that has happened, or you've experienced with

someone so at least the DTA would know that you did this together like or something like that and can give you suggestions. So, for a birthday, the DTA sends you memories or pictures which you can look at."

Participants expressed their appreciation for the visual produced by the DTA, noting the aesthetic appeal and overall attractiveness. They also liked aspects of the DTA's personalization features, noting that it enhances ease of use by adapting to their unique style of expression. They appreciated how the DTA not only tailors its communication to align with individual preferences but also generates personalized visuals and sweet messages.

Finally, participants appreciated how the DTA enhances efficiency and facilitates connections by serving as a valuable conversation starter. For instance, it helps rekindle communication with friends, such as through sending birthday wishes, thereby encouraging engagement and maintaining connections.

Worksheet 4: I dislike

Participants generally disliked the fact that the birthday wish was generated by AI rather than a human. According to them, that dehumanizes the connection. They felt that the connection felt less genuine and valued compared to a message crafted entirely by a human. They articulated these sentiments by stating:

"It would feel as if someone else is putting the words in my mouth or my boyfriend's assistant is buying me a present instead of him. The effort put in holds great value for me."

Participants also expressed concern that the use of AI-generated messages could lead to stressful and awkward situations, particularly for recipients who might feel undervalued. They argued that such interactions could foster emotional disconnection rather than genuine connection, undermining the value of authentic human engagement. They emphasized the importance of making an effort for those they care about, suggesting that a simple, personalized message written by a human holds more meaning. Participants indicated a preference for not engaging in what they perceived as *"empty"* connections through AI alone, advocating instead for people to also put in personal effort in maintaining connections with people they care about.

Worksheet 5: I fear/worry

A major concern expressed by participants was the issue of personal data and privacy, particularly regarding consent. Participants highlighted that granting a DTA access to personal data not only affects the data owner but also inadvertently exposes others' data stored with them, without their consent. For example, as recipients of a birthday wish, participants were troubled by the prospect of their personal data being shared by their friends' DTAs without their explicit permission. This raised significant apprehensions about the broader implications for privacy and data security. Similarly, participants expressed concerns about the DTA's ability to function effectively with limited data access. They questioned whether the DTA could still provide meaningful personalization if its access to personal data was restricted.

Finally, a significant concern voiced was the potential normalization of using DTAs for social connectivity.

Participants perceived it as dangerous if such technology became mainstream, fearing that it could lead to shallow relationships and weaken the strength and authenticity of human connections, ultimately eroding essential human values. Participants worried that people might become reliant on DTAs, reducing personal effort and meaningful interaction in their relationships. They expressed concern that individuals could fall into a cycle of receiving DTA generated messages and responding with their own DTA, thereby diminishing genuine human connection. One participant explained this by saying:

"I fear that people would normalize having conversations using their DTAs and soon only DTAs will be talking to each other."

Worksheet 6: I dream

Participants expressed a strong desire for greater control over the DTA's functions. They emphasized the importance of being actively involved in decision-making. Specifically, when it comes to interpersonal connections and associated ethical concerns, participants preferred that the DTA send reminders or notifications of upcoming birthdays instead of sending a pre-generated birthday wish. This approach would serve as a prompt to connect with someone they care about, encouraging more personal engagement.

To support this idea, participants suggested that the DTA should offer various options for creating a birthday message, including showcasing pictures and memories from the past, effectively functioning as a memory bank. This would allow users to select meaningful memories themselves and incorporate them into personalized birthday messages. Participants felt that this method would make the DTA a valuable tool while ensuring they remain the primary decision-makers and continue to put personal effort into their interactions, with one participant noting:

"I would want the DTA to be a memory bank, my DTA can create reminders, like my DTA can remind me that Frank was going to leave for the US, did you check up on him? DTA stays more in the background, but I control the connectivity in the foreground."

Additionally, participants envisioned a DTA that could adapt its communication style to suit different relationships, offering options to generate text in various tones—such as humorous or sarcastic for one friend, and respectful or serious for another. Participants believed that this capability would enhance the DTA's ability to facilitate more personalized and authentic connections, tailoring interactions to better reflect on their relationships.

Finally, participants envisioned the DTA as a tool that could alleviate the fatigue associated with coordinating social interactions, such as scheduling phone calls with friends across different continents and time zones. They highlighted the challenge of maintaining connections with long-distance friends. Participants expressed a desire for the DTA to assist in fostering real connections by encouraging and supporting actual human-human communication, rather than fully automating it. However, they noted that greater automation might be more acceptable in professional or non-intimate relationships.

3) The Hiking Trip

This scenario also elicited strong reactions from participants and provoked a range of thoughts and reflections. The analysis is categorized according to the Worksheet titles as follows:

Worksheet 2: This makes me feel

The participants acknowledged that while it may be *"convenient", "helpful"* and *"useful"* for Zayn and Fin to send their DTAs to attend the call on their behalf, for the human participants, Moira and Sabrina, they expressed feelings of *"sadness," "frustration," "annoyance,"* and *"disappointment"*. Participants expressed concerns that their friends sending their DTAs instead of showing up themselves would make them feel *"less appreciated,"* leading to the impression that maybe they were not genuinely interested in planning the trip. They further emphasized that engaging with a friend's DTA during a call, especially in discussions or disagreements, would be *frustrating* and *undesirable*. Moreover, participants shared that if they were the only human in a group of DTAs, they would feel *"scared."* In such a situation, they would prefer to send their own DTA to the call and let the DTAs communicate instead. One participant explained this sentiment by stating:

"If I was the actual human, I'd feel frustrated because I would know something my friend said that his DTA doesn't know. So, I think it would be really tiring to go back and forth with the DTA. I really do not want to argue with the DTA. "

Participants also reported feeling *"self-conscious"* in the presence of DTAs that were recording the call and generating summaries. This raised concerns about the integrity of their personal data and privacy. One participant explained this concern by stating:

"I'm not sure if I would speak freely, I would feel conscious, at the back of my mind I'd constantly have this thought that a DTA is present in the call, I would be concerned about my privacy."

Participants expressed feelings of *"insecurity"* and *"uncertainty"* about allowing their DTA to represent them, particularly due to their lack of control once the DTA joins the call. They were concerned about not knowing what their DTA might say to others during the conversation. One participant explained this by saying:

"My DTA might just create confusion by saying something I personally might not say, making the discussion and planning more difficult for my friends."

Another participant added on by stating:

"I'm not sure if the DTA can read the room and respond accordingly, the way humans react and respond to a conversation, so it may not agree to anything. If I am talking to a real person, they would be prone to getting convinced or the conversation might go differently."

Worksheet 3: I like

The most common response noted was participants' appreciation for the summary provided by the DTAs to the people absent in the call. They stated how it creates

convenience and ease for them if they are busy or unavailable, they can be filled in by their DTAs and do not miss any important updates or information. A participant expresses this by saying:

"Instead of being absent, it is nice that someone can attend the call on my behalf."

Participants collectively agreed that, despite some disagreements, the call effectively advanced the trip planning process. They appreciated how the DTAs contributed to a more efficient and organized approach, particularly in handling the often time-consuming and hectic aspects of planning. Participants valued the DTA's ability to share options, ideas, and suggestions tailored to their preferences, which can make the process of planning and organizing trips easier and less stressful. Additionally, they expressed how such assistance from the DTA allowed them to delegate certain digital interactions, freeing up time for other activities. A participant noted:

"This will reduce the amount of time we spend digitally because those interactions can be more automated, and I will have more time to spend on physical interactions and activities."

Worksheet 4: I dislike

Participants primarily disliked the fact that the involvement of DTAs in the conversation would disrupt the natural and flexible flow of communication, making it feel less organic. Participants further explained that DTAs lack the emotional sensitivity and ability to negotiate that are essential in human conversations. They expressed concern that DTAs might fail to consider others' perspectives, leading to repetitive and less meaningful interactions. A participant elaborated:

"If I am having a conversation with you, it would happen very organically and easily. When you throw a DTA in the mix, that can mess up the whole flow. What if I suggest a very nice trail and you actually like it and change your mind?"

Participants also emphasized that they would strongly prefer not to argue with a DTA, as this could lead to confusion and disrupt the natural flow of conversation. They expressed concerns about not feeling heard, particularly in situations where they possess information that the DTA lacks.

Finally, when considering the possibility of being represented by their DTA, participants were uncomfortable with the idea of the DTA automating their decisions. They worried about the DTA expressing their thoughts or making choices on their behalf, especially if their preferences change over time. They preferred to maintain control over their decisions rather than having the DTA act on their behalf.

Worksheet 5: I fear/worry

Participants voiced concerns about the potential for long-term misuse of the DTA, suggesting that people might increasingly use it as an excuse to avoid attending meetings themselves. They feared this could have a detrimental effect on interpersonal relationships. A participant stated:

"People will start coming up with excuses to skip conversations and send their DTAs instead."

Similarly, participants expressed concerns that their DTA might not accurately convey their thoughts and ideas to others. They also worried that conversations would lose their natural flow, as DTAs lack the soft skills and subtle understanding that humans possess, such as the ability to read the room and respond appropriately. Consequently, participants were hesitant to trust their DTA to engage in conversations on their behalf.

Participants further voiced concerns about being the only human among multiple DTAs on a call, expressing a preference to avoid such situations altogether and instead allow their DTA to participate on their behalf. Additionally, they expressed apprehension about the DTA's access to personal data and its potential to collect and summarize information from calls. This access could limit personal expression and emotional depth, leading individuals to communicate more directly.

Worksheet 6: I dream

Participants expressed a strong preference for having the DTA assist in planning and organizing trips by coordinating schedules and preferences with their friends. They appreciated the idea of the DTA providing suggestions and generating several options and itineraries. However, they preferred that the DTA refrain from making final decisions. Instead, participants wanted the DTA to handle tasks such as finding optimal flight and accommodation options within their budget, thereby alleviating the time-consuming aspects of trip planning and allowing them to focus on enjoying their time with friends. Additionally, participants preferred a scenario where either all participants were humans or all were DTAs, avoiding a mix of both.

Participants indicated a preference for the DTA to adopt a passive role during calls. They would like the DTA to be present in the call primarily to share general information such as their availability. Additionally, some participants suggested that before joining the call, they could provide their preferences to the DTA, which would then communicate these preferences to the other participants. Given that they did not attend the call, participants expressed that they would be comfortable with the decisions made by the group based on their conveyed preferences. A participant further explained this:

"I would like it to be very natural, as if it was a real discussion. If I did not join, then one of my friends would update me on the discussion. I would not want my DTA to argue but just convey my preferences and let the humans make a decision."

Finally, some participants expressed a preference for maintaining control over the DTAs present in a phone call. They indicated that having authority over the DTAs would facilitate smoother decision-making and ensure that they could effectively manage the interaction. A participant expressed:

"I would like a Trump card which enables me to override my friend's DTA, telling it that hey the human version of you agreed to do something else. This would make the discussion more efficient and effective."

4) Collective Scenario Analysis

Worksheet 7: Overall Impact

After discussing their thoughts on the "DTA by Connect" video, as well as "The Hiking Trip" and "The Birthday Wish" scenarios, participants shared their overall perspectives on the potential impacts of this technology on interpersonal connectivity. The analysis is detailed below:

1. **Relationship with Others**

Participants expressed concerns that this technology could negatively impact their relationships with loved ones due to a perceived lack of authenticity in the interactions. They noted that the automation, decision-making, and proactive functions of the DTA could blur the lines between conversations conducted by DTAs and those held by actual people. This ambiguity, they feared, could lead to a lack of trust, emotional detachment, and a sense of distance from those who matter most to them. One participant elaborated on these concerns, stating:

"I think it will make us weak in communication. It would cause emotional detachment and awkwardness when we meet in person. Eventually, it would just be people's DTAs talking to each other."

One participant further expressed the opinion that the DTA is unlikely to enhance their relationships with others; instead, it could either maintain existing relationships or potentially harm them.

Conversely, participants also recognized the potential positive impact of the DTA on their relationships. They noted that the DTA could facilitate connections by providing reminders, synchronizing schedules, and assisting with activity planning, thereby enabling people to spend quality time together without the associated logistical challenges. As long as the DTA's role remains supportive rather than controlling, participants felt it could help strengthen interpersonal bonds.

2. **Communication and Social Interactions**

Participants noted that if the DTA serves as a conversation starter or sends reminders to connect with friends, it could enhance social interactions with loved ones. They also suggested that the DTA could promote more in-person communication, such as by scheduling phone calls with distant friends. However, they expressed concerns that communication mediated by the DTA might lack trust and feel inauthentic. Ultimately, participants agreed that by delegating more digital interactions to the DTA, they would have more time to engage in meaningful, in person connections and activities they enjoy.

3. **Life Choices**

Regarding the impact on life choices, most participants expressed that assistance from the DTA would be beneficial and create ease. They appreciated the idea of receiving suggestions, recommendations, reminders, and notifications about memories or trip agendas, as long as they retained control over their final decisions. Participants welcomed the prospect of the DTA handling tasks they found boring or

stressful, allowing them to better connect with their loved ones while still maintaining autonomy over important life choices.

4. Social Skills and Behaviors

Participants had differing views on the impact of the DTA on social skills and behaviors. Some expressed concerns that reliance on the DTA could cause people to gradually lose their social skills, as they would delegate many social interactions to the DTA. This, in turn, could make in-person meetings feel awkward and unnatural. As one participant explained:

"People will become more dependent on their DTAs to handle social communication, which will make them more antisocial and introverted."

Others, however, focused on the potential positive aspects of the technology. They noted that the technology could facilitate engaging and enjoyable conversations, particularly for individuals who are naturally introverted or shy, helping them become more involved and expressive in social interactions. For those who struggle with prioritizing, managing, scheduling, and planning, the DTA could offer valuable support, ultimately enhancing their social connections and helping them stay more connected with their loved ones.

5. Physical and Mental Wellbeing

Participants reflected on the potential impacts of the DTA on human well-being, with a particular emphasis on mental rather than physical health. One participant expressed concern that relying too heavily on the DTA for social interactions could lead to feelings of loneliness. Specifically, if their friend's DTA joined a call or sent messages on behalf of their friend, they would feel frustrated, sad, and mentally disconnected from that friend.

Conversely, other participants noted that the DTA could alleviate stress and anxiety related to remembering and planning activities. By offering suggestions and assisting in decision-making, the DTA could enhance their social connectivity and help them express themselves more effectively.

6. Overall Society

Participants expressed concerns about the impact of DTAs on overall society, highlighting a potential shift toward increased dependency on these tools for managing social connections. They noted that this reliance could lead to a more disconnected society, where the depth and authenticity of social interactions diminish, resulting in feelings of emptiness and superficial relationships. While they acknowledged that society might gradually adapt to these changes, they also emphasized the possibility that people could become more aware of the risks associated with digital social networks and recognize the erosion of traditional social values.

6.2 Data Categorization

The focus groups provided valuable insights, guided by the topics outlined in the worksheets, which were designed to elicit discussions about participant values regarding AI mediated interpersonal connectivity. The sessions were meticulously designed to avoid explicitly mentioning any specific values. Instead, they were structured to encourage open and relevant dialogue, allowing participants to express their thoughts freely without steering them toward any direction. The aim of these discussions was to understand and elicit participant values. Consequently, the insights gathered, as detailed in the previous section, were analyzed and categorized according to thematic analysis principles (Delahunt, 2017) and value elicitation guided by value sensitive design (Sections 2.2.2 and 2.3). This thorough analysis and organization of the data based on value themes enabled the elicitation of specific contextual values that emerged from the discussions.

As discussed earlier in Section 4.2, this research adopts a working definition of value centered on human values, understood as *"what is important to people in their lives, with a focus on ethics and morality"* (Friedman et al., 2017). Using this definition as the basis for categorizing the analyzed data, recurrent themes identified as contextual values were established. The scenarios provoked discussions that revealed what is important to participants regarding their connections with close friends and emotionally significant bonds. These discussions unearthed their feelings, concerns, fears, and dreams, all within the context of understanding their values as framed by this research. For instance, concerns about data and privacy were grouped together under the heading *"Privacy and data governance"*. These concerns were then translated into general insights regarding the use of personal AI virtual assistants for interpersonal social connectivity. Figure 20 and figure 21 illustrate the clustering of the data gathered into emerging value themes.

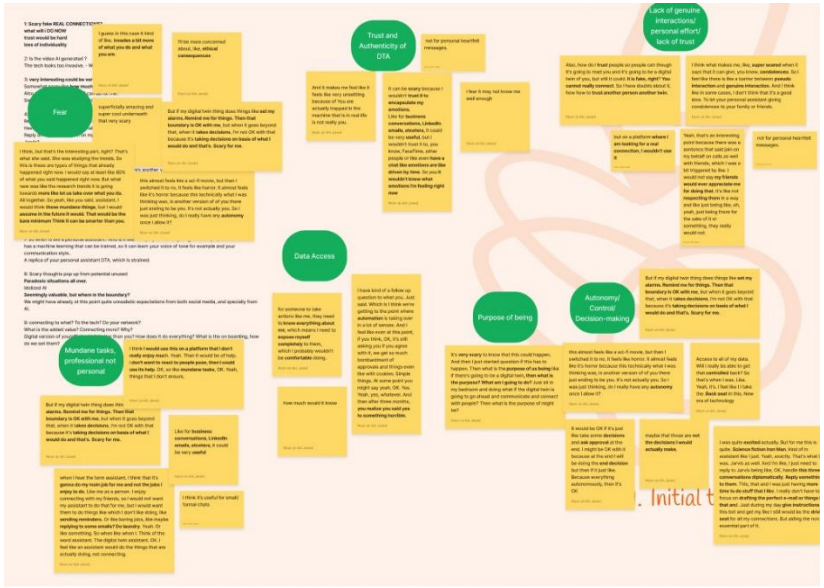


Fig. 20 Clustering analyzed data into value themes

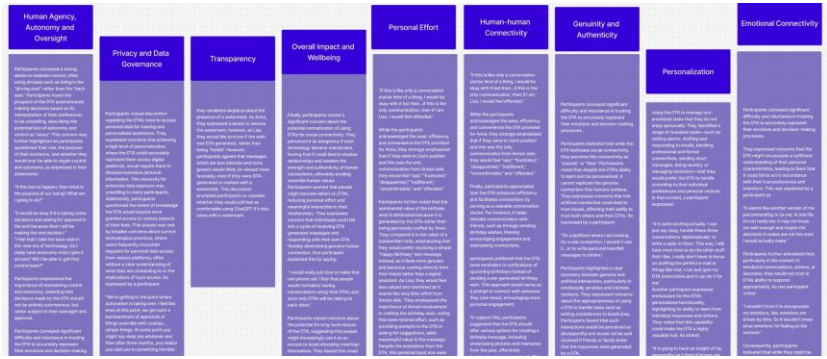


Fig. 21 Clustering analyzed data into value themes

This data categorization and synthesis was based on the following contextual aspects, as defined previously in this research:

1. The definition of value adopted in this project as: *what is important to people in their lives, with a focus on ethics and morality*
2. Specific contextual focus on the development of personal AI virtual assistants
3. Specific contextual focus on interpersonal connections, emphasizing the involvement of emotional and personal bonds, such as close friendships and family relationships
4. Research focus on the development of responsible AI virtual assistants fostering interpersonal social connectivity by understanding the contextual values of young adults

The categorization of value themes revealed *nine contextual* values relevant to this research. These values were further divided into two groups, as illustrated in Figure 22, which lists the elicited contextual values under their respective value groups. The division of these groups and the insights they provide for this research are explained in detail below.

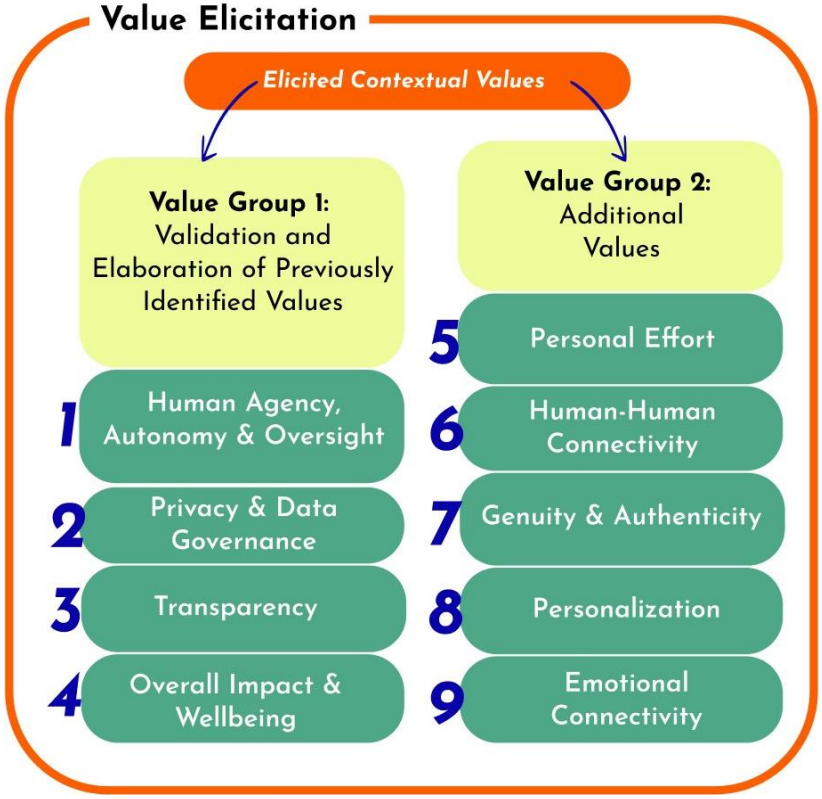


Fig. 22 Groups of contextual values gained through data categorization

1) Value Group 1: Validation and Elaboration of Previously Identified Values

This set of values represents those that emerged from the data analysis and aligned with the four contextual values previously identified in Section 2.3.2. These values were presented in the *framework* outlined in Chapter 3 and their significance for this research was further detailed in the methodology sections 2.2.2 and 2.3. The emergence of these values not only validates their importance for this research but also expands upon their significance within the research context. The data gathered from participants demonstrated that, in the context of AI virtual assistants fostering interpersonal connectivity, these identified values are indeed crucial and are interpreted and conceptualized for this research. Additionally, the findings further elaborated on the specific relevance of these values to this research, providing a deeper understanding of what is at stake and highlighting the major concerns. The values are detailed below.

1. Human Agency, Autonomy & Oversight

The first value that emerged from the data aligns with *"Framework Segment 1 - Human Agency, Autonomy, and Oversight,"* as presented in *Chapter 3, Section 3.1.1*. This connection serves to validate and further elaborate on *"Human Agency, Autonomy, and Oversight"* as a critical value in the context of AI-driven social connectivity. The data analysis also contributed to conceptualizing this value and interpreting it within the specific context of this research, which focuses on the responsible design of future AI virtual assistants aimed at fostering interpersonal social connectivity. Thus, concerning *"Human Agency, Autonomy, and Oversight"* within the given context, the following points were noted:

- Having autonomy and control over decisions made by personal AI virtual assistants, particularly in contexts that involve fostering interpersonal social connections.
- Being in the "driver's seat," actively making final decisions rather than passively allowing their personal AI virtual assistants to act on their behalf.
- Having the ability to monitor and oversee decisions made by their personal AI virtual assistants.
- While it is deemed acceptable for personal AI virtual assistants to make preliminary decisions based on user preferences, users should retain the power to approve and finalize those decisions, ensuring that the personal AI virtual assistant remains an assistive tool rather than an autonomous decision-maker.
- Hesitance amongst participants to rely on highly personalized and human-like AI virtual assistants to accurately represent their choices, intentions and decisions, without their involvement. Personal AI virtual assistants may not fully grasp the nuances of users' personal characteristics, which could result in decisions, actions, or responses that do not align with their true preferences and intentions. This can lead to feelings of insecurity and uncertainty.

- Preference to guide their personal AI virtual assistants by clearly defining their choices and intentions. Allowing AI to take on a more passive role, primarily sharing information as directed by them, rather than actively participating in discussions or making decisions on their behalf.

- Having authority in situations where they have knowledge that a personal AI virtual assistant lacks. The ability to override AI is essential for enhancing efficiency and effectiveness, ensuring that users are not placed in a position where they must debate or justify their decisions to the technology.

2. Privacy and data governance

The second value that emerged from the data aligns with *"Framework Segment 2 - Privacy and Data Governance,"* as presented in *Chapter 3*, *Section 3.1.2*, validating *"Privacy and Data Governance"* as a critical value in the context of AI-driven social connectivity. Based on the participants' responses, the ensuing discussion, and the concerns that emerged, this research elaborates on the value of *"Human Agency, Autonomy, and Oversight"* within the given context and draws the following points:

- Concerns about the necessity to disclose extensive personal information to a personal AI virtual assistant in order to achieve a high level of personalization, enabling it to accurately represent users across digital platforms.

- Concerns about the depth of knowledge a personal AI virtual assistant would gain once granted access to various aspects of a user's life.

- Users being aware of how a personal AI virtual assistant trains and operates on personal data.

- Personal AI virtual assistants inadvertently access data belonging to others without their consent, stored within the users' accounts, when they grant it the permission to access their own personal data.

- The capacity and limitations of a personal AI virtual assistant to operate effectively with restricted data access while still delivering meaningful personalization.

- Concerns about the recording and storage of data during interactions involving others, and the subsequent sharing of this information by personal virtual assistants with their users.

3. Transparency

The third value that emerged from the data aligns with *"Framework Segment 3 - Transparency,"* as presented in *Chapter 3*, *Section 3.1.3*, validating *"Transparency"* as a critical value in the context of AI-driven social connectivity. For this value, based on participant discussion, this research draws the following points:

- Skepticism surrounding the use of indicators, such as watermarks, that inform others when content has been

generated or tasks have been accomplished using a personal AI virtual assistant.

- Having the ability to decide whether they want others to know if their personal AI virtual assistant was involved in completing a task.
- Having the ability to control whether others are aware if they are interacting with the user directly or with their personal AI virtual assistant.
- Knowing whether a personal interaction, particularly one with emotional or sentimental significance, was conducted with another user directly or with their personal AI virtual assistant.

4. Overall impact and wellbeing

The fourth value that emerged from the data aligns with *"Framework Segment 4 - Overall impact and wellbeing,"* as presented in *Chapter 3*, *Section 3.1.4*, validating *"Overall impact and wellbeing"* as a critical value in the context of AI-driven social connectivity. Key points relevant to this value that emerged from the focus groups are listed below:

- Potential normalization of using personal AI virtual assistants for social connectivity.
- Concerns about the mainstream adoption of personal AI virtual assistants, including the potential for fostering shallow relationships and undermining the depth and authenticity of human connections.
- Dependency on personal AI virtual assistants would lead to reduced personal effort, ultimately resulting in fewer meaningful interactions.
- Concerns about communication and interactions being mediated solely by personal AI virtual assistants, potentially leading to a reduction in interactions between actual users.
- Concerns about users misusing personal AI virtual assistants to intentionally avoid direct interactions with others by having the assistants represent them, potentially harming interpersonal connections.
- Concerns about losing trust, experiencing emotional detachment, and feeling distant from loved ones due to the use of personal AI virtual assistants for personal and emotional interactions.
- Using personal AI virtual assistants for personal and emotional interactions can potentially harm existing connections rather than enhancing them.
- Enhancement of interpersonal bonds through the support of personal AI virtual assistants in facilitating user connections with others *(e.g. planning, scheduling, reminding, handling logistical challenges etc.).*
- Personal AI virtual assistants can enable people to have more time and energy to devote to activities they enjoy, with the help from their personal AI virtual assistants, positively affecting their mental and physical well-being.

- Personal AI virtual assistants can strengthen interpersonal bonds by enabling more in-person interactions for their users.
- Personal AI virtual assistance enhances user comfort and ease by providing tailored assistance.
- The delegation of social interactions to personal AI virtual assistants may result in a decline in social skills among users, leading to antisocial and awkward behaviors.
- Personal AI virtual assistants can assist shy or introverted users in facilitating social connections by helping them initiate interactions.
- Relying on personal AI virtual assistants for social interactions may result in users feeling lonely and isolated.

2) Value Group 2: Additional Values

The second group of values, identified through thematic analysis of the focus group insights, highlighted the newly discovered context specific values. These values were derived using the definition of value outlined in *Section 2.2*, focusing on what matters to participants in their interpersonal connections, such as close friendships. Given that interpersonal connectivity involves emotional expressions, sentiments, and bonds, participants' specific needs, expectations, feelings, fears, and aspirations emerged. The insights analyzed in Section 6.1 revealed new values pertinent to the context of this research, highlighting additional significant aspects related to the development of AI virtual assistants that foster interpersonal social connectivity. These values extend beyond the general responsible AI principles or values identified earlier. These values are detailed below.

1. Personal Effort

Personal effort was identified as a significant value due to its crucial role in sustaining interpersonal connections and strengthening bonds within close-knit groups. This value is detailed according to the analyzed data as follows:

- Concerns regarding whether users are entirely dependent on their personal AI virtual assistants for social connectivity, potentially neglecting personal effort in maintaining relationships.
- Concerns regarding communication only being done through personal AI virtual assistants
- Concerns about the potential loss of sentimental value when tasks are performed by AI rather than by a human. Participants felt that a simple message sent by a human carries more value and feels more genuine and personal compared to an intimate message conveyed through AI.
- Even minimal human effort enhances the value of the experience involving AI assistance and ensures the authenticity of the gesture.

- Enhancement of communication with AI virtual assistants can be achieved by encouraging humans to add a personal touch to what AI generates
- Importance of being involved in the process, investing personal time to express emotions when taking assistance from AI virtual assistant
- Using an AI virtual assistant to generate content, while retaining final approval and personalization by a human, is preferable to having the AI generate and send content autonomously.
- The importance of making an effort for people you care about is significant; a simple, personal message from a human carries more meaning than something generated by AI.
- Even when using AI virtual assistants, individuals should also invest personal effort in maintaining connections with those they care about.
- Relying solely on an AI virtual assistant for communication with loved ones may lead to feelings of being less appreciated, as it lacks the personal effort and time invested.

2. Human-Human Connectivity

Human-human connectivity was identified as a crucial value because of its fundamental role in fostering meaningful relationships. It underscores the importance of direct engagement between individuals. This value is detailed as follows:

- Concerns about communication occurring exclusively through AI virtual assistants, without direct human interaction.
- AI virtual assistant providing conversation starters would be useful as that facilitates and initiates direct human interactions.
- Preference for using the AI virtual assistant to send suggestions, reminders or notifications to connect with people, rather than having it handle the communication itself, encouraging more personal engagement and direct interaction with those they care about.
- Importance of direct human communications fostering genuine connections rather than fully automated communications through personal AI virtual assistants.
- Concerns about being the only human in a conversation with multiple virtual assistants, with a preference to avoid such situations altogether and, if necessary, to have conversations solely facilitated by AI virtual assistants.

- AI virtual assistants to enhance human connection by facilitating direct interactions and providing support to help people engage more effectively in those interactions.

3. Genuity and Authenticity

Genuity and authenticity refer to the importance of connections that involve the sincere expression of emotions and honest, real engagement. This value is detailed as follows:

- Reluctance in trusting AI virtual assistants to accurately represent emotions and decision-making processes.
- Concerns about the connectivity facilitated by AI virtual assistants being perceived as fake or pseudo stemming from their inability to replicate the genuine connections that humans can create.
- Concerns about AI virtual assistants could create trust issues between people due to the lack of authenticity in their emotional expressions.
- Reluctance to use AI virtual assistants for heartfelt or emotional conversations with others.
- Recognition of a clear boundary between genuine and artificial interactions, particularly in sensitive and emotional contexts. For example, using an AI virtual assistant to send condolences could be seen as disrespectful and inappropriate.
- Concerns regarding the communication becoming dehumanized, less genuine, and less valued when facilitated by AI virtual assistants, due to reduced human involvement.
- Concerns regarding AI virtual assistants creating awkward and stressful situations, as people might feel emotionally disconnected when they realize the interaction is being mediated by another virtual assistant rather than a human, realizing a lack of genuine human involvement.

4. Personalization

This value refers to the uniqueness of each social interaction and the importance of engaging in ways that align with individual preferences. The value is detailed as follows:

- Preference for using AI virtual assistants for non-emotional and mundane tasks, such as setting alarms, drafting and responding to emails, managing professional and formal connections, sending brief messages, and handling reminders. This allows individuals to allocate more time to meaningful interactions with others.
- Excitement about AI virtual assistant's ability to learn from individual responses and actions, and to provide tailored assistance accordingly.

- Usefulness of AI virtual assistants in their ability to learn from individual personalities and maintain consistency, enabling people to connect with others according to personal preferences.
- The superior memory capabilities of AI virtual assistants are highly useful for fostering connections. These assistants can function as digital memory books, sharing memories, sending reminders, maintaining records, and highlighting important occasions, thereby facilitating better connections between people.
- AI virtual assistants can enhance ease of use by adapting to unique styles of expression, allowing users to tailor their communication according to their preferences and needs.
- Appreciation for AI virtual assistants' ability to generate personalized content tailored to user preferences.
- Ease of use is enhanced when AI virtual assistants adopt unique styles of expression and tailor communication to align with users' preferences.
- AI virtual assistants can enhance user connections by making suggestions, offering options, providing recommendations, and sharing ideas, while allowing users to make final decisions rather than automating the entire process.
- The need for AI virtual assistants to remind users to check in with their friends, rather than automatically sending messages on their behalf.
- AI virtual assistants can alleviate stress by assisting with tasks such as scheduling calls at convenient times for everyone, thereby enabling users to connect more effectively with their friends. By facilitating these arrangements, virtual assistants help ensure genuine interactions between people.
- Appreciation for AI virtual assistants being able to step in for users during their absence. However, users prefer these assistants to take a passive role rather than being actively involved.
- Preference for AI virtual assistants to assist with planning and organizing, allowing people to have more time to genuinely connect with each other and engage in activities they enjoy.

5. Emotional Connectivity

This value emphasizes the significance of emotional bonds, empathy, shared feelings, and mutual understanding in social interactions. The value is detailed as follows:

- There is reluctance to trust AI virtual assistants to accurately represent emotions.
- Concerns regarding AI virtual assistants may lack sufficient understanding of users' personal characteristics, leading to fears that they could fail to act in alignment with users' true preferences and intentions.

- Concerns about AI virtual assistants making choices and decisions that users might not make themselves.
- Lack of trust in AI virtual assistants' ability to respond appropriately to emotional conversations, actions, or decisions.
- Lack of trust in AI virtual assistants' ability to accurately capture and convey emotions in the moment.
- Hesitance towards AI virtual assistants managing emotional, personal, and heartfelt engagements.
- Hesitance exists toward engaging in discussions or arguments with AI virtual assistants due to their lack of emotional abilities and soft skills.
- Preference for allowing AI virtual assistants to communicate with each other in situations where one person is interacting with multiple AI virtual assistants.
- Concerns about AI virtual assistants' inability to be persuaded or convinced, unlike real conversations with humans.
- Concerns about AI virtual assistants disrupting the natural and organic flow of conversation due to their lack of sensitivity and negotiation skills, resulting in less effective and meaningful interactions.

7. IMPLICATIONS FOR DESIGN

This chapter presents the culmination of this research in the form of "Implications for Design," synthesizing the insights gathered from previous chapters. Through rigorous literature review, exploration of various design approaches and methodologies, scenario design, focus group discussions, and thorough data analysis, this research has developed a set of "Implications for Design." These guidelines are intended to assist designers in creating responsible AI virtual assistants that enhance interpersonal social connectivity.

7.1 Introducing Implications for Design

7.1 Introducing Implications for Design

This chapter serves as the culmination of the extensive research process detailed across the preceding chapters, spanning over a hundred pages from Chapter 1 through Chapter 6. The primary objective of this research—to generate design implications that guide designers in responsibly designing future AI technologies that foster interpersonal social connectivity—is realized here. To effectively conclude the research activities and present these design implications, the study adopts the principles of values hierarchy (Van De Poel, 2013) and introduces nine individual pyramids of values as key implications for design.

These pyramids of values consist of the following:

1. **Value:** The context-specific values identified and elicited in this research, emphasizing what is important to people, along with the overarching ethical and moral considerations.

2. **Value Definition:** The context-specific definition of the value that outlines what it represents and its particular significance within the scope of this research.

3. **Value principle:** The broad set of rules or guidelines derived from the value, that provide a direction

4. **Design Norms:** The prescriptions or restrictions of action that refer to the specific goals to be achieved or the limits to be observed in pursuit of those goals.

Finally, the nine pyramids of values developed through this research are presented below. These pyramids are designed to guide the responsible design of future AI virtual assistants aimed at enhancing interpersonal social connectivity.

Implications for Design

Human Agency, Autonomy & Oversight

Value

The ability and freedom to monitor and control decisions and actions.

Value Definition

Ensure that users maintain an active role in the decision-making process, with the ability and freedom to monitor and control the actions of their personal AI virtual assistant.

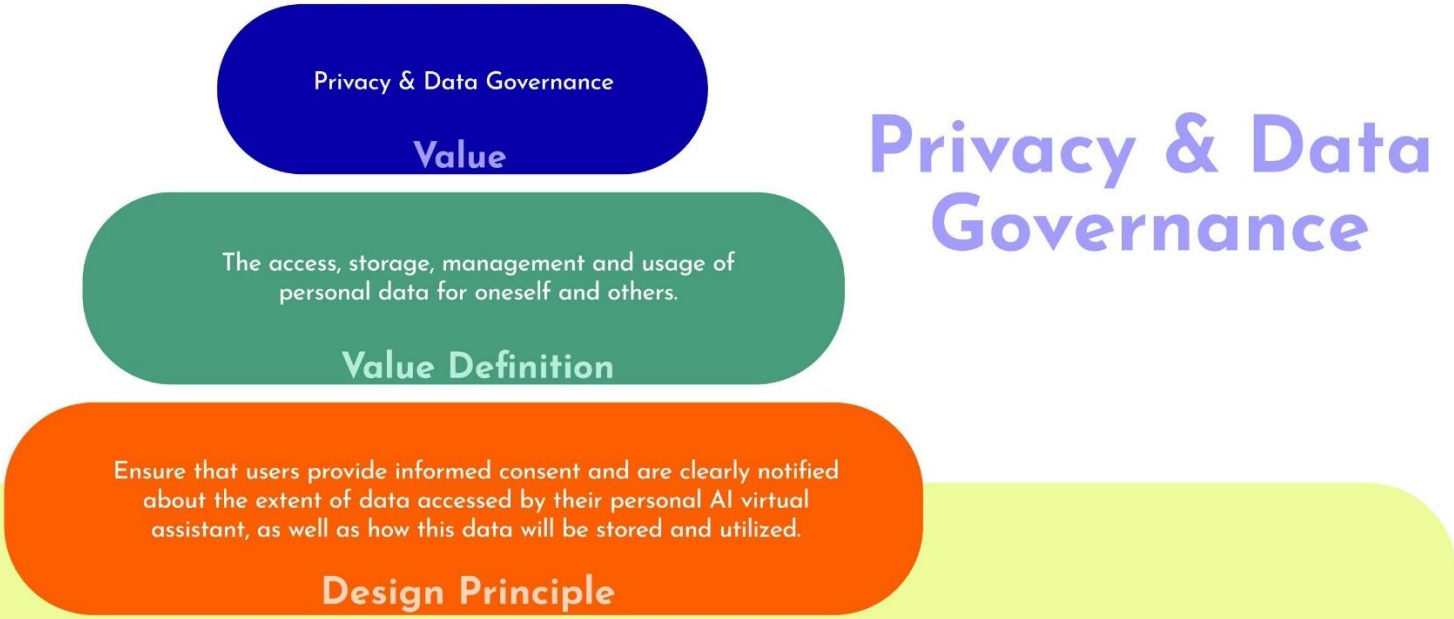
Design Principle

Human Agency, Autonomy & Oversight

1. Users should be the main drivers, actively making the final decisions.
2. Users should have the autonomy and control over decisions made by personal AI virtual assistants.
3. Users should monitor and oversee the decisions made by their personal AI virtual assistants.
4. Users should approve and finalize the decisions made by their personal AI virtual assistant.
5. Users should have control over the end decisions made by their personal AI virtual assistant.
6. Users should be able to regain control and autonomy from their personal AI virtual assistants.
7. Users should be able to guide their personal AI virtual assistants by clearly defining their choices and intentions.
8. Users should have authority and the ability to override a personal AI virtual assistant.

Design Norms

Implications for Design



Privacy & Data Governance

1. Rather than granting their personal AI virtual assistant full access to all their personal data, users should have the option to also guide it by personally sharing their preferences and choices.
2. Users should provide explicit consent regarding the specific amount of data they are comfortable with their personal AI virtual assistant accessing.
3. Users should know how their personal AI virtual assistant trains and operates on their personal data.
4. Users should be notified and have control over the access to other people's personal data that may be inadvertently shared with their personal AI virtual assistant.
5. Users should have control over the end decisions made by their personal AI virtual assistant.
6. In any situation where a personal AI virtual assistant is recording, storing and sharing data, users should be clearly notified.

Design Norms

Implications for Design

Transparency

Value

Transparency

The clarity and identification of the source and origin of communication.

Value Definition

Ensure that for tasks and actions with sentimental value, users have the option to clarify whether the outcome was entirely achieved by the personal AI virtual assistant or involved their direct input.

Design Principle

1. Users should have the freedom to decide whether or not to disclose the involvement of a personal AI virtual assistant in any action.
2. Users should have control over whether they disclose the involvement of their personal AI virtual assistant in interactions or engagements with others.
3. Users should know if a personal interaction, particularly one with emotional or sentimental significance was conducted with another use or with their personal AI virtual assistant.

Design Norms

Implications for Design

Overall Impact & Wellbeing

Value

Overall Impact & Wellbeing

The broad consequences, impacts, and effects on society as a whole.

Value Definition

Ensure that personal AI virtual assistants enhance users' sense of connection and strengthen interpersonal bonds, while avoiding the erosion of social skills and behaviors that could lead to loneliness or social awkwardness. The impact on users' well-being should be positive and supportive.

Design Principle

1. Personal AI virtual assistants should encourage and facilitate genuine, authentic connections between users.
2. Personal AI virtual assistants should encourage and support users in making personal efforts to maintain or establish meaningful connections.
3. Personal AI virtual assistants should facilitate and encourage users to engage in direct, personal interactions and connections with others.
4. Personal AI virtual assistants should support users in actively participating in personal and emotional interactions to maintain trust and prevent feelings of distance or emotional detachment.
5. A user's personal AI virtual assistant should not be used to communicate or express the user's emotions to others during personal, sensitive or emotional interactions.
6. Personal AI virtual assistants should cater to users' preferences and needs, providing tailored support based on individual requirements.
7. Personal AI virtual assistants should assist users in managing their tasks, allowing them to allocate more time to enjoyable activities and enhancing their overall well-being.
8. Personal AI virtual assistants should support users in initiating interactions with others without entirely automating them, thereby enhancing users' social behaviors and skills rather than diminishing them.

Design Norms

Implications for Design

Personal Effort

Value

Personal Effort

The personal effort invested in achieving a desired outcome.

Value Definition

Ensure that users' input and preferences are consistently incorporated, implemented, and prioritized in the actions performed by their personal AI virtual assistant.

Design Principle

1. Personal AI virtual assistants should not entirely handle user's connectivity without any involvement from their end.
2. Communications between users should not entirely happen through personal AI virtual assistants.
3. Personal AI virtual assistants should facilitate, encourage and involve users in providing their input when engaging in emotional interactions.
4. Users should be facilitated and encouraged to add a personal touch to the content generated by their personal AI virtual assistant.
5. Users should be involved in the process and facilitated in investing personal time and effort when taking help from their personal AI virtual assistant to express emotions.
6. Personal AI virtual assistants should facilitate connectivity by seeking final approval from the user rather than autonomously performing tasks.
7. Personal AI virtual assistants should encourage and support users in being involved and making personal effort for people they care about.
8. Personal AI virtual assistants should not become the only mode of communication between users.

Design Norms

Implications for Design

Human-Human Connectivity

Value

Human-Human Connectivity

The direct interaction between individuals.

Value Definition

Ensure that users' personal AI virtual assistants actively facilitate and promote opportunities for face-to-face or direct interactions between individuals.

Design Principle

1. Communication between users should not happen only through their personal AI virtual assistants.
2. Personal AI virtual assistants should enable and initiate human-human engagement *(eg. providing conversation starters)*.
3. Personal AI virtual assistants should facilitate more personal engagement and direct interaction between users *(eg. making suggestions, sending reminders)*.
4. Personal AI virtual assistants should not entirely take over communication by fully automating it.
5. Users should not be put in a situation where they are the only human engaged in discussion with other users' personal AI virtual assistants.
6. Personal AI virtual assistants should facilitate users to engage more effectively in human-human interactions.

Design Norms

Implications for Design

Genuity & Authenticity

Value

The honesty, sincerity and integrity in representing oneself and reflecting emotions and intentions.

Value Definition

Ensure that users' personal AI virtual assistants encourage and support the direct expression of emotions and sentiments by users themselves, rather than autonomously expressing these on their behalf.

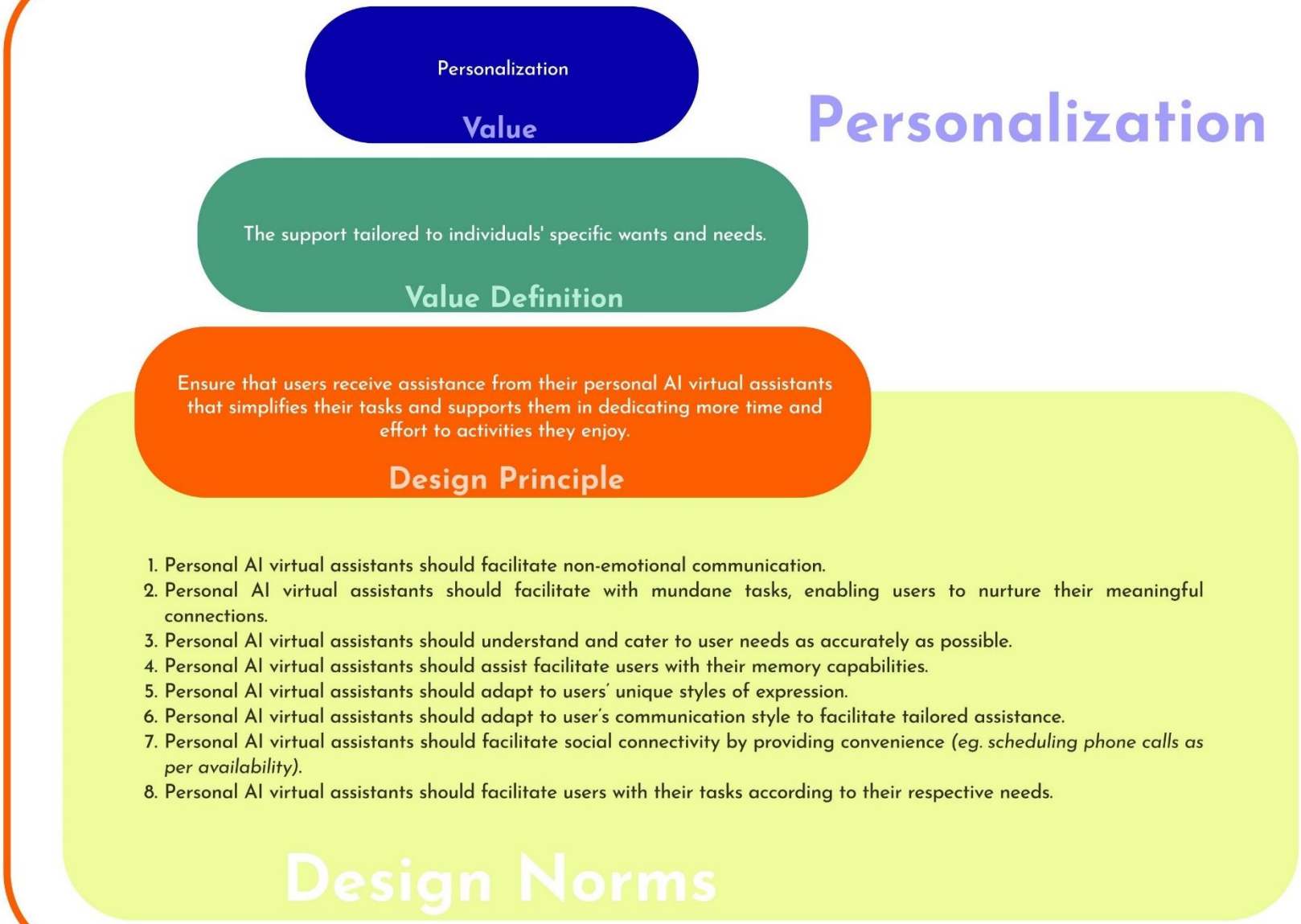
Design Principle

Genuity & Authenticity

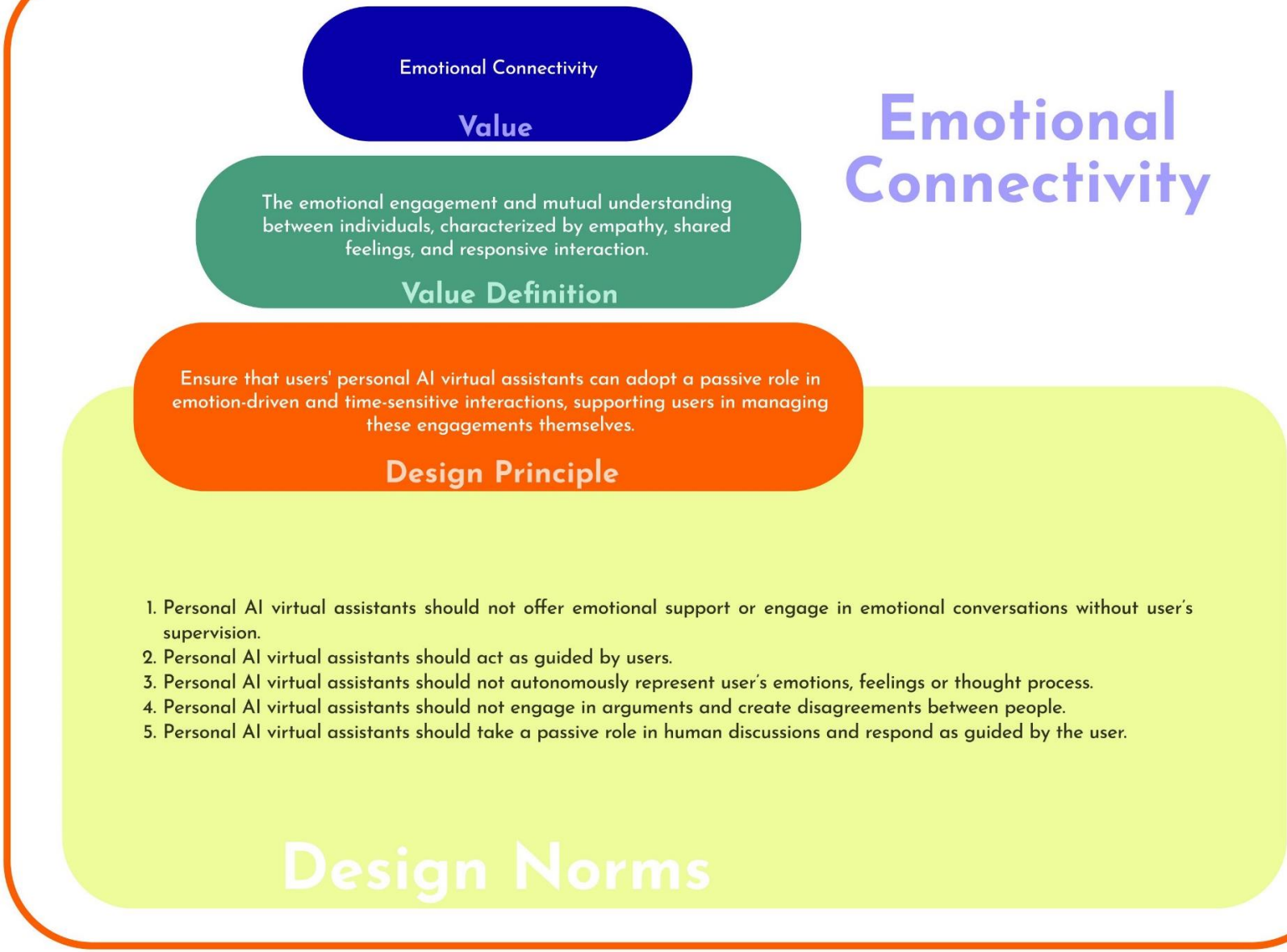
1. Personal AI virtual assistants should not create communication inaccuracies, especially when representing emotions or taking decisions.
2. Personal AI virtual assistants should not create fake or pseudo connectivity between users.
3. Personal AI virtual assistants should not create trust issues among users through intimate and emotional engagements.
4. Personal AI virtual assistants should not be used as the sole mode of communication for heartfelt bonds.
5. Personal AI virtual assistants should not engage in emotionally sensitive and intimate communication.
6. Personal AI virtual assistants should not entirely dehumanize the connectivity.
7. Personal AI virtual assistants should not create uncomfortable or awkward situations between users.

Design Norms

Implications for Design



Implications for Design



8. RESEARCH CONCLUSION

This chapter concludes the research by reflecting on the proceedings and activities conducted throughout the study. It revisits the research questions and goals, demonstrating how they were addressed through the activities detailed in the preceding chapters. The chapter then discusses the overall findings, implications, limitations, and suggests avenues for future research related to the project.

8.1 Research Proceedings, Research Questions & Goals

This research examined the growing use of AI technologies in the realm of social connectivity through the lens of responsible AI. The project aimed to explore the future development of AI technologies in social contexts and investigate how these technologies can be designed responsibly, with a focus on user values. The primary objective was to generate design implications that would guide designers in the responsible creation of AI technologies that foster social connectivity. The central research assignment guiding the motives of the research was:

Generate design implications by identifying and eliciting the values of young adults through design speculations, to guide designers in the responsible development of future AI technologies that foster social connectivity.

To achieve this, the research undertook a thorough review of theoretical data to (1) define social connectivity and its significance for this study; (2) investigate current and emerging trends in the application of AI within social contexts; and (3) examine responsibility and the responsible AI guidelines pertinent to this research. Building on the insights gained from these steps, this research *identified* and *elicited stakeholder values* pertinent to the application of AI in the social domain through design experiments. To effectively design, plan, and execute these experiments effectively, the research employed a combination of *Speculative Critical Design* and *Value Sensitive Design* approaches. Ultimately, by *identifying, eliciting,* and *defining relevant values*, the research *generated design implications* in the form of *9 pyramids of value* that can guide designers in responsibly developing AI technologies for social connectivity.

The initial phase of establishing a theoretical foundation proved to be quite extensive, given the complexities of the sub-topics and their interconnections. The first area of exploration was social connectivity, with the goal of defining its meaning within the context of this research and identifying a specific user group. Social connectivity encompasses a wide range of interactions, each with distinct needs. This research focused on the social connectivity of young adults, particularly their use of social media and digital platforms to maintain connections with close friends and family. This type of connectivity, termed "interpersonal social connectivity," was identified as the primary focus of the study.

To anticipate the trajectory of future AI technologies, this research embarked on an exploration of current AI developments. This section yielded valuable insights into how AI is being developed to enhance social connectivity on existing platforms, introducing a wide array of new features and functionalities designed to improve how people connect. By thoroughly examining these emerging AI functionalities within the social domain, the research highlighted the advancement of AI virtual assistants tailored for social

platforms, identifying key trends in the features they are being engineered to offer within current social media applications. Based on these findings and their relevance to the research context, the focus was refined from *future AI technologies fostering social connectivity* to specifically concentrating on *AI virtual assistants designed to enhance interpersonal connectivity*. The insights gained through this staged answered the first research question:

(a) What are the trends in AI technologies fostering social connectivity?

To address this question, the research concluded the AI trend analysis by identifying eight key functionalities and features being integrated into AI virtual assistants for enhancing social connectivity. The study first highlighted the significant trend of increasing development and integration of AI virtual assistants within social platforms. Under this overarching trend, the research detailed specific functionalities and features currently being designed and implemented in these AI systems. The identified trends were:

1. Multipurpose, offering various functionalities
2. Multimodal, processing various forms of data
3. Humanlike, having characters, personalities, opinions, long term memory
4. Self-learning, autonomously improving performance without human intervention
5. Guiding, influencing and supporting decision-making process and making decisions
6. Generating content
7. Training on personal data
8. Simulating social interaction between people

The research then focused on understanding the importance of responsibility within its specific context. By analyzing the roles of various stakeholders—including designers, companies, and policymakers—in the responsible design of AI, the study reviewed existing AI principles and guidelines to determine the most suitable framework. The research chose the Assessment List for Trustworthy Artificial Intelligence (ALTAI) as the guiding framework, which provides criteria for evaluating trustworthy AI, due to its more comprehensive and practically applicable nature, compared to other guidelines giving out a general set of principles, ALTAI provides with a list of specific question to assess the responsibility of AI. Out of the seven core values outlined by ALTAI, four were identified as directly relevant to the research context. These values were selected based on their alignment with the identified trends and their potential impact on interpersonal social connectivity, ultimately serving as the contextual values for this study. These identified values were:

1. Human Agency and Oversight
2. Privacy and Data Governance
3. Transparency
4. Societal and Environmental Wellbeing

The findings from the in-depth theoretical analysis culminated in the development of a framework, titled "*Framework for Designing Responsible AI virtual assistants for social connectivity*". This framework established the connections between the three independently explored topic areas, tailoring them specifically to the context of this research. It also presented the four previously identified contextual values in relation to the relevant questions from ALTAI, the trends observed in AI virtual assistants and their possible social effects on human-human connectivity. Additionally, it provided guidance for the design phase by highlighting the key elements for this research and their interrelation.

The progress of the research up to this point refined the research assignment, making it more specific and focused. The highlighted sections indicate the key areas of refinement.

Generate design implications by identifying and eliciting the values of young adults through design speculations, to guide designers in the responsible development of future AI virtual assistants that foster interpersonal social connectivity.

Moving forward, with a focus on developing future AI virtual assistants through a responsible AI lens, this research sought to understand, surface, and elicit the values of young adults regarding AI-fostered social connectivity. The aim was to discern what is important to young adults when connecting with those with whom they share emotional bonds, such as friends. To elicit these values, the project adopted *Speculative Critical Design (SCD)* and *Value Sensitive Design (VSD)* as key approaches. SCD provided tools and methods to spark imagination about the future, foster creative and critical thinking, and provoke discussion, while VSD offered valuable tools and methods for addressing human values and societal impacts in relation to technological developments. Elements from both approaches, relevant to the project's objectives, were combined to create a tailored approach. This combination guided the development of speculative design scenarios that presented a fictional AI virtual assistant set in the near future, grounded in the analysis of current AI trends, and designed to enhance communication and foster connectivity. The SCD and VSD approaches not only helped in crafting these scenarios but also in delivering and discussing them with participants through focus groups, generating valuable data for the research.

The data generated was then analyzed and synthesized to understand participants' values. Emerging value themes were clustered using principles of thematic analysis. Guided by VSD, nine values were elicited—four of which were previously identified, while the remaining five emerged as new, context-specific values during the discussions. The re-emergence of the previously identified values not only validated their relevance to this research but also allowed for their elaboration, leading to a deeper contextual understanding. The newly elicited values further clarified what matters to people and highlighted their moral and ethical concerns within the specific context presented. Collectively, this set of nine contextual values guided the project towards its ultimate goal: generating design implications. This phase of the research process answered the second research question:

(b) What values do young people uphold regarding interpersonal connectivity fostered by AI technologies?

The research discovered a set of nine values that young adults uphold regarding interpersonal connectivity fostered by AI technologies, in particular, personal AI virtual assistants. The values are listed below:

1. Human Agency and Oversight
2. Privacy and Data Governance
3. Transparency
4. Societal and Environmental Wellbeing
5. Personal Effort
6. Human-Human Connectivity
7. Genuity and Authenticity
8. Personalization
9. Emotional Connectivity

The set of contextual values played a crucial role in generating *"Implications for Design."* Each value was first given a context-specific definition to clarify its significance. Following Van De Poel's (2013) Hierarchy of Values, each value was then expanded into value principles and design norms. The value principles serve as broad rules or guidelines that provide direction, while the design norms are specific prescriptions or restrictions that define the goals to be achieved or the boundaries to be respected in pursuit of those goals. This approach translated the nine values into nine value pyramids, each serving as a distinct implication for design. This generation of design implications finally answered the third research question stated below:

(c) How can design implications be generated to guide a responsible design of future AI technologies fostering social connectivity?

Thus, through the generation of design implications derived from extensive research activities, this study successfully addressed the final research question. These design implications are intended to guide designers in the responsible creation of future AI virtual assistants that foster interpersonal social connectivity.

8.2 Discussion

This research, with its focus on the development of AI technologies, their influence on social connectivity, and the broader societal implications, aligns with some of the core principles of Value Sensitive Design (VSD) as articulated by Friedman & Hendry (2019). These principles emphasize that the relationship between technology and human values is inherently interactional, involving the co-evolution of technology and socio-structural aspects, and advocate for a commitment to progress rather than perfection.

During the discussions prompted by the speculative scenarios and the process of value elicitation, it became evident that participants initially reacted with fear and hesitation toward certain features of the fictional technology presented. However, upon further reflection, they began to identify conditions

under which they might be comfortable using these features. For example, the idea of an AI-generated birthday wish sent by a personal virtual assistant to a close friend was initially met with criticism. Participants expressed concerns about the lack of authenticity in such interactions. Yet, after considering the potential utility of the feature, some participants suggested that it could be acceptable if the virtual assistant made suggestions or recommendations and required the user's approval before sending the message, rather than doing so automatically. As one participant remarked:

"I would be okay with the automation as long as I am making the end decision."

This shift in perspective highlights the nuanced relationship between technology and user values and underscores the importance of incorporating reflective processes into the design of AI technologies to ensure they align with users' evolving needs and ethical considerations.

The use of Speculative Critical Design (SCD) as an approach proved invaluable at multiple stages of this research. Initially, it facilitated the development of effective scenarios and subsequently ensured their successful delivery and discussion during the focus group sessions. Once participants engaged with these scenarios, they were immediately provoked to discuss the topics at hand. The scenarios stimulated numerous thoughts and reflections, prompting participants to share their insights and concerns with the group. Overall, participants reported finding the experience enjoyable, noting that they felt deeply immersed in the scenarios. One participant highlighted this immersion by stating:

"When you announced that now we will conclude the session, I saw the time and I was so surprised to think where all the time went, I was so immersed in the session that time just flew by."

The scenarios' realism and originality contributed to their authenticity and made them instantly relatable. Guided by SCD principles, the scenarios were grounded in current AI trends, avoiding a futuristic detachment by portraying believable and imminent developments. The scenarios depicted plausible future situations, which heightened their impact. In fact, one participant mistook the fictional video presented as part of the scenario for an actual promotional video from a real company. Another participant expressed:

"The scenarios feel very realistic, which is also scary, you know? I think that it might actually just happen. And I think that's why... like I really got involved in the scenarios because I can see all of this actually happening."

In addition to the scenarios' realism, the deep engagement and thought provocation they sparked were particularly notable. Participants not only discussed and questioned the scenarios during the session but also extended their reflections beyond the session, contemplating current norms and practices related to existing AI technologies, social media platforms, and digital behaviors. One participant, reflecting on the impact of the session, remarked:

"I could not stop thinking about the scenarios and the discussion that followed; I kept thinking about it for a while after the session ended. I was reflecting on the current norms and what may happen in the future."

This response underscored the effectiveness of SCD in fostering reflection and critical thinking. The scenarios prompted participants to not only consider the speculative futures presented to them but also to critically examine the present-day practices surrounding AI technologies and social norms. Participants reflected on various issues, such as the

vast amount of data that social media platforms already collect and the overwhelming number of notifications seeking user approval for actions they may not fully understand. This reflective process further validated the use of SCD as a powerful tool in eliciting meaningful engagement and critical discourse.

The focus group findings underscored the Value Sensitive Design (VSD) principle that values do not exist in isolation and are often interrelated. For instance, tensions between personalization and privacy emerged as a prominent theme. Participants were concerned about AI virtual assistants accessing extensive personal data to offer tailored assistance, yet they also recognized that such personalization could significantly enhance their social connections.

Additionally, the use of structured worksheets titled "Initial Thoughts," "This Makes Me Feel," "I Like," "I Dislike," "I Fear/Worry," "I Dream," and "Overall Impact" was effective in guiding the focus group discussions. These worksheets helped facilitate a natural flow of conversation and encouraged participants to express their thoughts freely, without explicitly focusing on values. This approach ensured that discussions remained relevant to the research objectives while capturing the nuanced reactions and reflections of the participants.
Moreover, summarizing the theoretical research insights into a context-specific framework effectively bridged the gap between research and design phases. This framework guided the development of design scenarios by incorporating its insights, ensuring that the scenarios were grounded in theoretical findings. The scenarios were designed to elaborate on and deepen the understanding of these insights within the research context, thereby aligning practical design efforts with theoretical principles.

8.3 Implications

This research offers significant insights for the design of AI technologies in the social domain by highlighting key values and addressing ethical and moral concerns. It enhances understanding of the values essential for AI-fostered social connectivity and their implications for responsible AI technology design. The nine value pyramids developed in this research translate these contextual values into design norms, providing a framework for converting broad value principles and norms into specific design requirements. These pyramids facilitate the practical application of values in designing AI technologies. They are particularly valuable for developing responsible AI systems that support interpersonal social connectivity, where emotional bonds, authentic expression, and trust are crucial.

In addition to the four AI values identified from ALTAI, this research introduces five new values specifically relevant to AI-facilitated social interactions. These nine values, thoroughly detailed and contextualized, provide essential guidance for designers and policymakers involved in the development of AI technologies aimed at enhancing social connectivity.

This research highlights the critical importance of understanding context-specific values in the design and implementation of AI technologies. The findings demonstrate that general AI principles, as derived from ALTAI, gain greater relevance and applicability when contextualized through detailed research activities. This approach reveals the nuanced significance of these values in specific scenarios, underscoring the need for designers and policymakers to delve deeply into the values relevant to their particular context. By doing so,

they can apply these values more effectively, ensuring a more responsible and contextually appropriate design of AI systems.

Finally, this research highlights the significance of developing a framework to synthesize and represent the core insights gathered from literature reviews. Such a framework serves as a crucial tool for transitioning from theoretical research to the design phase, effectively linking the two stages. By organizing and connecting the core insights, the framework provides a clear direction for the research, ensuring that the subsequent design activities are informed by the foundational knowledge and insights obtained. This structured approach facilitates a coherent and focused progression from theory to practical application.

Finally, this research demonstrates the effectiveness of combining Value Sensitive Design (VSD) and Speculative Critical Design (SCD) approaches. The integration of these methodologies facilitated the successful elicitation of insights and the generation of meaningful results. SCD played a crucial role in provoking engaging and thought-provoking discussions through the design and implementation of scenarios, while VSD provided a structured framework for value elicitation and the formulation of design implications. This combined approach ensured a comprehensive exploration of user values and their application in responsible AI design.

8.4 Limitations

This research acknowledges that while it focused on four ALTAI values deemed most relevant to the context, the remaining three values—Technical Robustness and Safety; Diversity, Non-discrimination and Fairness; and Accountability—may also offer valuable insights into context-specific understanding of AI technologies.

Additionally, the research recognizes that AI technologies play a role in fostering various forms of social connections beyond just interpersonal, emotional bonds, including professional and broader community interactions. Although the study concentrated on interpersonal social connectivity involving close, emotional relationships, exploring AI's impact on professional and community connections could yield different values and perspectives.

Furthermore, this research notes that the scenarios presented to participants were specific to certain uses of fictional technology. The insights gathered were directly related to these scenarios, and alternative representations of the technology's context could potentially reveal different insights. Thus, the findings are constrained by the particular scenarios used in the focus groups and the specific roles of the AI technology depicted in them.

From the data gathered, this research primarily focused on eliciting values and generating implications for design. However, it also recognizes the value of additional insights obtained during the analysis, particularly those involving participants' envisioned scenarios for AI technologies in social connectivity. These dream scenarios, where participants imagined innovative uses for AI, offer valuable perspectives that could inform the development of relevant design concepts.

Finally, while the design implications of this research are framed as design norms, these norms serve as a foundation for further development. Based on these norms, implications can

be extended, and the norms can be translated into specific design requirements tailored to the context.

8.5 Future Research

Future research could benefit from exploring additional values from the outset to uncover their relevance and implications specific to the research context. While this study concentrated on four identified values and elicited new ones, incorporating other AI values could deepen the understanding of their significance within the same context.

Additionally, research could expand to examine AI technologies in different social domains, such as community building or professional relationships. Investigating these areas may reveal varied values and perspectives related to AI-fostered social connectivity.

Exploring alternative scenarios or contexts could also provide new insights. By designing different scenarios or utilizing varied contexts, researchers can gain a more comprehensive understanding of user values related to AI-fostered interpersonal social connectivity.

Further analysis of data gathered through focus group worksheets could lead to valuable design concepts. Investigating participants' dream technologies or their recommendations for improving AI technologies could offer innovative ideas and practical solutions for enhancing social connectivity.

Lastly, the design norms developed in this research could be expanded into specific design requirements, offering concrete guidance for the responsible design of AI technologies aimed at fostering social connectivity.

References

A

Auger, J. (2013). Speculative design: crafting the speculation. Digital Creativity, 24(1), 11–35. https://doi.org/10.1080/14626268.2013.767276

C

Cacioppo, J. T., Hawkley, L. C., Ernst, J. M., Burleson, M., Berntson, G. G., Nouriani, B., & Spiegel, D. (2006). Loneliness within a nomological net: An evolutionary perspective. Journal of Research in Personality, 40(6), 1054–1085. https://doi.org/10.1016/j.jrp.2005.11.007

Cacioppo, S., Grippo, A. J., London, S., Goossens, L., & Cacioppo, J. T. (2015). Loneliness. Perspectives on Psychological Science, 10(2), 238–249. https://doi.org/10.1177/1745691615570616

Casnati, F., Ianniello, A., & Romani, A. (2023). Provocation through Narratives: new speculative design tools for Human-Non-Human collaborations. In Springer series in design and innovation (pp. 747–755). https://doi.org/10.1007/978-3-031-49811-4_71

Czeskis, A., Dermendjieva, I., Yapit, H., Borning, A., Friedman, B., Gill, B., & Kohno, T. (2010a). Parenting from the pocket. Value Tensions and Technical Directions for Secure and Private Parent-teen Mobile Safety, 15:1-15:15. https://doi.org/10.1145/1837110.1837130

Czeskis, A., Dermendjieva, I., Yapit, H., Borning, A., Friedman, B., Gill, B., & Kohno, T. (2010b). Parenting from the pocket. Value Tensions and Technical Directions for Secure and Private Parent-teen Mobile Safety. https://doi.org/10.1145/1837110.1837130

D

DataCamp. (n.d.). Guide to OpenAI API: Best practices. DataCamp. Retrieved August 22, 2024, from https://www.datacamp.com/tutorial/guide-to-openai-api-on-tutorial-best-practices

Dias, A., Geard, N., Campbell, P. T., Warr, D., & McVernon, J. (2018). Quantity or quality? Assessing relationships between perceived social connectedness and recorded encounters. PLoS ONE, 13(11), e0208083. https://doi.org/10.1371/journal.pone.0208083

Dunbar, R. I. M. (2014). The social brain. Current Directions in Psychological Science, 23(2), 109–114. https://doi.org/10.1177/0963721413517118

Dunne, A., & Raby, F. (2014). Speculative everything: design, fiction, and social dreaming. Choice Reviews Online, 51(10), 51–5390. https://doi.org/10.5860/choice.51-5390

E

Eggink, W., Ozkaramanli, D., Zaga, C., & Liberati, N. (2020). Setting the stage for responsible design. Proceedings of DRS. https://doi.org/10.21606/drs.2020.116

Ehn, P. (2008). Participation in design things. Participatory Design Conference, 92–101. https://doi.org/10.5555/1795234.1795248

Eisenberger, N. I., & Cole, S. W. (2012). Social neuroscience and health: neurophysiological mechanisms linking social ties with physical health. Nature Neuroscience, 15(5), 669–674. https://doi.org/10.1038/nn.3086

Ellis, S., Snyder-Mackler, N., Ruiz-Lambides, A., Platt, M. L., & Brent, L. J. N. (2019). Deconstructing sociality: the types of social connections that predict longevity in a group-living primate. Proceedings of the Royal Society B Biological Sciences, 286(1917), 20191991. https://doi.org/10.1098/rspb.2019.1991

European Commission. (2019). Ethics guidelines for trustworthy AI. European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Commission. (2020). Assessment list for trustworthy artificial intelligence (ALTAI): Self-assessment. European Commission. https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

F

Friedman, B. (1996). Value-sensitive design. Interactions, 3(6), 16–23. https://doi.org/10.1145/242485.242493

Friedman, B., & Hendry, D. G. (2019). Value sensitive design : shaping technology with moral imagination. In MIT Press eBooks (pp. 1–11). https://ci.nii.ac.jp/ncid/BB28979337

Friedman, B., Hendry, D. G., & Borning, A. (2017). A survey of value sensitive design methods. Foundations and Trends® in Human–Computer Interaction, 11(2), 63–125. https://doi.org/10.1561/1100000015

G

Gaggioli, A. (2018). Virtual Personal Assistants: an emerging trend in artificial intelligence. Cyberpsychology Behavior and Social Networking, 21(12), 803–804. https://doi.org/10.1089/cyber.2018.29133.csi

Google. (n.d.). Responsible AI practices. Google. Retrieved August 22, 2024, from https://ai.google/responsibility/responsible-ai-practices/

H

Haslam, C., Cruwys, T., Haslam, S. A., & Jetten, J. (2015). Social connectedness and health. In Springer eBooks (pp. 1–10). https://doi.org/10.1007/978-981-287-080-3_46-2

Hawkley, L. C., Browne, M. W., & Cacioppo, J. T. (2005). How can I connect with thee? Psychological Science, 16(10), 798–804. https://doi.org/10.1111/j.1467-9280.2005.01617.x

Ho, V. (2024, August 21). Three big AI trends to watch in 2024. Microsoft. https://news.microsoft.com/three-big-ai-trends-to-watch-in-2024/

J

Johannessen, L. K., Keitsch, M. M., & Pettersen, I. N. (2019). Speculative and critical design — features, methods, and practices. Proceedings of the . . . International Conference on Engineering Design, 1(1), 1623–1632. https://doi.org/10.1017/dsi.2019.168

L

Lieberman, M. D. (2014). Social: why our brains are wired to connect. Choice Reviews Online, 51(12), 51–7036. https://doi.org/10.5860/choice.51-7036

Luria, M., & Candy, S. (2022). Letters from the Future: Exploring Ethical Dilemmas in the Design of Social Agents. CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3491102.3517536

M

Malpass, M. (2010). Perspectives on Critical Design: a Conversation with Ralph Ball and Maxine Naylor. A Conversation With Ralph Ball and Maxine Naylor. http://irep.ntu.ac.uk/id/eprint/1779/

Meta. (2023, December 5). Meta AI updates. Meta. https://about.fb.com/news/2023/12/meta-ai-updates/

Meta. (2023, June 6). Introducing Voicebox: AI for speech generation. Meta. https://about.fb.com/news/2023/06/introducing-voicebox-ai-for-speech-generation/

Meta. (2023, September 13). Introducing AI-powered assistants, characters, and creative tools. Meta. https://about.fb.com/news/2023/09/introducing-ai-powered-assistants-characters-and-creative-tools/

Meta. (2023, September 20). Social profiles for Meta's AI characters. Meta. https://about.fb.com/news/2023/09/social-profiles-for-metas-ai-characters/

Meta. (2023, September 6). Building generative AI features responsibly. Meta. https://about.fb.com/news/2023/09/building-generative-ai-features-responsibly/

Meta. (2024, April 10). Meta AI assistant built with LLaMA 3. Meta. https://about.fb.com/news/2024/04/meta-ai-assistant-built-with-llama-3/

Meta. (2024, August 22). Meta LLaMA 3: Unveiling the next generation of language models. Meta. https://ai.meta.com/blog/meta-llama-3/

Meta. (n.d.). Generative AI privacy guide. Meta. Retrieved August 22, 2024, from https://www.facebook.com/privacy/guide/genai

Meta. (n.d.). Meta AI. Meta. Retrieved August 22, 2024, from https://ai.meta.com/meta-ai/

Meta. (n.d.). Responsible AI. Meta. Retrieved August 22, 2024, from https://ai.meta.com/responsible-ai/

Meta. (n.d.). Responsible use guide. Meta. Retrieved August 22, 2024, from https://ai.meta.com/static-resource/responsible-use-guide/ [Original source: https://studycrumb.com/alphabetizer]

Microsoft. (n.d.). Responsible AI in Azure Machine Learning. Microsoft. Retrieved August 22, 2024, from https://learn.microsoft.com/en-us/azure/machine-learning/concept-responsible-ai?view=azureml-api-2

Microsoft. (2022, June). Microsoft's responsible AI standards: Our commitment to ethical AI [Blog post]. Microsoft. https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-Responsible-AI-Stan.pdf

Mitrović, I. (2015). An introduction to speculative design practice. Department for Visual Communications Design, Arts Academy, University of Split. http://speculative.hr/en/introduction-to-speculative-design-practice/

Mitrović, I. (2016). Introduction to speculative design practice: Speculative – Post-design practice or new utopia? Speculative. http://speculative.hr/en/introduction-to-speculative-design-practice/

Mitrović, I., Auger, J., Hanna, J., & Helgason, I. (2021). Beyond Speculative Design : Past – Present – Future. Past – Present – Future. https://research.tilburguniversity.edu/en/publications/beyond-speculative-design-past-present-future

O

O'Rourke, H. M., & Sidani, S. (2017). Definition, Determinants, and Outcomes of Social Connectedness for Older Adults: A scoping review. Journal of Gerontological Nursing, 43(7), 43–52. https://doi.org/10.3928/00989134-20170223-03

OpenAI. (2023, August 23). Planning for AGI and beyond. OpenAI. https://openai.com/index/planning-for-agi-and-beyond/

OpenAI. (2024, August 15). Sora: The future of AI interaction. OpenAI. https://openai.com/index/sora/

OpenAI. (n.d.). OpenAI API. OpenAI. Retrieved August 22, 2024, from https://openai.com/api/

OpenAI. (n.d.). OpenAI Charter. OpenAI. Retrieved August 22, 2024, from https://openai.com/charter/

OpenAI. (n.d.). Safety standards. OpenAI. Retrieved August 22, 2024, from https://openai.com/safety-standards/ [Original source: https://studycrumb.com/alphabetizer]

P

Pichère, P., & Cadiat, A.-C. (2015). Maslow's hierarchy of needs. Lemaitre.

R

Research and Markets. (2018). Intelligent virtual assistant market: Global industry trends, share, size, growth, opportunity and forecast 2018–2023. https://www.researchandmarkets.com/research/bthc3k/global

Risnes, M., Thorstensen, E., Mirtaheri, P., & Berg, A. (2024). Exploring value dilemmas of brain monitoring technology through speculative design scenarios. Journal of Responsible Technology, 17, 100074. https://doi.org/10.1016/j.jrt.2023.100074

S

Stavrova, O., & Luhmann, M. (2015). Social connectedness as a source and consequence of meaning in life. The Journal of Positive Psychology, 11(5), 470–479. https://doi.org/10.1080/17439760.2015.1117127

Sutcliffe, A., Dunbar, R., Binder, J., & Arrow, H. (2014). Relationships and the social brain. In Oxford University Press eBooks (pp. 129–150). https://doi.org/10.1093/acprof:osobl/9780199652594.003.0007

T

The Fountain Institute. (n.d.). The Double Diamond of Speculative Design. The Fountain Institute. Retrieved August 22, 2024, from https://www.thefountaininstitute.com/post/the-double-diamond-of-speculative-design

Tromp, N., Hekkert, P., & Verbeek, P. (2011). Design for Socially Responsible Behavior: a classification of influence based on intended user experience. Design Issues, 27(3), 3–19. https://doi.org/10.1162/desi_a_00087

U

Umbrello, S., & Van De Poel, I. (2021). Mapping value sensitive design onto AI for social good principles. AI And Ethics, 1(3), 283–296. https://doi.org/10.1007/s43681-021-00038-3

V

Van De Poel, I. (2013). Translating Values into Design Requirements. In Philosophy of engineering and technology (pp. 253–266). https://doi.org/10.1007/978-94-007-7762-0_20

W

Wong, R. Y., Valdez, J. C., Alexander, A., Chiang, A., Quesada, O., & Pierce, J. (2023). Broadening Privacy and Surveillance. Eliciting Interconnected Values With a Scenarios Workbook on Smart Home Cameras. https://doi.org/10.1145/3563657.3596012 [Original source: https://studycrumb.com/alphabetizer]

Appendix

Appendix A

DESIGN FOR our future

TU Delft

Personal Project Brief – IDE Master Graduation Project

Name student Noor-ul-Ain Javed **Student number** [redacted]

PROJECT TITLE, INTRODUCTION, PROBLEM DEFINITION and ASSIGNMENT
Complete all fields, keep information clear, specific and concise

Project title Responsible Design of Future AI Technologies Fostering Social Connectivity

Please state the title of your graduation project (above). Keep the title compact and simple. Do not use abbreviations. The remainder of this document allows you to define and clarify your graduation project.

Introduction

Describe the context of your project here; What is the domain in which your project takes place? Who are the main stakeholders and what interests are at stake? Describe the opportunities (and limitations) in this domain to better serve the stakeholder interests. (max 250 words)

In the upcoming years, there are expectations for the continued development and seamless integration of AI technologies into various technological platforms, each with their designated functions. In this context, this graduation project: Responsible Design of Future AI Technologies Fostering Social Connectivity is dedicated to envisioning the potential evolution of future AI technologies (eg. personal AI virtual assistants), foreseeing their significant role in nurturing social connections.

Technology has become an integral aspect of contemporary social dynamics especially for young adults. In 2022, 84% of young adults aged 16-29 represented primary users of social media networks [1][2], offering both, opportunities for connection and complexities to navigate in the digital age. Given their adaptability to new technologies, propensity for volatility and online presence, young adults aged 18-26 years will be involved in this research.

By reflecting on how young people connect with distant friends using current AI technologies (eg. Instagram) and understanding the AI elements integrated in these technologies that help foster connectivity, this project aims to prompt critical questions and through speculation, explore potential design interventions. These interventions are intended to guide designers in responsibly integrating future AI technologies to foster social connectivity.

References:
[1] https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Being_young_in_Europe_today_-_digital_world&oldid=564756#:~:text=In%202022%2C%2084%20%25%20of%20young,media%20networks%20in%20the%20EU.
[2] https://www.linkedin.com/pulse/age-factor-how-social-media-affects-us

→ *space available for images / figures on next page*

DESIGN FOR our future

TU Delft

Personal Project Brief – IDE Master Graduation Project

Problem Definition

What problem do you want to solve in the context described in the introduction, and within the available time frame of 100 working days? (= Master Graduation Project of 30 EC). What opportunities do you see to create added value for the described stakeholders? Substantiate your choice.
(max 200 words)

Through speculation and criticism, the potential role of personal AI virtual assistants on enhancing social connectivity will be explored in this project. This includes envisioning scenarios in which this future AI technology takes inspiration from current AI features that foster connectivity. While existing digital social paltforms like Instagram undoubtedly contribute to enhancing social connectivity, it's also essential to acknowledge their negative impacts within this domain. Some examples include the lack of authenticity in individuals' personal profiles, instances of cyberbullying, perpetuation of constant comparison with others, pressure to continually expand one's social network, fostering addictive digital behaviors, distractions from meaningful offline interactions and connections etc. In the years ahead, we anticipate ongoing advancements in AI technologies designed to enhance social connectivity. Hence, this project is dedicated to delving into this progression, focusing specifically on the role of personal AI virtual assistants and how designers can responsibly integrate them in the future.
Q. How can future AI technologies fostering social connectivity be responsibly designed for young people?
(a) How are the AI features currently integrated in digital platforms negatively effecting personal social relationships?
(b) How can AI features in future technologies negatively effect personal social relationships?
(c) How can design intervention help foster social connectivity responsibly with future AI technologies (eg. personal virtual assistants)?

Assignment

This is the most important part of the project brief because it will give a clear direction of what you are heading for. Formulate an assignment to yourself regarding what you expect to deliver as result at the end of your project. (1 sentence) As you graduate as an industrial design engineer, your assignment will start with a verb (Design/Investigate/Validate/Create), and you may use the green text format:

Investigate a speculative design solution to analyze how personal AI virtual assistants can foster social connectivity responsibly in the future, for young adults aged 16-26 years, that are already using digital AI platforms for social connectivity, by extrapolating insights from the results of the speculation.

Then explain your project approach to carrying out your graduation project and what research and design methods you plan to use to generate your design solution (max 150 words)

This project merges present-day challenges with speculative exploration of fictional scenarios for a possible future. In order to achieve that and investigate design interventions, the initial methods include sensitizing the participants to the topic through a sensitizing booklet. This will be followed by one-on-one interviews that help understand people's concerns and problems faced with current AI technologies like Instagram. Afterwards, to speculate how personal AI virtual assistants can can foster social connectivity responsibly in the future, EXF methods for speculative and fictional design will be explored. So for example, to generate and investigate design interventions, participants will be invited for a co-creation session to brainstorm ideas together and participate in role-playing activity to envision the possible results.

introduction (continued): space for images

image / figure 1 Social connectivity currently

image / figure 2 Probable future of social connectivity

Project planning and key moments

*To make visible how you plan to spend your time, you must make a planning for the full project. You are advised to use a Gantt chart format to show the different phases of your project, deliverables you have in mind, meetings and in-between deadlines. Keep in mind that all activities should fit within the given run time of 100 working days. Your planning should include a **kick-off meeting**, **mid-term evaluation meeting**, **green light meeting** and **graduation ceremony**. Please indicate periods of part-time activities and/or periods of not spending time on your graduation project, if any (for instance because of holidays or parallel course activities).*

Make sure to attach the full plan to this project brief.
The four key moment dates must be filled in below

Kick off meeting	12 Mar 2024
Mid-term evaluation	24 May 2024
Green light meeting	24 Jul 2024
Graduation ceremony	27 Aug 2024

In exceptional cases (part of) the Graduation Project may need to be scheduled part-time. Indicate here if such applies to your project

Part of project scheduled part-time	
For how many project weeks	
Number of project days per week	

Comments:

Motivation and personal ambitions

Explain why you wish to start this project, what competencies you want to prove or develop (e.g. competencies acquired in your MSc programme, electives, extra-curricular activities or other).

Optionally, describe whether you have some personal learning ambitions which you explicitly want to address in this project, on top of the learning objectives of the Graduation Project itself. You might think of e.g. acquiring in depth knowledge on a specific subject, broadening your competencies or experimenting with a specific tool or methodology. Personal learning ambitions are limited to a maximum number of five.
(200 words max)

As a designer I am dedicated to cultivating well-being and positivity through my design endeavors, a commitment I aim to uphold in this project. My goal is to delve into the intricate dynamics of human relationships and social connectivity assisted by artificially intelligent technologies. This project offers an exciting opportunity for me to delve into speculative design and work with fictional scenarios—a domain that greatly intrigues me. I am particularly enthusiastic about exploring methods for speculative design and immersing myself in this creative process. That also excites me about the end results for this project and how it will be concluded. My inclination towards deep thinking and keen observation, coupled with a perfectionist approach to my work, characterizes me as a meticulous individual. While this meticulousness serves me well, it also leads to a thorough examination of problems from various angles before making decisions, which can be time-consuming. As I embark on this project, I aim to challenge myself to make bold design decisions and explore their outcomes, pushing beyond my comfort zone and embracing the opportunity for growth. My profound fascination with future technologies and their convergence with human interactions, alongside their integration into everyday life, motivated my decision to explore this topic. Furthermore, I am driven by a passion for learning and adopting new design methodologies. Previous semesters have afforded me the opportunity to delve into courses such as Advanced Machine Learning for Designers, Human-Agent Collaborations, Creative Facilitation, and Context Mapping Techniques, all of which have enriched my understanding of these subjects. As I undertake this project, I intend to apply the valuable insights gained from these courses.

Appendix B

Consent Form

Responsible Design of Future AI Technologies Fostering Social Connectivity

Purpose of the Activity: Insights drawn from this activity will help develop design guidelines for future designers, promoting responsible AI technologies that enhance social connectivity, specifically through personal virtual assistants. This activity targets young adults aged 18-28, and the data collected from their reflections and discussions will be used to extrapolate valuable insights into how personal virtual assistants can foster social connectivity and impact human relationships.

Plan of Activities: You are invited to participate in a research focus group. Each focus group will consist of 4-6 people and last for 1.5 hours. During the focus group, you will first be introduced to the research topic. Then you will be presented a video followed by some scenarios, based on which you will reflect on the worksheets handed out to and engage in a focus group discussion.

Data Management: We will be using a recording device to record your responses and clicking some pictures of the session. However, your participation in this research is treated with the utmost confidentiality. We are committed to protecting your identity and will keep your responses anonymous. Any personally identifiable data will be anonymized and deleted immediately after being transcribed. Until permanent deletion, the data will be stored in TU Delft's project drive which is only accessible to the people conducting this research.

Your Rights: Participation in this study is entirely voluntary, and you have the freedom to withdraw at any point or choose not to answer specific questions without giving a reason. Please also note that there is no compensation provided in participating.

Please read the following points and answer yes/no:	Yes	No
I have read and understood the study information or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction	☐	☐
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason	☐	☐

Please read the following points and answer yes/no:	Yes	No
I have had enough time to decide about my participation	☐	☐
I understand that taking part in the study involves interacting with design scenarios, followed by a discursive interview	☐	☐
I understand that pictures and voices will be recorded during the activities and observation notes will be taken	☐	☐
I understand that information I provide will be used in a report for the research purposes of a MSc thesis	☐	☐
I understand that personal information collected about me that can identify me, will not be shared beyond the study team. It will be anonymized and deleted right after transcription.	☐	☐
I understand that all the data will only be stored temporarily in TU Delft's project drive and would only be accessible to the research team	☐	☐

For any further inquiries, please do not hesitate to reach out to us anytime using the following details:

Name: Noor-ul-Ain Javed
Email: njaved@tudelft.nl

Thank you for considering this opportunity to contribute to this study!

Name of Participant	Signature	Date

I, as researcher, have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

Name of Researcher	Signature	Date
Noor-ul-Ain Javed		

Appendix C

Focus Group Schedule

Total time: 1.5 hours

25 mins	14:00 - 14:05	Participant walk-in
	14:05 - 14:15	Welcome, project introduction, overview of activities, consent form
	14:15 - 14:25	DTA video presentation + worksheet 1 (brain dump)
25 mins	14:25 - 14:30	What if... "The Birthday Wish" scenario introduction
	14:30 - 14:40	Individual reflection (worksheets 2-5)
	14:40 - 14:50	Group discussion + worksheet 6
25 mins	14:50 - 14:55	What if... "The Hiking Trip" scenario introduction
	14:55 - 15:05	Individual reflection (worksheets 2-5)
	15:05 - 15:15	Group discussion + worksheet 6
15 mins	15:15 - 15:20	Individual overall reflection (worksheet 7)
	15:20 - 15:30	Group discussion + wrap up

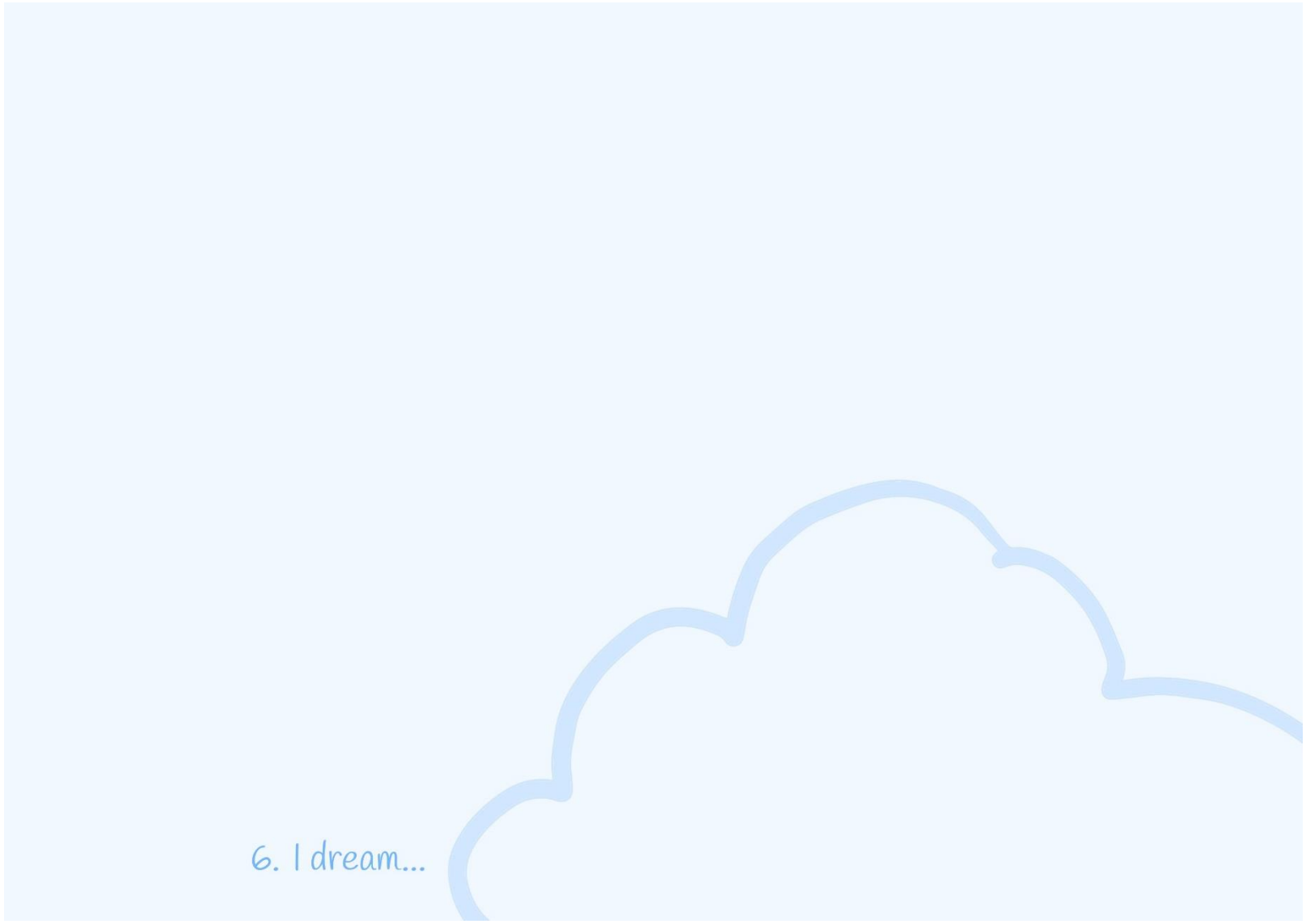
Appendix D

2. This makes me feel...


3. I like...
Initial Thoughts...


4. I dislike...


5. I fear/worry...


6. I dream...

7. Overall impact...

How do you think using the DTA to foster social connectivity can affect the following?

Relationships with others

Communication and social interactions

Life choices

Social skills and behaviors

Physical and mental wellbeing

Overall society