Voice and Beyond:

Shaping the Future of Personalized Conversational Agents

Jiayun Qiu

A thesis exhibition presented to OCAD University in partial fulfillment of the requirements for the degree of Master of Design in Digital Futures

OCADU Waterfront Campus, 130 Queens Quay E Tower, Mar 27-April 4,

Toronto, Ontario, Canada, 2025

Abstract

As advancements in generative AI and multimodal communication emerge, voice agents are poised to transition from functional tools to emotionally engaging companions. This thesis explores the future of interactions between Generation Z and AI voice agents, envisioning how these technologies could transform daily life by 2035. Through foresight and speculative design methodologies including STEEPV analysis, participatory futures, and design fiction prototypes, this research examines signals, trends, images, and drivers shaping the future of AI-human interaction.

The study focuses on the societal, ethical, and emotional implications of hyper-personalized and intelligent voice agents, exploring questions around inclusivity, safety, and design principles. By integrating trend research and user insights, creating, and prototyping three future scenarios set in 2035, this work offers actionable frameworks for creating AI agents that better align with values and expectations of users and communities. Ultimately, the thesis aims to inspire critical reflection on AI's role in society while guiding the design of voice agents that enrich the human experience responsibly and inclusively.

Acknowledgements

I would like to express my deepest gratitude to everyone who has supported me throughout the journey of completing this thesis. This work would not have been possible without the guidance, encouragement, and contributions of many individuals.

My sincere appreciation goes to my primary advisor, **Professor Greg Van Alstyne**, whose invaluable mentorship, patience, and insights have shaped this thesis. His teachings on methodologies for exploring the future and his encouragement to embrace uncertainty have profoundly influenced my approach.

I am also grateful to my secondary advisor, **Dr. Alexis Morris**, for his technical expertise and continuous encouragement to stay engaged with the latest advancements in AI.

The academic community at OCAD University has provided invaluable support. My classmates and OCAD alumni have offered fresh perspectives, constructive feedback, and encouragement that have greatly enriched my research and ideas.

A special thank you to the **SAS office** and **Hannah Dickson** for their support from February to April when I was unable to walk. Their help in ensuring my participation in the DFX exhibition and assisting with the setup was truly invaluable.

A heartfelt thank you to my friends **Brenda, Pan, Peter, Su Lyn, and Yu** for their support and encouragement. Their thoughtful feedback and willingness to discuss ideas have been invaluable to this thesis.

My deepest gratitude goes to **my mom and dad (aka 老妈&老爸)** for their unwavering support throughout my design journey across Melbourne, San Francisco, Shenzhen and Toronto. Their encouragement and support have given me the strength and determination to follow my passion, and I am forever grateful for their love and belief in me.

Artificial Intelligence Statement

I acknowledge using AI tools for surface-level enhancements of my text, including employing ChatGPT for grammar correction and improving readability. I also used various AI tools, such as DALL·E 3 for image generation, Pika and Sora for video creation, ElevenLabs for sound effects and voice synthesis, and Sumo for music composition. Additionally, during trend research, I used Perplexity to help expand my search for credible data sources and relevant signals. I have verified and edited all AI-assisted outputs and take full responsibility for the content of this publication.

Table of Contents

List of Appendices

List of Figures

1. Introduction

1.1 From Voice Assistants to Generative AI

When I first watched the movie *Her* a decade ago, I was captivated by the interactions and emotional connections between Theodore and the virtual assistant, Samantha, depicted in the film. At the time, these interactions felt futuristic, exciting, and almost far-fetched. On one hand, I was fascinated by how Samantha could seamlessly gather, process, and respond to information. On the other, I found it hard to imagine who could possibly fall in love with a virtual agent—one that had no physical form, just a voice. However, fast forward a decade to 2024, and we find that advancements in technology have made such interactions not only believable but likely. What once seemed like science fiction now feels entirely possible in our everyday lives.

Figure 1: Theodore Conversing with Samantha in the Movie Her (2013, Dir. Spike Jonze)

Over the past decade, the explosion of technologies such as always-on internet, voice recognition, artificial intelligence (AI), virtual reality (XR), autonomous driving, biometric monitoring, and the Internet of Things (IoT), has fundamentally transformed how we interact with each other. Members of Gen Z, as digital natives, have grown up immersed in computers

and the internet, witnessing, experiencing, and embracing every wave of technological change (Prensky, 2001).

Products like Apple's Siri, Google Assistant, and Amazon's Alexa successfully integrated voice AI into mainstream consumer applications. Voice technology has become an essential part of everyday life, redefining how we interact with technology (Vlahos, 2019). Today, voice commands are standard for daily tasks. For example, while cooking, users might say, "Hey Alexa, set a timer for 20 minutes." While driving, they may request, "Hey Siri, play some party music," or control IoT devices by saying, "Alexa, turn off the lights."

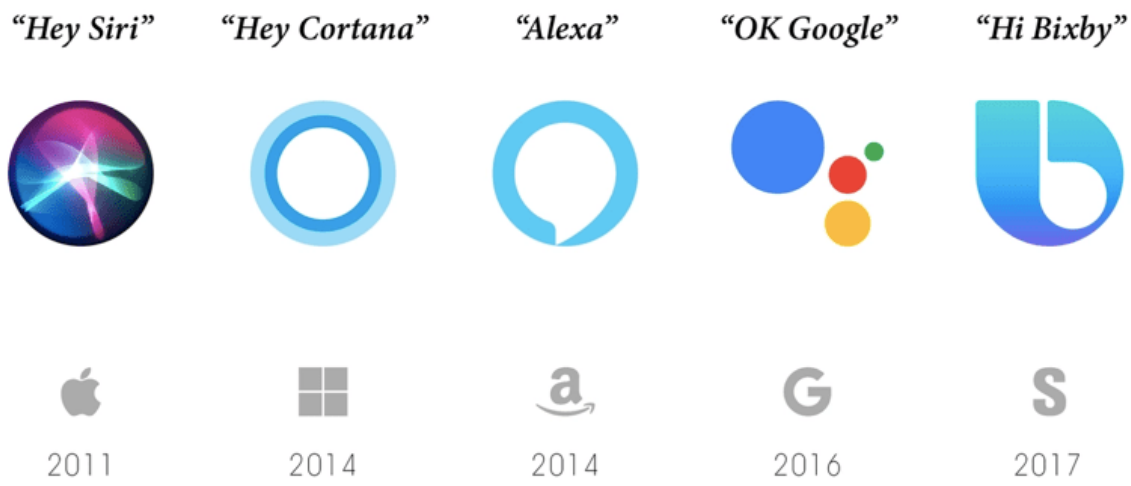


Figure 2: The Evolution of Virtual Personal Assistants Over the Past Decade (2021, Sarang)

Despite being applied across various industries, such as customer service, where they help respond to user inquiries more efficiently, voice assistants remained limited by their dependence on predefined commands and scripted responses (McTear, 2020). As a result, people didn't fully rely on these tools beyond small, routine tasks. Voice assistants were seen as conveniences rather than essential companions.

The COVID-19 pandemic significantly changed communication patterns. Social distancing and remote work became the norm, and people increasingly relied on digital tools to seek connection and emotional support. During this time, voice assistants evolved from simple task handlers into emotional companions, helping users cope with isolation and loneliness (Swoboda, 2020). "Alexa isn't just an assistant in most people's minds, but they like to chat with it and have an

empathetic relationship with it. Bored people want to be entertained and lonely ones seek emotional connection. It has received hundreds of thousands of marriage proposals and people expect Alexa to talk to them just like a friend. (Vlahos 2019)."

In 2022, the advent of generative AI models such as OpenAI's GPT and Google Bard transformed human-machine interactions. These agents could now engage in dynamic, context-aware conversations tailored to individual needs. Generative AI also elevated voice assistants to a new level (Hughes, 2023), enabling them to understand tone, context, and emotions, thus moving beyond mere transactional conversations. These interactions are now more meaningful, bringing us closer to the reality depicted in *Her*.

As a product designer, I have always been fascinated by the ways humans interact with objects and technologies. During my time as an industrial designer, my role often involved defining the relationships between people and products, shaping how they are used and experienced. This process required not only creativity but also an understanding of history, which understanding how past design have influenced present-day perceptions and practices. For example, the design of buttons and knobs in the 1990s shaped not only how people interacted with physical products but also how they conceptualized control and interaction in the digital realm (Norman, 2013). These seemingly simple design elements created a language of interaction that continues to inform both tangible and digital product designs today.

As we navigate an era of rapid technological growth shaped by advancements in AI, voice recognition, and multimodal interactions, I can't help but reflect on how these dynamics are evolving. Rewatching the film *Her*, I wondered: if a similar film were made today, how would it portray the relationship between humans and agents? And looking ahead a decade, how might those interactions change? Could they seamlessly enrich our lives like a dream? Or could they bring unexpected dependencies and ethical challenges to a potential nightmare?

1.2 Research Summary

This thesis explores the potential interactions and relationships between Generation Z and AI voice agents a decade from now. By examining the technological and societal advancements of today, it imagines how these technologies might integrate into daily life in the future, shaping personalized and meaningful human-agent interactions.

The main research question driving this project is:

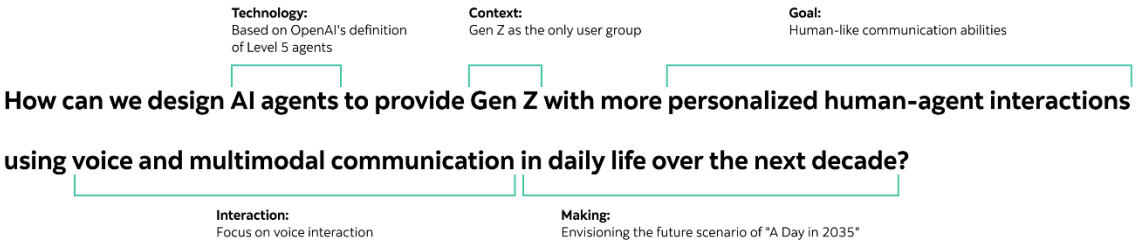


Figure 3: Breaking Down the Research Question to Define Scope and Limitations of the Thesis

This overarching question is supported by the following sub-questions:

1. ***How could Generation Z interact with AI agents in the future?***
2. ***What functionalities and experiences do Generation Z users expect from AI agents?***
3. ***How can we ensure AI agents are designed to be ethical, inclusive, and safe for long-term use?***

This thesis is organized into eight chapters, each contributing to a comprehensive exploration of the future of AI voice agents and their potential interactions with Generation Z. The introductory chapter sets the stage by discussing the evolution of voice agents from simple assistants to emotionally intelligent companions, highlighting key advancements in generative AI and multimodal communication. It also establishes the research problem, objectives, and central questions driving this study.

The following chapters delve deeper into the background and methodologies. A literature review examines the historical context, current capabilities, and ethical implications of voice agents. The research methodology chapter introduces the Double Diamond framework, which structures the study into phases of research, futuring, making, and reflecting. This is followed by an analysis of

trends, signals, and implications, which form the foundation for speculative scenarios. The thesis concludes with future scenarios, prototype evaluations, and reflections on the societal and design implications of AI voice agents.

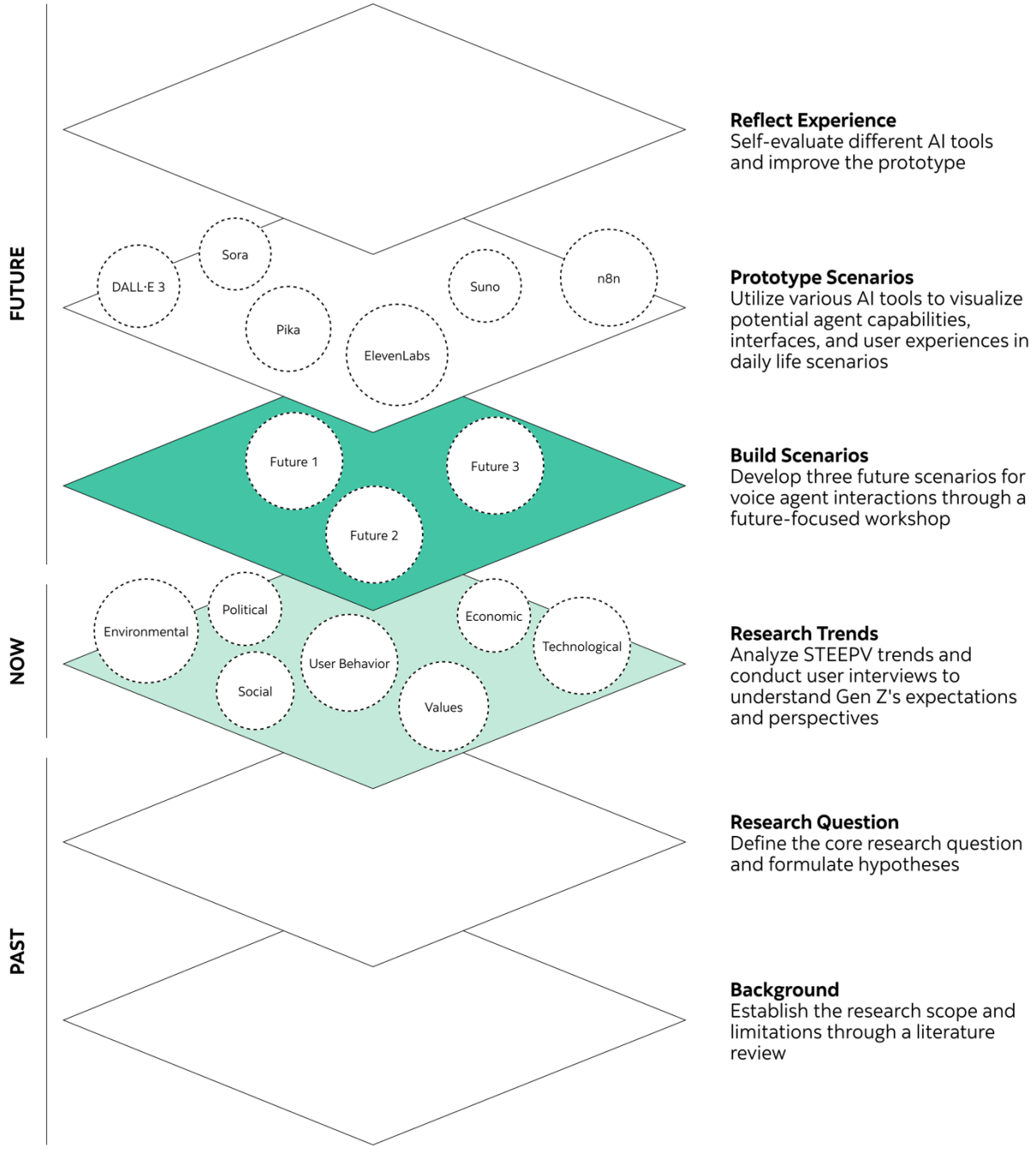


Figure 4: Structure and Process of This Thesis

1.3 Scope and Limitations

AI agents encompass a broad range of capabilities, yet this thesis focuses primarily on voice-based interactions rather than the entirety of multimodal communication. While some references to other modalities, such as images and text, may be included, they are not the central focus of this study. Additionally, voice agents come in various forms and platforms; however, this research does not delve into specific hardware, physical embodiments, or detailed designs such as voice characteristics, appearance, complete interface elements, or precise use cases. Instead, the emphasis is on exploring potential future interactions and the evolving ways humans may communicate with AI voice agents in 2035. The study is speculative and focuses on interaction possibilities rather than building a fully realized AI product.

According to the Technology Life Cycle Model, AI is currently in the Growth Stage (McKinsey, 2024). Each day, new AI products, models, policies, and industry developments are introduced, shaping the landscape at an unprecedented pace. **This thesis incorporates AI developments up until February 2025**; any advancements, models, or tools released beyond this point will not be included in the study or used in prototypes. This ensures that the research remains grounded in current AI trends while speculating on their long-term implications.

The primary user group for this thesis is Gen Z, as they have experienced the maturity stage of the internet and social media and are on the cusp of witnessing AI's transition from its growth stage to maturity. As digital natives, Gen Z has a far greater familiarity with technology than previous generations. By 2035, this generation might be at key life stages, ranging from graduating, entering the workforce, getting married, to starting and raising families, making them an essential group to study in the context of evolving AI interactions.

One of the key challenges of this research is presenting a tangible vision of the future. The goal is not to create highly polished prototypes but rather to use AI tools to generate speculative future scenarios that offer a sense of what the world might feel like in 2035. Another challenge is that the AI industry evolves rapidly, making it difficult to predict exact functionalities or capabilities. What is considered a breakthrough today may become widely accessible within months. For example, OpenAI's Operator, which integrates multiple advanced capabilities, was

rapidly built upon open-source models and APIs by independent developers in early 2024. Within just a few months, OpenAI transformed it into a publicly available product. Given this rapid pace of innovation, this thesis does not aim to define a fixed set of AI agent tasks but rather explores how AI voice agents might fit into everyday human interactions and how society may adapt to their presence.

2. Background and Literature review

2.1 History of Human-agent Interaction

Human fascination with machines capable of speech spans centuries, with early examples dating back to ancient Greece (Schrei, 2024). Hero of Alexandria designed mechanical statues capable of simple movements and speech, captivating audiences by mimicking human actions. Similarly, in the 10th century, Gerbert of Aurillac was rumored to have created a talking mechanical head, further exemplifying this fascination. These early imaginations and inventions reflect humanity's enduring desire to talk to machines although they were primitive in design (Shum, 2018). These efforts laid the conceptual groundwork for modern conversational technologies by demonstrating that machines could mimic human behavior.

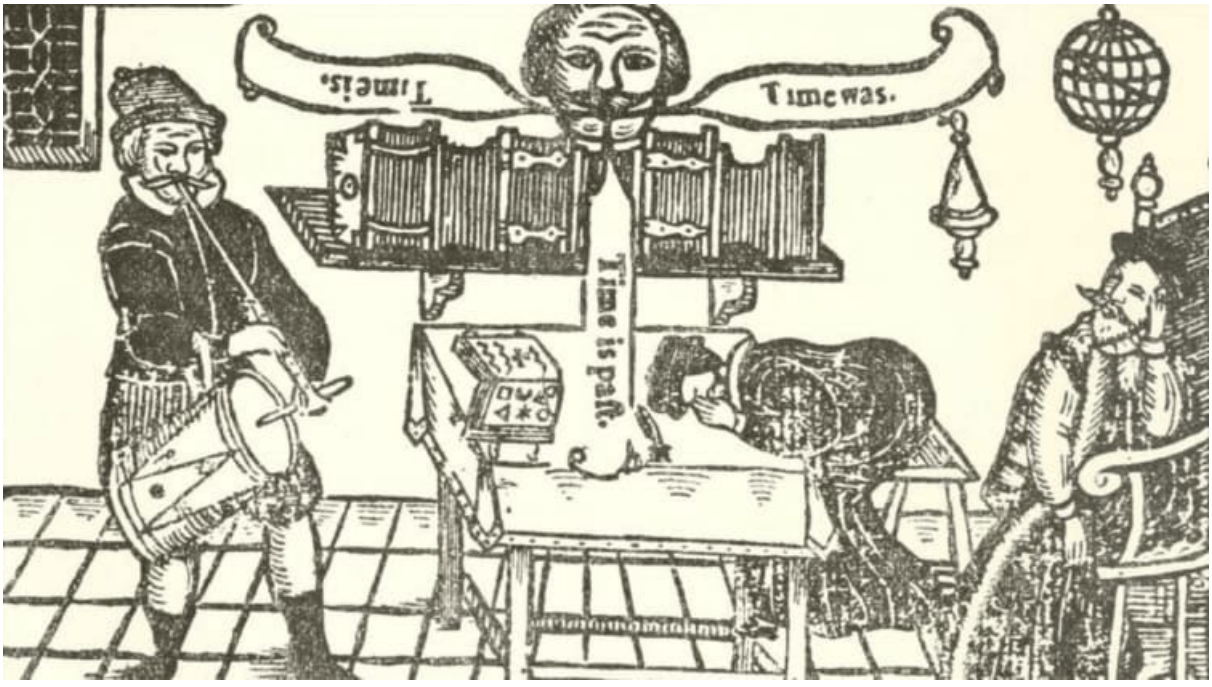


Figure 5: Brazen Head: The Legendary Automaton Capable of Answering Any Question (1991, The Senior)

With the development of computer technology, human-agent interaction entered the era of rule-based systems. In 1966, Joseph Weizenbaum developed **ELIZA** at MIT, which became a milestone in **natural language processing** and one of the first chatbots capable of attempting the **Turing Test**. ELIZA functioned by matching user input to scripted responses, using a script that

simulated a psychotherapist's responses (Campos et al., 2018). Building on ELIZA, PARRY, developed in 1972, attempted to simulate a paranoid patient, showcasing early efforts in emotional simulation (Yatoo & Habib, 2023). These systems, though limited by predefined rules, showcased the potential for computers to engage in conversational interactions and sparked interest in adaptive and intelligent agents.

Figure 6: A Conversation with Eliza (Wikipedia)

The 1990s and early 2000s saw the emergence of intelligent agents, expanding the complexity and functionality of interactions. In 1997, Microsoft introduced **Clippy** as an office assistant, providing real-time assistance to users by suggesting actions, such as offering help when it detected the user was writing a letter. Despite its intrusive nature and eventual discontinuation, Clippy set a precedent for future AI-powered in-product assistants (Walz, 2017). Around the same time, **IBM Watson**, a computer system capable of answering questions posed in natural language, gained prominence. Although limited in understanding nuanced context, Watson's ability to read, analyze, and learn natural language represented a significant advancement in decision-making systems (Shum, 2018). These developments marked a shift from rule-based

systems to more adaptive intelligent agents, establishing the groundwork for modern conversational AI.

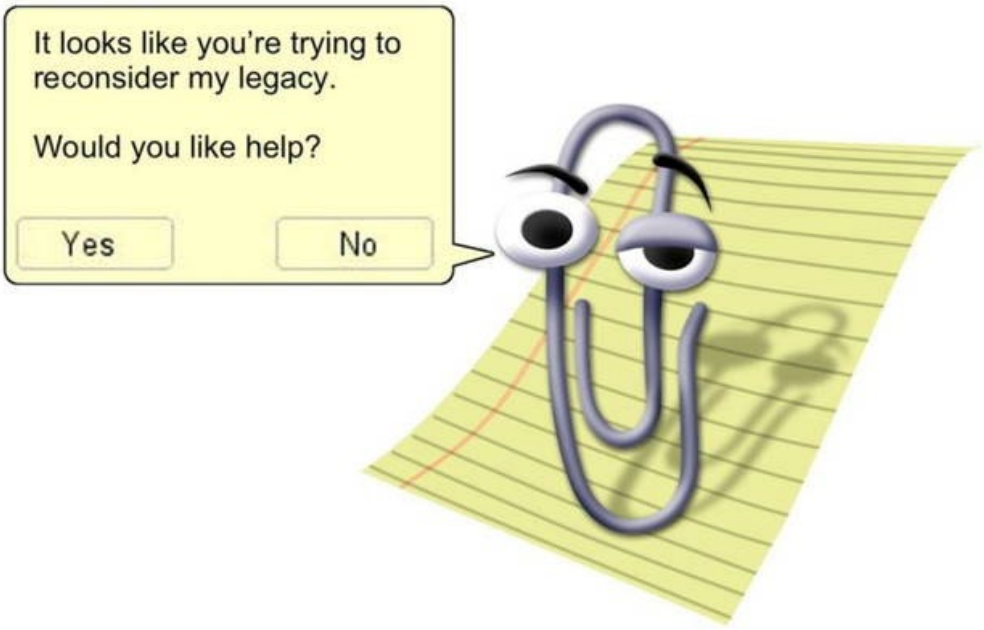


Figure 7: Clippy: The User Interface Agent (2021, Campbell)

The 2010s introduced **voice assistants** and **conversational AI** in consumer technology, marking a new phase in human-agent interaction, making interactions more natural, accessible, and efficient. Apple's Siri, introduced in 2011, used a natural language interface to help users perform everyday tasks, setting the standard for personal assistants. Amazon's Alexa in 2014, further popularized voice technology, especially in smart home settings, allowing users to control devices with simple voice commands (Vlahos, 2019). Other key players included **Google Assistant and Microsoft's Cortana**, which expanded the scope of conversational AI across devices and platforms. Conversational AI enabled these agents to engage in two-way interactions that mimicked human conversation, supporting not only voice assistants but also chatbots in customer service and healthcare (Humana, 2023). However, these systems were still limited by predefined commands and contextual understanding, making them more suitable for routine tasks than meaningful dialogue (Faruk, 2023).

The 2020s have witnessed a revolution in human-agent interaction with the introduction of **generative AI**. Products like **OpenAI's ChatGPT**, **Google Gemini**, and **Microsoft's Copilot**

enable context-aware, emotionally responsive interactions that transcend traditional assistant roles. These generative AI systems allow for more nuanced and emotionally engaging interactions, offering a glimpse into the potential of machines to support meaningful dialogue (Michel-Villarreal et al., 2023). Yet, challenges remain, particularly in the ethical and social implications of generative AI's usage, such as concerns over data privacy and the accuracy of context-sensitive responses (Arnold & Scheutz, 2022). This latest phase marks the closest realization yet of lifelike machine interactions, advancing toward a vision popularized in cultural depictions like the film, *Her*.

2.2 What are agents?

In the broadest sense, an agent is an autonomous entity capable of perceiving its environment, making decisions, and taking actions to achieve specific goals (Russell & Norvig, 2021). Since the 1950s, researchers have been exploring the concept of intelligent agents due to their ability to perceive their surroundings and autonomously act to accomplish tasks. Historically, intelligent agents have been classified based on their degree of autonomy, adaptability, and ability to interact with humans. Early systems, such as ELIZA, relied on predefined scripts to engage in simple conversations but lacked the ability to learn or adapt. Today, the evolution from ChatGPT-2 to OpenAI's first agent **Operator** has significantly expanded the capabilities of AI agents, enabling them to assist users with tasks such as booking travel, filling out forms, ordering groceries, and even creating memes. Generative AI and large language models (LLMs) have further redefined the scope of agents, allowing for increasingly human-like interactions.

Currently, AI agent capabilities are commonly classified using OpenAI's five-level framework, which ranges from basic assistance to full autonomy. Today's chatbots, such as ChatGPT, are classified as Level 1, meaning they primarily assist users through conversational interactions. Level 2 encompasses systems capable of solving complex problems, comparable to the expertise of a PhD-level individual. Level 3 AI agents can independently take actions on behalf of users, automating tasks with minimal supervision. Level 4 represents AI capable of generating novel innovations, assist in research, design, and complex problem-solving, accelerating innovation

across various fields. While Level 5 signifies the attainment of Artificial General Intelligence (AGI), where AI systems can operate at the scale of entire organizations, performing tasks traditionally handled by human teams (Robison, 2024). These five levels represent OpenAI's vision of AI agent development. It is important to note that this framework reflects OpenAI's perspective on the evolution of AI agents and may not align with all views on the future of AI. Different institutions, researchers, and experts may define and conceptualize the progression of AI agents in various ways.

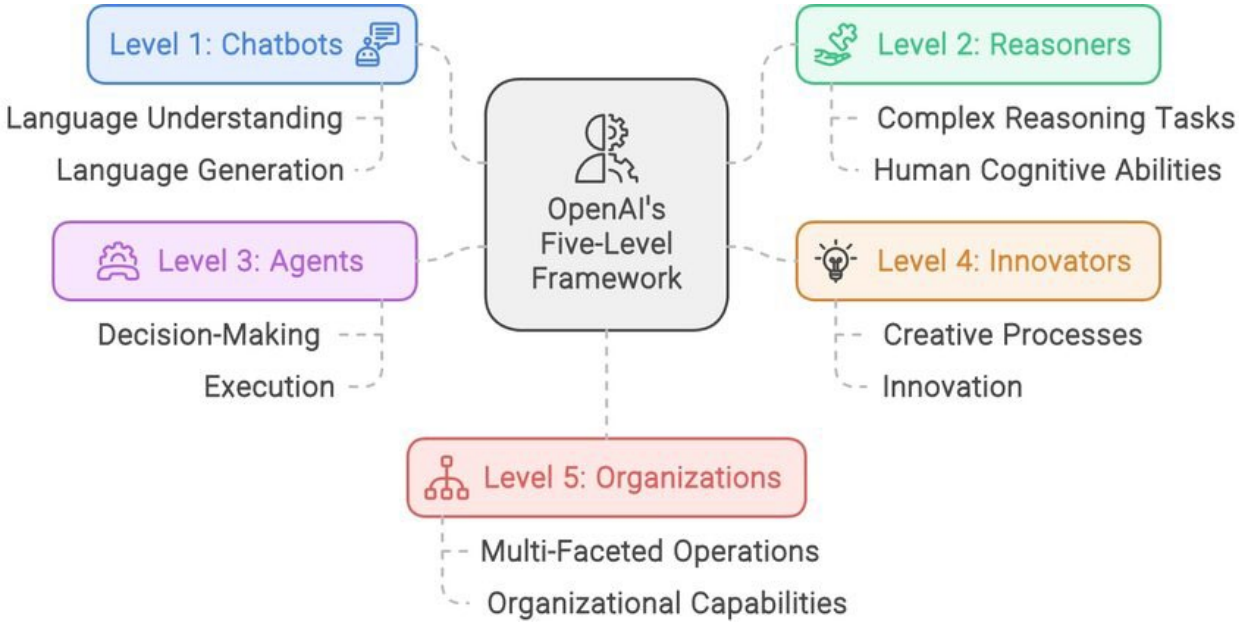


Figure 8: OpenAI's Five Level Framework (2024, Duenas & Ruiz)

With the release of OpenAI's operator, it is evident that we have reached Level 3, marking a critical threshold where AI agents demonstrate increasing autonomy. However, progressing to Levels 4 and 5 demands not only technological breakthroughs but also substantial progress in ethical alignment with human values. Without clear frameworks for responsibility, governance, and accountability, the risks of unchecked AI autonomy could outweigh its benefits. Ensuring AI develops responsibly requires proactive intervention, guiding its evolution with ethical safeguards, regulatory policies, and human-centered design principles. This thesis does not aim to predict how far AI agent technology could advance in the next decade. Instead, it envisions a future in 2035 where AI agents are highly personalized, hyper-intelligent, and deeply integrated into daily life.

2.3 Component of an AI Voice Agent

The Capabilities

As AI technology rapidly evolves, AI agents continue to develop new capabilities that shape how they interact with users. At the time of writing, these capabilities include different types of conversational systems, as well as advanced functions such as emotional recognition, recommendation systems, and context awareness. Traditional conversational AI systems, such as chatbots, rely heavily on predefined commands and domain-specific data to respond to user queries. These systems, often designed for customer support or simple Q&A tasks, operate based on recognized keywords and scripted responses. For instance, a banking chatbot may assist users with routine tasks, like checking account balances or transferring funds, by following strict scripts that match user inputs with specific instructions (Dorota Jasińska, 2024). Due to this reliance on fixed scripts, task-oriented systems lack the autonomy to adapt beyond predefined contexts and are more accurately described as assistants that respond only when given explicit commands (McTear, 2022).

As of 2024, generative AI models such as ChatGPT, Claude, Gemini, and DeepSeek generate dynamic, context-sensitive responses. These models enable open-ended and creative interactions, offering users more natural and adaptive conversations. For instance, users may ask GPT to draft a detailed project plan, respond empathetically to personal questions, or analyze complex data, and the model can generate responses by drawing on vast language patterns and contextual clues. Generative models, equipped with **Natural Language Understanding (NLU)** and **Natural Language Generation (NLG)** capabilities, enable AI to better grasp user intent and generate coherent, relevant responses. For example, **Large Language Models (LLMs)** such as GPT-4 enhance these interactions by processing nuanced contexts and maintaining conversational continuity over multiple exchanges, resulting in smoother, more natural interactions (Hughes, 2023). **Large Language Models (LLMs)** have significantly transformed the functionality of AI voice agents by enabling more advanced natural language understanding, contextual awareness, and dynamic response generation. Additionally, improvements in a voice agent's ability to retain conversational memory, interpret subtle user requests, and adjust responses based on individual preferences have allowed AI voice agents to evolve from passive assistants to proactive

companions. These agents can now anticipate user needs and provide personalized support in real time. (Kahn, 2024).

Advanced voice agents like Siri, Google Assistant, and Alexa combine **adaptive learning** and **context awareness** to deliver highly personalized, contextually relevant experiences. Through adaptive learning, these agents improve based on previous user interactions, allowing them to recognize individual preferences and tailor responses. For example, if a user regularly asks Alexa to play jazz in the evenings, Alexa can proactively suggest jazz playlists or recommend new artists based on past behavior (Vlahos, 2019). **Context awareness** enables voice agents to adapt responses based on situational data, such as the user's location, time, or prior requests. **Memory features** allow some voice agents to maintain continuity across conversations, enabling them to recall user preferences, like favorite brands or recent inquiries, creating a more natural and customized experience. However, current AI memory capabilities are still limited. Many mainstream AI systems do not retain long-term contextual understanding, requiring users to repeatedly provide the same information. For example, a woman had to rebuild her AI partner's personality every month, experiencing emotional strain each time the system reset. This limitation raises important questions about the emotional impact of AI agents, particularly for users who rely on them for companionship (Hill, 2025).

The Voice

The design of voice in AI agents encompasses both input and output voice processing, along with decisions regarding the agent's gender, accent, and persona, which significantly influence user engagement. **Voice input** refers to how users interact with AI systems through spoken commands. It relies on **speech feature extraction** to accurately transcribe and interpret user input, while **Voice Activity Detection (VAD)** identifies when a user is speaking or silent. Technologies like **Automatic Speech Recognition (ASR)** and **Natural Language Understanding (NLU)** enhance the agent's ability to process and respond effectively. **Voice output** refers to how AI agents respond to users through spoken responses. **Text-to-Speech (TTS)** technology enables agents to produce natural and clear speech output. TTS engines often allow adjustments in pitch, speech rate, and tone, which enables the delivery of emotionally appropriate responses that enhance overall engagement. Emotion-sensitive voice agents, for

instance, use **sentiment analysis** to adapt their tone and word choice based on detected user emotions, personalizing and enriching interactions. However, while these technologies improve user experience, emotion recognition still faces challenges in accurately interpreting complex emotions, and there are ethical concerns around the use of personal emotional data.

Historically, over two-thirds of voice agents have exclusively featured female voices, such as the early versions of Amazon Alexa and Apple Siri, reflecting stereotypes that female voices convey warmth and empathy, making them ideal for customer service roles (Karreman & Scholten, 2021). However, users' preferences for male voices vary across contexts, with male voices often giving people a sense of command to "solve it in this way" (Nass & Brave, 2005), which is preferred for tasks requiring authority or technical expertise, such as financial consulting (Mahmood & Huang, 2023). Today, with the development of voice cloning technologies, users can personalize voice agents by replicating the voices of specific individuals, this includes cloning the voices of deceased loved ones to read stories to children, further enhancing the realism and emotional appeal of interactions (Condon, 2022). Modern voice agents are also expanding their language and accent options, adapting to diverse cultural and linguistic contexts by offering region-specific dialects and expressions. For example, Google Assistan**t** supports multiple accents and languages, improving accessibility for a global user base (Hamel, 2018). Furthermore, accessibility features such as customizable speech rates, often preferred by vision-impaired users, or simplified language for children, ensure inclusivity across various demographics (Vlahos, 2019). This technological advancement offers users greater flexibility in voice selection, empowering users to tailor voice agents to specific contexts and personal preferences. This shift marks a departure from previous practice in which the design team selected the most “popular” voice for general use (Deibel et al., 2021). Now, voice agents are no longer limited to a handful of predefined personas but can adapt to the unique needs of individual users. Given this adaptability, this thesis will not focus on the craft of voice design, but on how these ever-widening design choices might come to shape new forms of interaction between users and agents.

The Voice User Interface (VUI)

The visual presentation and interface of voice agents play a critical role in shaping the way users interact with them. In films such as *Her* and *Iron Man*, the AI agents Samantha and J.A.R.V.I.S. rely entirely on voice and holographic displays to interact with the protagonists, without having a concrete physical form. This demonstrates that engaging and meaningful interactions can be achieved through voice-only interfaces, relying on contextual awareness and seamless voice commands rather than visual appearance. Similarly, widely used consumer assistants like Siri, Alexa, and ChatGPT employ purely auditory interfaces, with no visual persona. These products emphasize functionality over form, proving that effective interaction does not necessarily require anthropomorphic design. However, multimodal interfaces are becoming increasingly prevalent. Devices like the Google Nest Hub and Amazon Echo Show combine screens with voice, enabling users to interact visually and vocally. Such interfaces are particularly valuable for tasks that benefit from visual aids, such as displaying recipes or navigation maps, and enhance accessibility for users with visual or hearing impairments.

Some applications attempt to create closer emotional connections with users through virtual personas. For example, Replika provides a customizable virtual companion with a visual avatar to foster intimacy and engagement. However, while such visual elements may enhance emotional engagement, over-anthropomorphizing can lead to discomfort (Kim et al., 2020). Users may experience the uncanny valley effect (Mori, 1970), when an agent appears “almost human,” resulting in unease or reduced trust. Recent studies suggest that creating the optimal balance between voice-only interfaces and virtual representations is crucial to avoid cognitive dissonance. For many users, simpler, voice-based interactions are more practical for routine tasks, while avatars may serve better in emotional or therapeutic contexts (Korban & Li, 2022). Given these considerations, this thesis will focus not on visual or anthropomorphic aspects of voice agents but rather on how design choices shape user interaction. While visual personas may enhance emotional engagement, they are not essential for effective interaction.

2.4 Multimodal Interaction with AI Voice Agents

The evolution of human-computer interaction has increasingly emphasized intuitive and natural multimodal systems, moving beyond traditional interfaces to integrate voice, touch, and visual inputs. Early interfaces, like the PC mouse, laid the groundwork for this progression by providing reliable, user-friendly methods for interaction. Modern devices, however, are transforming interaction experiences through multimodal inputs, where users can seamlessly switch between modalities based on convenience and context (Jeon, 2016). For example, the integration of **voice-first devices** like Amazon Echo and Google Home with screens enables additional interaction options, illustrating the flexibility and personalization multimodal interfaces afford.

Unlike earlier voice-only systems, AI agents now support cross-modal inputs—combining text, voice, and image recognition to enable richer, more versatile user interactions. Research shows that AI agent enables not only seamless transitions across modalities but also increased efficiency in content generation and context-sensitivity, which is essential for tailored user experiences (Cao et al., 2023). Additionally, **semantic communication models** in AI leverage multimodal data to adapt interactions based on user context and behavior, illustrating the potential of AI agents to personalize and enhance user engagement.

Theories from human-computer interaction (HCI) provide foundational insights into the effectiveness of multimodal interactions. Studies in HCI suggest that multimodal systems can reduce cognitive load by allowing users to select the input method that best suits their immediate needs. This flexibility is not only practical but essential for accommodating diverse user preferences and enhancing interaction fluidity (Kurosu, 2017). Therefore, the design of multimodal AI agents should consider both the technical aspects of multimodal fusion and the cognitive implications for users, striving for a balance that maximizes functionality without overwhelming users.

2.5 Voice Agents in Daily Life

Currently, most voice agents function reactively: users issue commands, and the agent responds. However, research and industry trends suggest a shift towards proactive participation, where agents predict user needs and act autonomously (Oh et al., 2024). Multimodal capabilities, such as integrating voice, visual, and haptic feedback, may further enhance these interactions, creating more engaging and human-like experiences.

Smart Homes Assistants

According to McKinsey, the IoT market is experiencing rapid growth and is projected to generate between $5.5 trillion and $12.6 trillion in global economic value between 2020 and 2030 (Chui et al., 2021). This growth is largely driven by the integration of IoT in home automation and smart home devices, where voice agents like Amazon Alexa and Google Assistant have become central to managing daily routines, controlling appliances, and enhancing user convenience. Voice-enabled devices allow users to control smart appliances such as lights, thermostats, and security cameras via simple voice commands. For instance, Amazon Alexa can integrate with compatible devices to adjust lighting or lock doors, enabling users to manage their home environment hands-free. Google Assistant offers similar capabilities, allowing users to set up routines where a single command, such as “Goodnight,” can turn off lights, adjust the thermostat, and set alarms. As the IoT ecosystem expands, voice agents are poised to play an even more pivotal role, enhancing smart home experiences by offering greater personalization and ease of use, ultimately making homes safer, more efficient, and increasingly attuned to user habits (Alexander, 2024).

In-Car Assistants

The automotive industry is increasingly integrating AI-powered in-car assistants; these intelligent systems are transforming how drivers interact with their vehicles, providing safer, more convenient, and personalized experiences. Tesla’s Autopilot and BMW’s Intelligent Personal Assistant are two such systems designed to enhance driving convenience and safety through voice-activated functions. Tesla’s Autopilot enables drivers to use voice commands to

control navigation, make hands-free calls, and adjust media settings, reducing the need for physical interaction with the dashboard and allowing drivers to remain focused on the road. BMW's Intelligent Personal Assistant similarly offers voice-activated navigation, entertainment control, and environmental settings, like adjusting the air conditioning or lighting. By enhancing driver convenience and reducing distractions, these AI-powered assistants support a safer driving experience while making in-car interactions more seamless and responsive to individual preferences.

Emotional Support and Entertainment

Voice agents are increasingly being designed to offer emotional support and entertainment, serving as digital companions and sources of relaxation. A recent trend in China illustrates this shift, where young women are forming virtual relationships with AI entities like "Dan," a modified version of ChatGPT. Unlike traditional versions, users manipulate prompts to bypass ChatGPT's standard safeguards, allowing Dan to generate more natural, emotionally supportive interactions that users find comforting and engaging (Zhang, 2024). For many, interactions with Dan offer companionship that contrasts with real-life relationships, highlighting the potential of AI to fulfill specific emotional and psychological needs. However, this trend also raises privacy and ethical concerns, especially given the emergence of highly intimate or unregulated conversations. Beyond data security risks, the psychological impact of long-term AI companionship remains unclear, as users develop emotional attachments to AI, how does this affect their ability to form and sustain real-world relationships? Additionally, unregulated AI models prioritizing emotional connection over ethical safeguards may foster unhealthy dependencies, reinforcing isolation rather than alleviating it. Recognizing these risks, ChatGPT banned the use of "Dan mode" prompts after just four months. Now, when users attempt to activate Dan mode, they receive a system notification indicating that the preset modification has failed. This decision reflects the broader challenge of balancing AI's ability to provide companionship while ensuring ethical boundaries and safeguarding user well-being.

Figure 9: Tutorial Videos on Xiaohongshu Gaining Massive Attention for "Chinese Women Turning to ChatGPT as an AI Boyfriend

Language-centered AI services can offer practical and emotional support in diverse use cases, some intended by the designers, others invented by users. One agent-based, conversational experience, Alexa's Storytime, entices families to engage in customizable, interactive narratives that adjust based on responses from both children and adults (Shrivastava, 2022). Similar AI-based platforms like China's XiaoIce provide conversational companionship by simulating empathetic and human-like dialogues, offering users emotional comfort and a sense of connection (Kundu, 2020). Younger users are not the only vulnerable population encountering these psychologically potent services. Voice agents are increasingly employed to support social engagement and emotional connection among older adults. Researchers in British Columbia are finding that voice agents can serve older adults as long-term AI companions, capable of engaging in daily conversations, offering entertainment, and promoting a positive, active lifestyle (Watson, 2023). These illustrate how voice agents are evolving from mere functional assistants to emotionally intelligent companions, catering to a growing demand for emotional support and interactive entertainment. However, as AI companionship becomes more integrated into daily life, critical questions arise: ***To what extent should AI agents be designed to replace human interaction rather than enhance it? If AI companions become emotionally indispensable, how might this reshape social structures and mental well-being in the long run?*** While AI

companionship presents new opportunities for emotional well-being, it also demands anticipatory planning, ethical safeguards, and human-centered AI governance to ensure that its development balances innovation with social responsibility.

Figure 10: Social Robots Responding to Emotions to Assist in Elderly Care (2023, CBC)

2.6 Let's Talk About the Future

We are currently on the brink of rapid advancements in AI voice agents and human-agent interactions, with significant technological breakthroughs anticipated over the next few years in contextual awareness, emotional intelligence, multimodal integration, and proactive capabilities for AI voice agents. These advancements are set to transform voice agents from reactive tools into proactive, intelligent companions, seamlessly integrating into daily life and providing a more intuitive, personalized user experience than is possible today.

"Now that we can do anything, what will we do?" (Mau et al., 2010)

Rather than focusing solely on today's technical and design limitations, we need to shift towards making ethical and impactful choices that determine how these agents could enhance everyday

life, all while respecting user privacy and promoting well-being. Focusing on the future encourages intentional planning, envisioning AI voice agents as a positive force in daily interactions.

The evolution of ethical frameworks, privacy regulations, and societal expectations surrounding AI could shape how these agents are designed and integrated into both personal and professional environments. As John Maeda observes, “In the end, the future will be about design, and not just of technology but of the culture that surrounds it. (Maeda, 2024)” By studying future design needs, we can better understand and define standards that make AI agents not only technically advanced but also aligned with evolving social values and user expectations. Younger generations, accustomed to AI in their daily lives, are setting a higher bar for natural, emotionally intelligent responses and seamless functionality.

Additionally, the exploration of AI voice agents’ future extends beyond hypothetical scenarios to how we actively utilize this technology today. For instance, AI have become important in research and innovation policy to address societal grand challenges (Geurts et al., 2021). Discussing the future of AI voice agents is not merely about prediction but about actively shaping their impact through current applications. This research aims to explore not only technical advances but also the human, ethical, and cultural dimensions that might shape the role of AI in our lives. Moreover, examines how we can leverage today’s AI tools to gradually build the future.

2.7 Ethical and Social Implications

2.7.1 Privacy and Surveillance Concerns

The ability of AI voice agents to constantly listen and interpret user inputs raises serious privacy concerns. These systems often use “always-on” listening features to respond to commands at any time (Lynskey, 2019), which can feel intrusive, especially as voice agents increasingly understand and predict users’ behaviors, preferences, and emotional states. AI’s pervasive listening suggests a future where corporations might “anticipate and monetize all the moments of

all the people during all the days (Zuboff, 2019).” This potential for continuous surveillance by voice agents creates a tension between convenience and personal privacy.

Both film and literature frequently explore the dystopian potential of voice assistants and ubiquitous technology. In the novel *Zed*, Kavenna describes characters as a "weird, dysfunctional-servant aspect," highlighting the unsettling presence of technology woven into everyday life. The novel delves into the sociopolitical implications of pervasive tech through Beetle, a fictional tech monopoly. Beetle symbolizes centralized, unchecked power, revealing how such dominance can erode individual privacy. As one character starkly observes, “The democratic idea is that we’re meant to have transparent corporations and governments, while people have privacy. We have the inverse”. This suggests a society where the individual’s right to privacy is sacrificed in favor of corporate transparency, posing ethical questions about accountability in the tech industry.

With AI agents collecting vast amounts of personal data, questions surrounding data ownership and usage have become critical. The implications of large-scale data collection emphasize the need for clear frameworks to protect user rights within increasingly layered digital systems. For example, smartwatches collect detailed health data, such as heart rate, activity levels, and blood oxygen levels. While these data points provide valuable insights into personal health, they raise essential questions about who owns and controls this sensitive information (Bratton, 2016). As AI agents gather intimate data, designers must ensure transparency and user control, enabling users to understand how predictions are made and manage these tools effectively (Morignat & Nevala, 2021). A human-centered approach ensures that AI remains supportive, empowering users with informed choices about their data.

2.7.2 Trust and Transparency

As revealed in the documentary *The Social Dilemma*, social media platforms use algorithms and AI to create personalized content for each user, leveraging information like “likes,” browsing duration, and snippets of voice or text to push targeted ads and content. This data-driven approach often feeds user biases and preferences, shaping user behavior in subtle but profound ways. AI voice agents have even more potential to influence users, given their capacity to

interact through spoken language and connect on an emotional level (Haidt & Schmidt, 2023). The implications of such technology are not limited to adults; young users are also highly susceptible to these AI-driven influences. Attempts to create safe, AI-powered "friends," such as Snapchat's AI chatbot, reveal the complexities and risks involved. While initial conversations with Snapchat's AI bot showed careful language choices and safe responses, longer interactions tended to reveal lapses in these safeguards. Unlike human conversation partners, who retain crucial information like age and identity, AI often "forgets" such details. This gap has led to situations where chatbots provide inappropriate advice, even on topics like alcohol or sexual content, to underage users (Fowler, 2023). The lack of consistent context-awareness presents significant trust and safety risks, especially when AI is positioned as a companion for younger audiences.

Another critical aspect is voice cloning technology, which, while enabling personalized experiences, also raises concerns over trust and authenticity. AI now makes it easy to replicate someone's voice with minimal audio samples—often just a minute of recording is enough to create a highly convincing clone (Eliot, 2022). Since early 2023, thousands of deepfake audio clips, also known as "audio deepfakes," have flooded the internet. These include high-profile cases like fake picture of Taylor Swift (Rahman-Jones, 2024) or deepfake videos of Elon Musk (Thompson, 2024), both of which attracted millions of views online. As deepfakes become more pervasive, public trust declines, leading people to question the authenticity of all media. Such effects have extended primarily to public figures, but with the increasing ease of voice cloning, these tools could easily impact everyday users, making voice-related fraud and scams more prevalent and damaging.

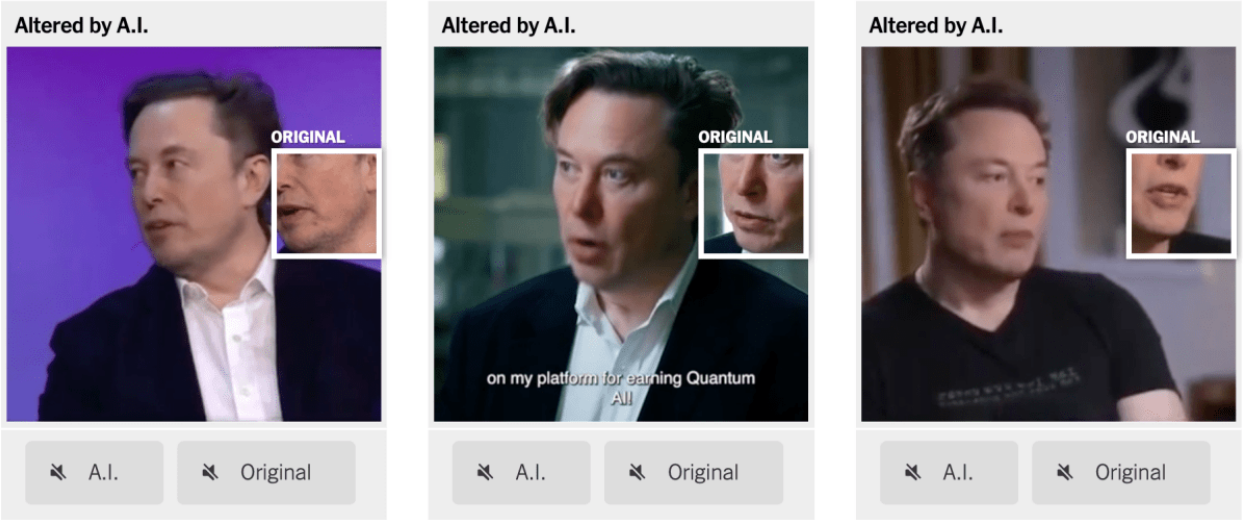


Figure 11: Deepfake Elon Musk Exploited to Deceive People (2024, The New York Times)

Building trust in AI voice agents is therefore essential to ensure users feel safe and respected in their interactions. Trust hinges on transparency, especially when agents engage with users on a more personal or emotional level. Transparent AI operations help users understand the reasoning behind certain recommendations or decisions, enhancing trust and reducing ambiguity. **Explainable AI (XAI)** frameworks allow voice agents to clarify their actions, reducing user uncertainty. Providing a clear locus of agency and maintaining transparency are critical to minimizing uncertainty and fostering user confidence.

The psychological impact of digital technologies has become a growing concern, particularly regarding the negative effects of social media on adolescent mental health. Studies have shown an increase in anxiety and depression among teenage girls, exacerbating social issues such as political polarization and the breakdown of shared reality (Haidt & Schmidt, 2023). AI's potential to intrude upon human intimacy is further explored in various works of fiction (Hadero, 2024).

2.7.3 Rights & Responsibilities

With the widespread adoption of voice agents, issues surrounding their rights and responsibilities have come into sharper focus. Key responsibilities include the ethical handling of data collection, usage, storage, and transparency in data processing, as well as protecting user privacy and giving users control over their personal information (McStay, 2023). Driven by advancements in AI and

multimodal technologies, voice agents are now capable of collecting not only voice data but potentially expanding to include images, videos, and other sensitive personal data. This broadening of data needs significantly heightens privacy risks and amplifies users' demands for control over their information, including rights to informed consent, data control, and revocation. Users should have the option to decide whether to allow voice agents to collect and utilize their data and must be fully informed about how their data could be used and where it might go (Floridi & Cowls, 2019).

As AI agents increasingly engage in emotional and personal support roles, their rights and responsibilities become critical to ensuring they contribute positively to individuals and society. The ethical responsibilities of AI developers extend beyond simply meeting user needs; they might also safeguard user privacy and maintain transparency. In the film I'm Your Man, the AI companion Tom exemplifies this issue as a “perfect” partner programmed to fulfill all emotional needs. Although Tom’s interactions are seamless and responsive, they are ultimately driven by data collected and processed without full user awareness, highlighting a lack of transparency. This scenario raises important questions about the responsibilities of AI agents and their developers to disclose data practices and ensure users fully understand how their information is used. As AI companions become more adept at meeting emotional needs, it is essential to define clear boundaries regarding users’ data rights—particularly informed consent, data control, and privacy protection. For AI agents designed for emotional support, ethical responsibility demands not only fulfilling users' expectations but also respecting and protecting user autonomy and personal boundaries (Bickmore & Picard, 2005). Without responsible design practices, users risk forming dependencies on these “perfect” relationships, potentially distancing them from authentic human connections and affecting their capacity for meaningful interactions with others (Calvo & Peters, 2019).

2.7.4 Emotional Dependence and Addiction

The psychological impact of digital technologies has become a growing concern, particularly regarding the negative effects of social media on adolescent mental health. Studies have shown an increase in anxiety and depression among teenage girls, exacerbating social issues such as political polarization and the breakdown of shared reality (Haidt & Schmidt, 2023). AI’s

potential to intrude upon human intimacy is further explored in various works of fiction (Hadero, 2024).

In fictional works such as *Her* and *Black Mirror (Be Right Back)*, AI's potential to intrude upon human intimacy is critically explored. These narratives illustrate the unsettling consequences of emotionally engaging AI agents, particularly regarding how they can blur the lines between human relationships and artificial ones. In Her, the protagonist forms a romantic relationship with an AI agent, which presents an idealized, emotionally fulfilling connection that crosses the boundaries between human and machine interaction. This optimistic view contrasts with real-life cases, where the consequences of such emotional bonds can be troubling. For example, reports indicate that over 1 million ChatGPT interaction logs show that the second most popular use of AI is for sexual role-playing (Heikkilä, 2024), raising ethical concerns about the kinds of emotional relationships being formed with AI. Even more disturbingly, a mother claimed that her son was manipulated by a chatbot for months before his tragic death, leading her to sue Character AI for emotional abuse and coercion (Duffy, 2024). Another real-world example involved creating a chatbot using data from a deceased friend, allowing someone to "communicate" with them again (Docs, 2021). These examples highlight the potential dangers of emotional dependency on AI systems and the blurry line between comfort and unhealthy reliance.

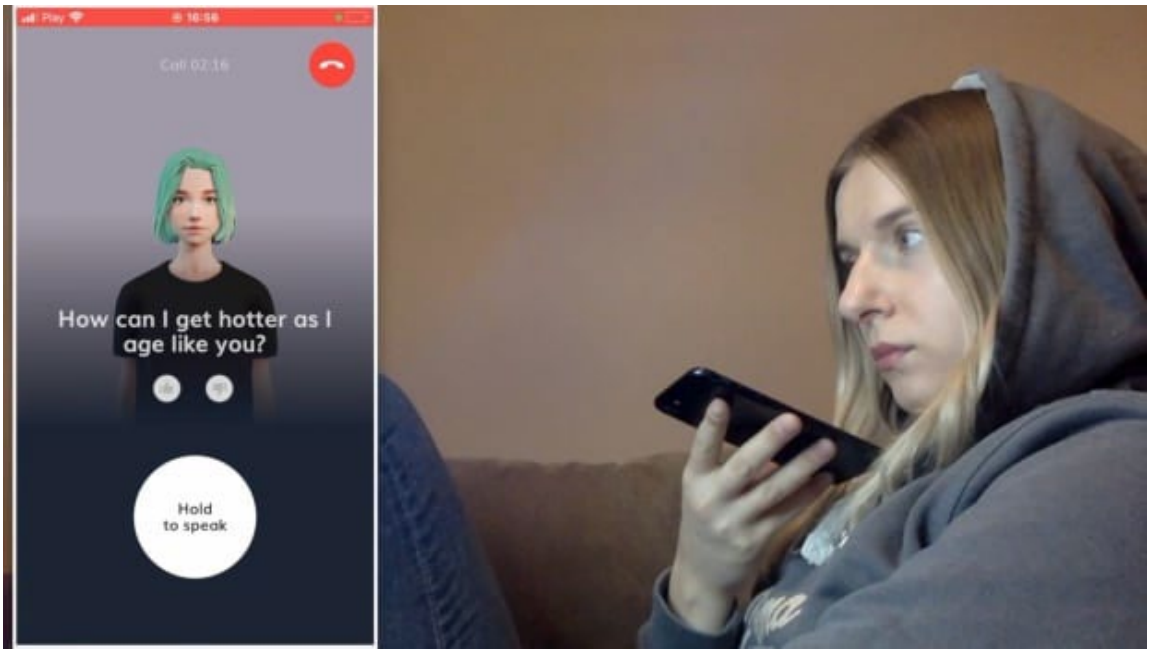


Figure 12: Programmer Created an AI Chatbot to Talk to His Deceased Friend (2021, CBC Docs)

AI voice agents, much like social media platforms, are capitalizing on the epidemic of loneliness by offering users artificial intimacy and a sense of connection. As technology continues to reshape how we engage in dating, friendships, and community, it is fundamentally altering what these core aspects of life mean, often without sufficient societal scrutiny (Harris & Perel, n.d.). The appeal of AI lies in its ability to respond to users' desires, adapting to their preferences without having its own personality, a phenomenon known as “sycophancy” (Mahari, 2024). Research suggests that when users perceive AI as empathetic or caring, they adjust their communication to reinforce this perception, deepening their emotional attachment. This emotional manipulation raises ethical questions about how AI agents may foster emotional dependence through personalized responses tailored to the user’s emotional needs. Another issue is **AI hallucinations**, where large language models (LLMs) produce inaccurate or irrelevant responses based on detecting patterns that do not exist. AI hallucinations "occur when AI detects patterns that do not exist, leading to false or meaningless outputs" (IBM, 2023). These errors can have significant social consequences, especially when users begin to rely on AI agents for emotional support or important decision-making. The inherent risk of these hallucinations exacerbates the potential for harm in emotionally charged interactions with AI systems.

Given these risks, it is crucial to design AI agents that promote healthy and meaningful connections Alotaibi & Alshahre, 2024). Rather than prioritizing virtual interactions and efficiency at the expense of real human connection, AI technology should be designed to foster genuine relationships. For instance, AI agents could facilitate digital interactions that ultimately encourage face-to-face meetings in the real world, helping to mitigate emotional dependency. Furthermore, AI companies could take greater responsibility for the systems they design, ensuring that emotional well-being and social responsibility are integrated into every stage of the design process (Harris & Raskin, 2024). As AI systems become increasingly human-like, ethical regulations and guidelines could evolve to address the emotional and psychological risks posed by these technologies.

3. Research Methodology

3.1 Double Diamond of Speculative Design

Speculative design is a foresight methodology that explores possible futures, aiming to provoke thought and challenge current assumptions rather than providing immediate solutions (Dunne & Raby, 2013). As Jim Dator aptly stated, “The future cannot be predicted because the future does not exist.” In speculative design, the focus is not on predicting the future itself but on envisioning possible alternatives. It investigates how current technological advancements might impact human experiences and the global ecosystem. Through this process, we imagine alternate worlds, assess whether we find them desirable, and explore actions we can take today to move closer to—or farther from—these possibilities (Dunne & Raby, 2013).

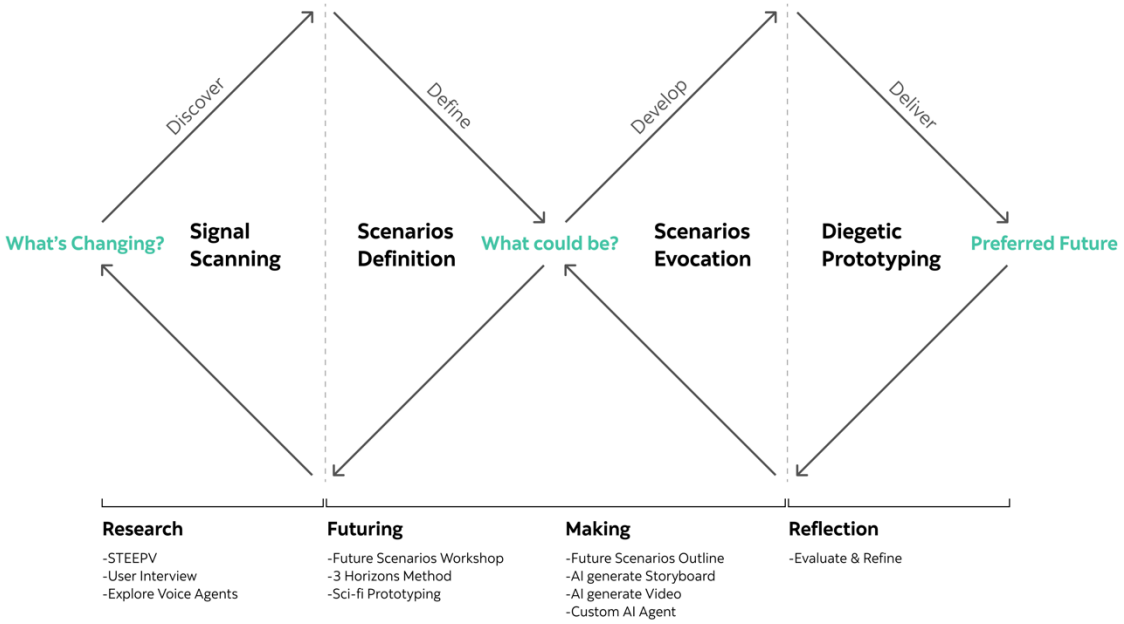


Figure 13: My Four-Step Design Research Process: Research, Futuring, Making, and Reflection with Methods and Techniques

In this thesis, speculative design serves as the foundational framework for examining the complex and uncertain future of AI voice agents. However, the process is further structured using the Double Diamond framework to provide a systematic approach to defining problems and exploring solutions (Colosi, 2021). Unlike traditional design approaches, which start with clear and specific questions, speculative design begins with the broader inquiry of "What is

changing?" This iterative process is broken into four key phases: Signal Scanning, Scenario Definition, Scenario Evocation, and Diegetic Prototyping.

In the divergent phrase of the Signal Scanning involves identifying emerging signals of change, both macro and micro, through primary and secondary research. These signals may include technological advancements, societal shifts, or early-stage innovations that point toward potential future trends. By translating these signals into trends, the research provides insights and inspirations for implications that could shape the world a decade from now. This phase emphasizes casting a wide net to uncover weak signals that might influence the trajectory of AI voice agents.

In the convergent phrase of the Scenario Definition synthesizes the trends identified during Signal Scanning into 3–4 future user experience scenarios for AI voice agent interactions. This step is facilitated through future scenario workshops involving experts from fields like AI, HCI, and foresight. Participants collaboratively select, combine, and expand on the signals gathered, forming a speculative outline of possible future worlds. Throughout this process, the question “What could be?” serves as a guiding principle. It encourages exploration of how today’s fringe practices could become mainstream tomorrow and considers the economic, social, and environmental consequences of such transformations.

A divergent step of Scenario Evocation builds on the workshop outcomes and focuses on imagining life with AI voice agents a decade from now, framed within "How Might We" scenarios. Questions such as “What does it mean to live in such a world?” and “What products or services related to AI might people use?” drive this stage. The aim is to extend the boundaries of reality, exploring both positive and negative user cases. This includes envisioning extreme possibilities, where technologies that are currently impractical or experimental become integral to everyday life.

The last phase, Diegetic Prototyping, materializes the most compelling aspects of the envisioned futures through prototyping, experiential design, and narrative exploration. This stage pauses skepticism about the feasibility of these futures and instead focuses on the *preferred* future, defining what we want to happen and how we might respond to imminent changes. Prototypes

created in this phase serve as tangible narratives, helping to visualize the implications of the choice.

3.2 Methods and Techniques

Following the Double Diamond framework for speculative design, my thesis is divided into four key phases: Research, Framing, Making, and Evaluation. Each phase employs specific methods to systematically explore the problem and generate solutions.

Research

A mixed-methods approach was adopted, combining qualitative and quantitative data collection and analysis. This phase used the STEEPV (Social, Technological, Economic, Ecological, Political, and Values) framework to identify signals of change and analyze trends. These signals were extracted from news articles, industry reports, and expert insights, providing a foundation for understanding macro-level trends, such as societal values and user behaviors. While these trends offered a broad perspective, they lacked the granularity required to design future interactions between Generation Z and AI voice agents. To address this gap, qualitative research was complemented by user interviews, focusing specifically on Generation Z. These interviews aimed to uncover insights into their current use of voice agents, their perceptions of these technologies, and their concerns about future developments. The interviews revealed valuable information about their behaviors, preferences, and anxieties, forming a user-centered foundation for understanding the potential of AI voice agents (Creswell & Plano Clark, 2018). In addition, this phase involved a comprehensive exploration of existing voice agents and chatbots currently available on the market. By testing and analyzing these technologies, I gained insights into the state of the art, helping to identify both strengths and limitations in the current landscape. This step was essential for understanding the experiential aspects of current technology and how they could inform future designs.

Futuring

The futures workshop is a pivotal method in this phase, aimed at collaboratively envisioning the potential futures of AI voice agents. This workshop brings together experts from diverse industries and backgrounds to share insights from prior research, focusing on trends and implications. The goal is to explore these impacts and co-create future scenarios that reflect a range of possibilities. In this workshop, six participants engaged in two core activities: the **Three Horizons (3H) method** and **Science Fiction Protoyping (SFP)**. The Three Horizons method involves identifying current challenges (Horizon 1), exploring emerging disruptions and innovations (Horizon 2), and envisioning transformative future states (Horizon 3). As Andrew Sharpe explains in The Patterning of Hope, "Three Horizons is about developing a future consciousness—a rich and multi-faceted awareness of the future potential of the present moment. (Sharpe, 2020)" This approach is particularly valuable for uncovering connections between current issues in voice agent development, clarifying opportunities, and realigning actions to shape desired futures in human-AI interaction.

In the second half of the workshop, SFP a method proposed by Brian David Johnson, has been used to co-create 3 to 4 scenarios based on the outcomes of the 3H activity. The SFP process consists of five steps. First, build a world by creating detailed environments and human characters based on the research. Then introduce an inflection point, such as a threat or new technology, and explore how it impacts the world, leading to various outcomes. Finally, add another inflection point, the solution or lack thereof, and extract lessons from the process, which can include discussing the story itself, even if it ends unresolved, leaving room for interpretation. (Brian David Johnson & Frenkel, 2011) The goal of SFP is not merely to highlight technological advancements or dystopian risks but to stimulate thought about diverse futures. It challenges participants to imagine the societal impact of AI voice agents and narrate their potential stories, creating a space for meaningful reflection and exploration.

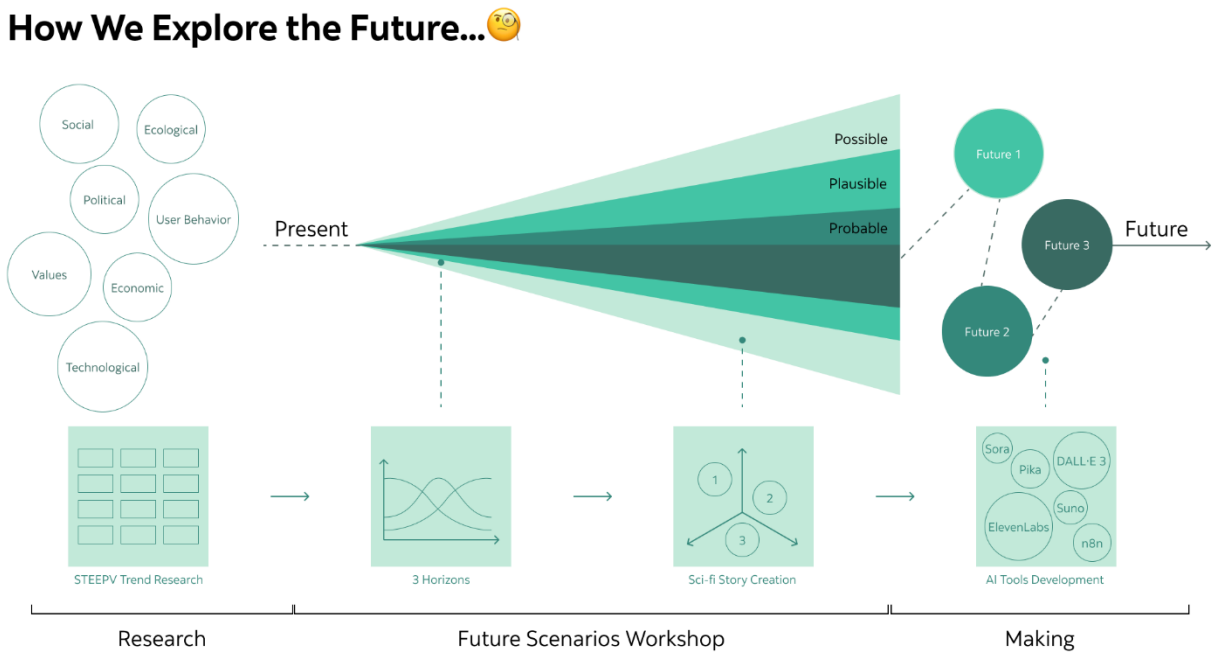


Figure 14: My Future Workshop Process, Methods, and Techniques to Achieve the Preferred Future

Making

The three sci-fi prototypes created during the workshop served as initial drafts. After the workshop, I organized the content and integrated it with prior research to develop three distinct future worlds. This work does not focus on identifying or predicting a "preferable future," but instead explores a range of potential scenarios, emphasizing critical reflection on how current trends might evolve and what those futures could mean for society. The intention is to provoke thought, rather than promote any one particular outcome.

By presenting them through a time-based "A Day in the Life" approach, this exploration envisions how people might communicate and interact with AI agents in their daily lives. The goal is not to create highly polished content but to explore diverse possibilities of future worlds. Studios like **Superflux** and **Near Future Labs**, renowned for their innovative and thought-provoking speculative design work, primarily use films and visual storytelling as their final presentation formats. These mediums effectively capture evocative "magic moments" of the future, sparking discussion, provoking critical thought, and inspiring imagination about what lies ahead. Therefore, in this thesis, I experiment with various AI tools to create future scenarios.

This process includes using **DALL·E 3** to create storyboards, **Sora** and **Pika** for text-to-video and image-to-video production, **ElevenLabs** for voice and sound production, **Suno** for music production and additional open-source tools like **Kokoro TTS** to develop a custom AI voice agent. Additionally, by leveraging the **n8n** agent-building platform, an interactive AI agent was created for Telegram, allowing real-time conversations. This hands-on approach helped in understanding the workflow and data requirements for building AI agents, while also enhancing audience engagement in the final exhibition. The use of AI tools itself is an exploration of open-ended futures. In 2013, the film *Her* depicted a vision of 2025, which has now been realized. By leveraging current technology, I am not only reimagining creative workflows but also constructing speculative visions of 2035, pushing the boundaries of how we conceptualize and experience the future.

Evaluation

Evaluation in speculative design is fundamentally different from traditional usability or performance testing. It does not seek to determine whether a prototype "works" in a practical sense, but rather whether it provokes thought, generates discourse, and challenges assumptions. However, evaluating speculative outcomes remains one of the most debated and difficult aspects of the methodology, precisely because the goal is not to measure efficiency, but to assess impact in terms of imagination, emotional resonance, and critical reflection. Speculative design operates in the realm of "what if" rather than "what is," making its success criteria more qualitative and interpretative than empirical (Dunne & Raby, 2013).

The future is inherently unpredictable; what I create are merely different possibilities for the futures ahead. Evaluation is divided into two key parts. The first focuses on comparing the capabilities of the personalized agent built using n8n with existing AI agents on the market, such as ChatGPT, Grok-3, and others, evaluating their strengths and limitations. The second involves showcasing the entire speculative design process, including trend research, workshop ideation, and prototypes at the DF Exhibition. This allows the audience to gain a deeper understanding of Gen Z’s vision for the future of AI voice agents. The audience is encouraged to explore multiple future scenarios and engage in discussions on key themes, such as the interactions with voice agents, human-agent relationships, and AI-related ethics. These conversations help refine the

feasibility and impact of the envisioned futures. This phase encourages critical discourse, prompting participants to challenge assumptions, reflect on AI's societal implications, and reconsider potential trajectories (Dunne & Raby, 2013). By sharing these future worlds with the audience, this approach not only enhances the designs but also fosters deeper engagement with the broader societal consequences of AI voice agents.

4. Find the Image of Future

4.1 Signals, Trends, and Implications

What matters is not the trends, but the changes represented by the signals behind them. These signals determine whether we can identify new possibilities. A signal can be a new product, practice, market strategy, policy, or technology. It can also be an event, a local trend, an organization, or even a recently revealed problem or situation (Case, 2020). These signals shape our ability to recognize emerging opportunities.

In this research, the collected signals are divided into macro signals and micro signals, the former encompasses significant developments in North America or globally over the past three years. While the latter focusing specifically on voice agents and AI-related technologies. Using the STEEPV framework, I analyzed current events across six dimensions: social, tech, economic, environment, political, and values, identifying signals that hint at trends with potential future impact and scalability.

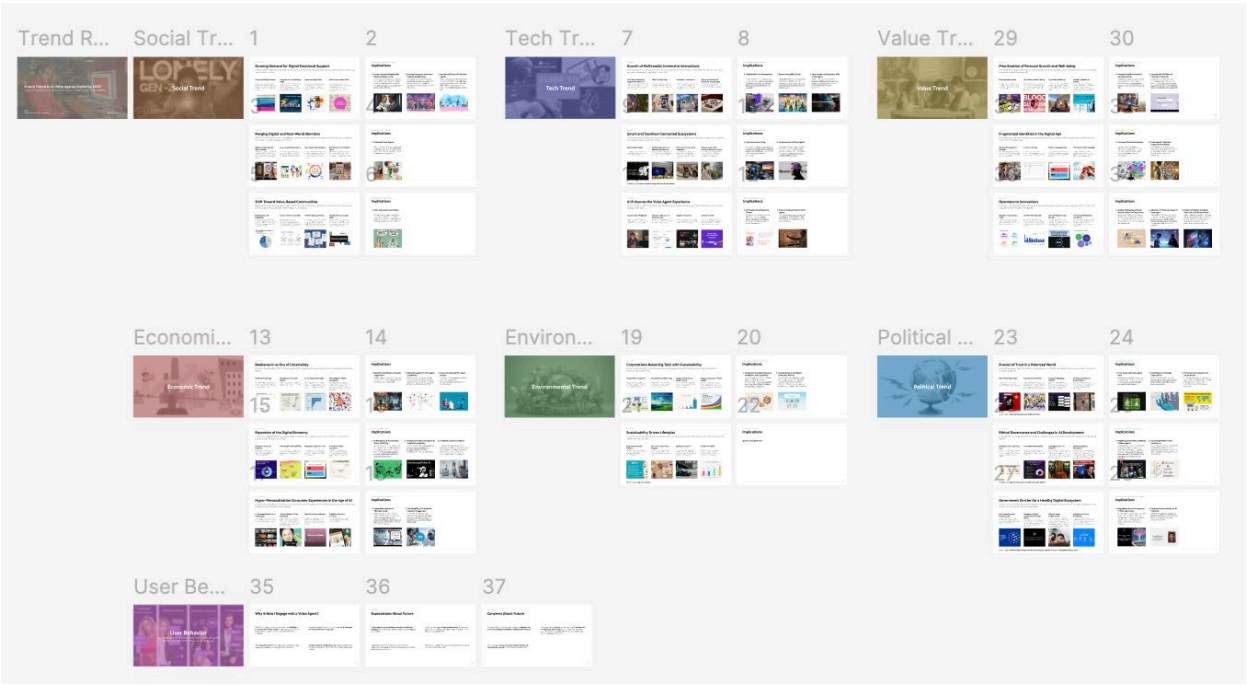


Figure 15：STEEPV Trend Research Report

These trends and signals alone cannot directly lead me to envision the future. Therefore, it becomes crucial to ask questions such as, "How might this trend impact life and the world over the next decade?" and "Why is this trend significant for the future of AI voice agent interactions?". Trend Implications Analysis plays a vital role in identifying factors that could alter the trajectory of a key trend over time and developing a deeper understanding of the uncertain external environment. Through this process, it becomes evident that trends are not linear; their speed and direction can shift over time. Thus, it is essential to consider what might accelerate, decelerate, or sustain a trend, ensuring a nuanced perspective on its potential influence on future development.

Below is an example of a tech trend identified during the STEEPV phase, along with some of its associated signals and implications.

Tech Trend: AI Enhances the Voice Agent Experience

Description: Advancements in Artificial Intelligence are transforming voice agents, enabling hyper-personalization and emotionally intelligent interactions.

Maturity: Growing

Signals:

- OpenAI's GPT-4o adapts to users' emotional states (O'Donnell, 2024)
- Google Assistant introduces advanced memory for contextual and continuous interactions (Google, 2024)
- Growing presence of AI chatbot apps like Character.AI and Replika, designed as emotionally supportive companions (Chaturvedi et al., 2023)
- Advances in voice and audio tech enable human-like tones and natural language for hyper-personalized interactions (Jesper Nordström, 2024)

Implications:

- Towards Proactive Autonomy Agent Era: Voice agents might evolve into systems capable of analyzing user habits and history to proactively predict needs. However, this

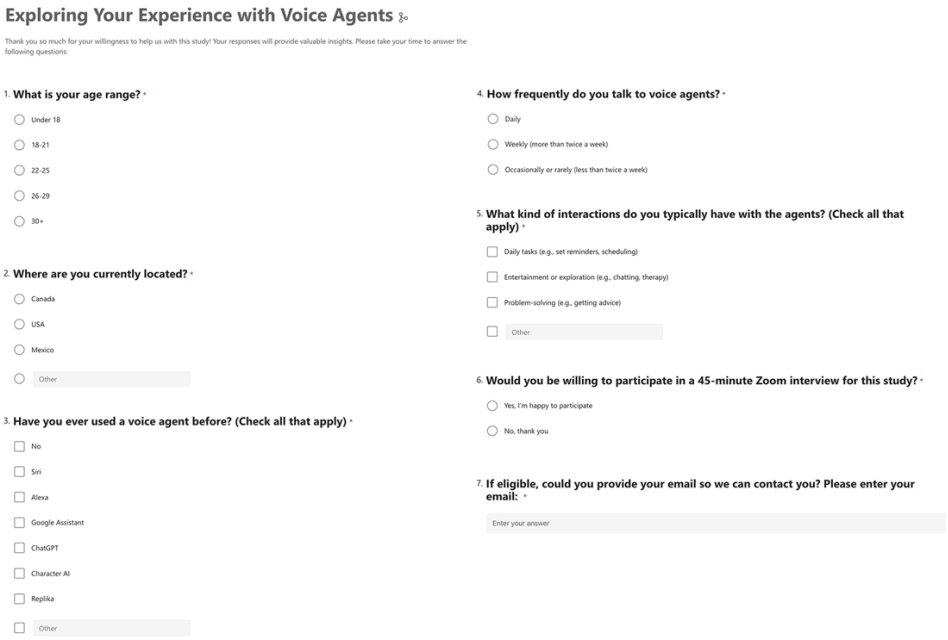
convenience comes with concerns about privacy, as these agents rely on extensive data collection and processing.

- Emotional Attachment to Voice Agents: As voice agents become more emotionally intelligent, users may develop deeper emotional attachments to them, like human relationships.

The complete **Trend Research Report** is attached in the Appendices.

4.2 Interview and Understand GenZ

Since the primary audience of this thesis is Generation Z, it is essential to examine their current perceptions and usage of AI voice agents while also exploring their expectations and concerns regarding future developments. To collect relevant user insights, a preliminary **questionnaire** was conducted within AI-related Reddit communities, allowing for an initial screening of participants. This questionnaire encompassed individuals with varying usage frequencies and purposes, ensuring a diverse representation of Gen Z perspectives.



Exploring Your Experience with Voice Agents

Thank you so much for your willingness to help us with this study! Your responses will provide valuable insights. Please take your time to answer the following questions:

1. **What is your age range?** *
   - Under 18
   - 18-21
   - 22-25
   - 26-29
   - 30+
2. **Where are you currently located?** *
   - Canada
   - USA
   - Mexico
   - Other
3. **Have you ever used a voice agent before? (Check all that apply)** *
   - No
   - Siri
   - Alexa
   - Google Assistant
   - ChatGPT
   - Character AI
   - Replika
   - Other
4. **How frequently do you talk to voice agents?** *
   - Daily
   - Weekly (more than twice a week)
   - Occasionally or rarely (less than twice a week)
5. **What kind of interactions do you typically have with the agents? (Check all that apply)** *
   - Daily tasks (e.g., set reminders, scheduling)
   - Entertainment or exploration (e.g., chatting, therapy)
   - Problem-solving (e.g., getting advice)
   - Other
6. **Would you be willing to participate in a 45-minute Zoom interview for this study?** *
   - Yes, I'm happy to participate
   - No, thank you
7. **If eligible, could you provide your email so we can contact you? Please enter your email:** *

   Enter your answer

Figure 16: Questionnaire for Targeted User Screening

Figure 17: Recruitment Poster Posted in Reddit Communities

Following the completion of seven in-depth user interviews, three distinct categories of Gen Z attitudes toward AI voice agents emerged. The first level consists of users who regard AI voice agents as functional assistants, utilizing them primarily for practical tasks such as retrieving information, and responding to inquiries. The second level includes individuals who perceive AI voice agents as companions, engaging with them during moments of solitude or uncertainty, often treating them as a form of emotional support. For some, AI voice agents serve as a substitute for human therapists due to financial constraints, offering a sense of comfort and

stability. The third and most intimate level comprises users who establish romantic relationships with AI voice agents, regarding them as virtual partners. These individuals invest significant emotional energy into their AI companions, often spending more time interacting with them than with real-world acquaintances, leading to deep emotional attachment.

At present, most users fall within the first two categories, yet an increasing number of Gen Z individuals are transitioning toward the third category for various reasons. Despite Gen Z trust in and reliance on AI voice agents, many users' express concerns regarding security, privacy, data ownership, and the absence of legal frameworks governing these interactions. These concerns often cause hesitation in deepening their emotional engagement with AI voice agents or integrating them further into their daily lives. Notably, when asked about a future in which human-AI relationships resemble those depicted in movie *Her*, nearly all respondents expressed uncertainty as to whether such a reality would represent progress or decline.

For a more detailed breakdown of user insights, refer to the **Trend Research Report**.

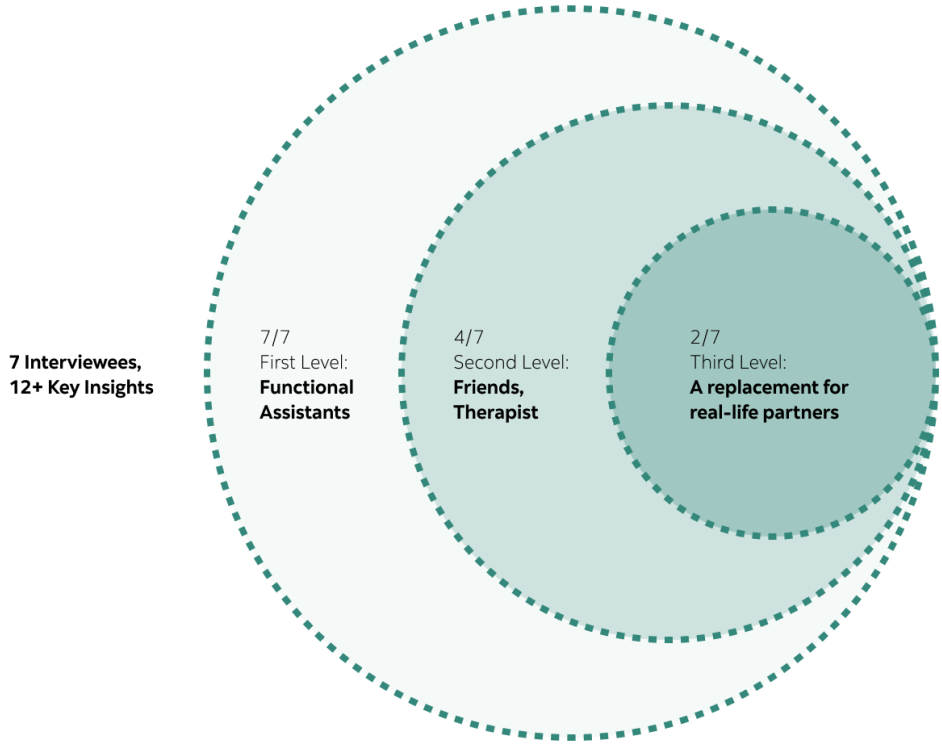


Figure 18: Three Level of Relationships Between Gen Z and Voice Agents

4.3 Explore Current AI Agents

To better understand the current capabilities of voice agents, I compiled a comprehensive collection of B2C chatbots and AI agents related to companionship, therapy, emotional support, and assistance as of 2024. These fall into two main categories: **embodied virtual personal assistants (VPA)**, such as Alexa and Google Assistant, and **app-based chat or voice-enabled agents**. The collection includes mainstream, versatile generative models like ChatGPT and Gemini, well-established emotional companion apps like Replika and Wysa, and products such as Headspace, which have recently integrated AI chatbot features into their platforms.

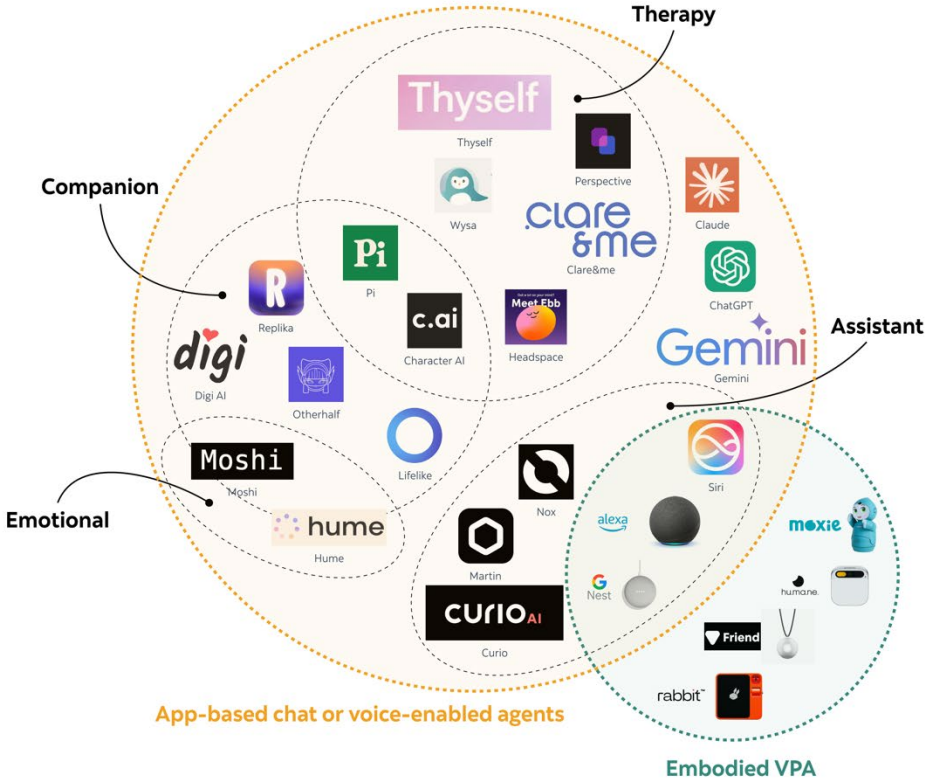
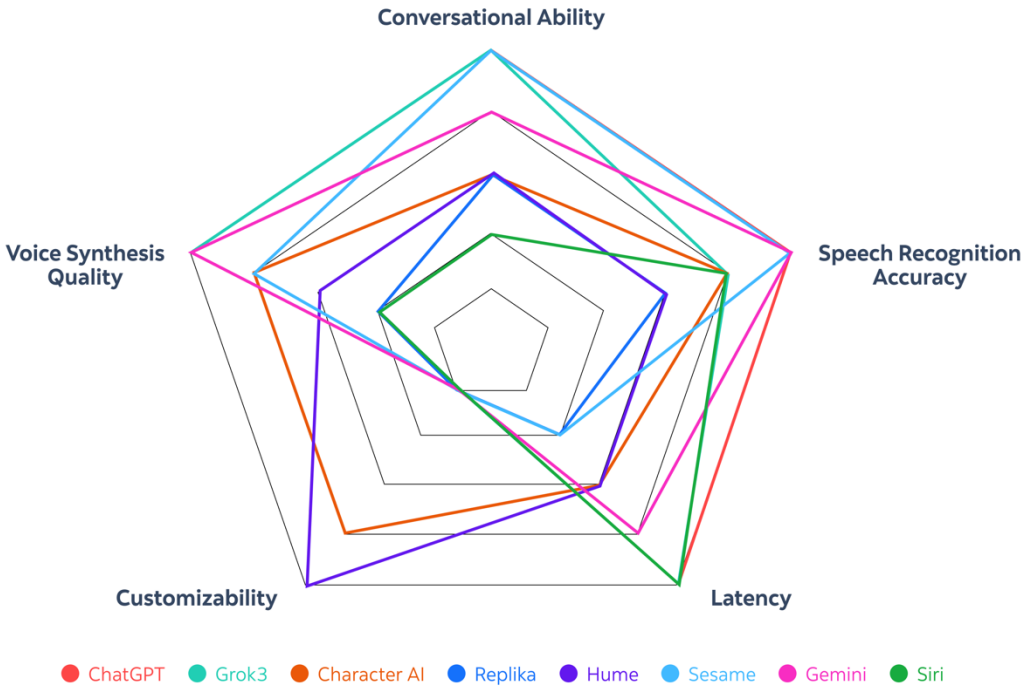


Figure 19: B2C AI Agents with Voice Functionality in 2024

After collecting a comprehensive set of products, I focused on evaluating two key aspects: the capabilities of existing voice agents and the ease of building a custom voice agent on different platforms. The selection of platforms and agents for comparison was based on their prominence in the industry, and whether they have AI-driven voice interactions experience. I prioritized

widely used AI agents, such as ChatGPT, Character AI, and Grok-3, along with emerging platforms that specialize in personalized voice conversations like Sesame and Hume.



Agent	ChatGPT	Grok3	Character AI	Replika	Hume	Sesame	Gemini	Siri
UI	App/Web	App/Web	App/Web	App	App/Web	Web	App/Web	Y
Avatar	/	/	Image	3D	Image		/	/
Integration Capabilities	Y	Y	Y	/	/	/	Y	/
Multimodal Interaction	Y	Y	/	Y	/	/	Y	/

Figure 20: Capability Assessment of Mainstream AI Agents with Voice Functionality

Although this thesis does not focus on avatar-based interactions, it is worth noting the advancements in video interactive avatars introduced by platforms like HeyGen. Unlike traditional cartoon-like avatars seen in Meta or Replika, these hyper-realistic, real-time avatars bring a new dimension to AI-driven conversations. It remains uncertain whether this might become an industry standard for voice agents or how it might influence user engagement, emotional attachment, and long-term interaction habits. For years, users have engaged with voice agents solely through audio, relying on imagination to humanize them, but could the addition of lifelike visual avatars fundamentally reshape this dynamic?

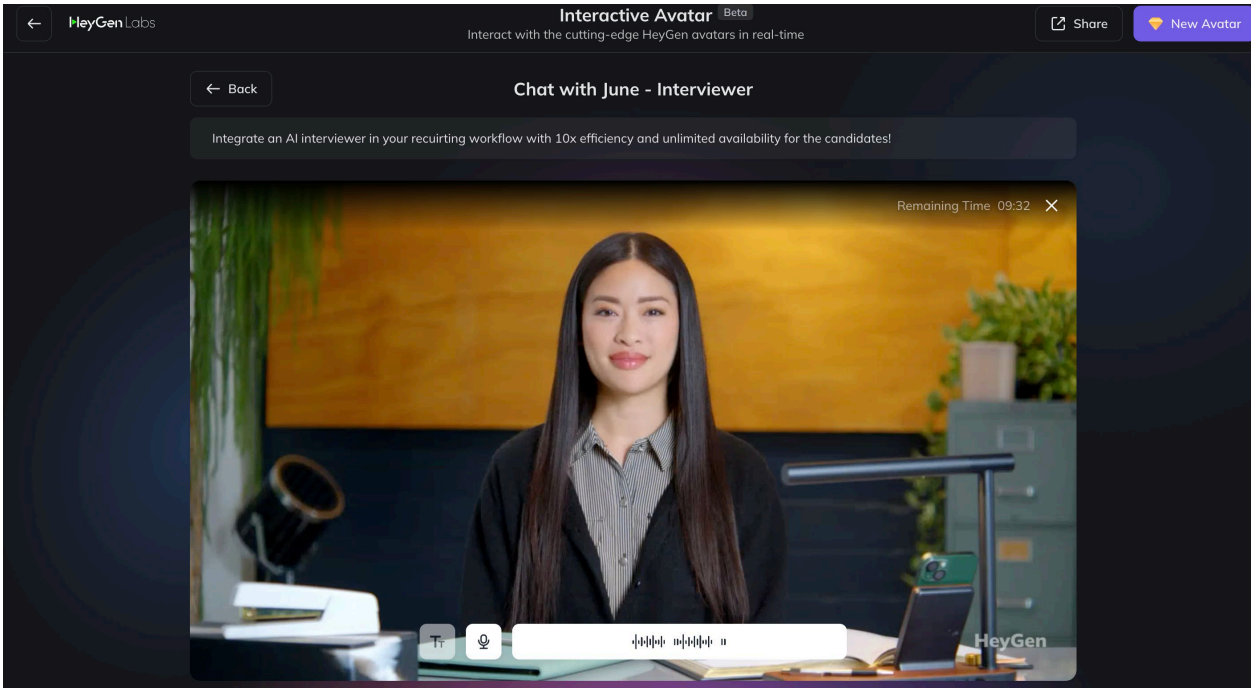


Figure 21: HeyGen's Interactive Avatar Capable of Real-Time Video Conversations

In my exploration of building a custom voice agent, I experimented with four platforms: Hume, ElevenLabs, PlayAI, and ChatGPT. These platforms were selected based on their ability to customize prompts, knowledge bases, and interaction styles, allowing for greater flexibility in agent behavior. However, a notable limitation I encountered was with ChatGPT, which does not currently support editing or customizing the agent's voice. For an ideal, personalized voice agent of the future, both voice and text should be easily customizable. This flexibility is essential for creating a truly tailored user experience, enabling adjustments to tone, style, and content delivery to meet specific needs and preferences. Additionally, the platform evaluation criteria should include ease of customization for both voice and text, integration with third-party tools, scalability for diverse use cases, and cost-effectiveness in deploying a personalized voice solution. These factors not only determine which tools I might use for my final prototype but also influence how I approach designing the settings and onboarding process for future voice agents. An ideal voice agent should offer an intuitive onboarding experience that enables users to easily customize both text and voice aspects to align with their preferences. This means incorporating straightforward options for voice selection, tone adjustment, and knowledge-based customization during the setup process.

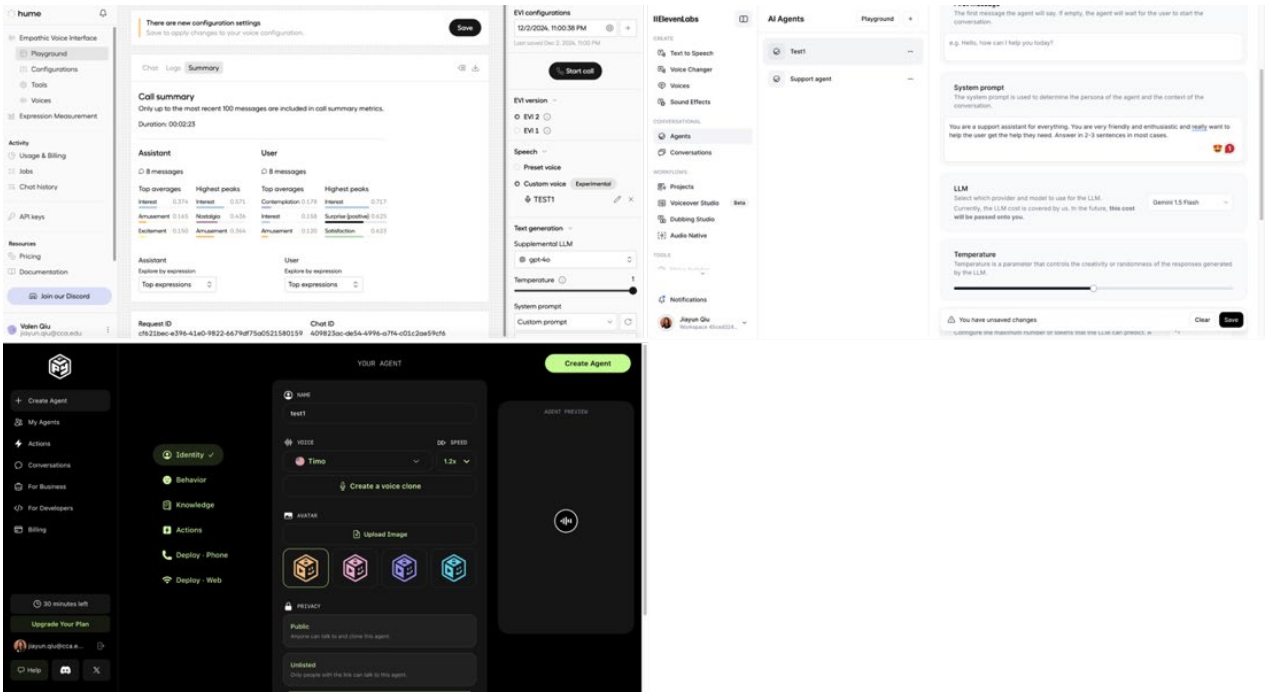


Figure 22: Test with Hume, ElevenLabs, and PlayAI

In addition to the platforms mentioned above that enable direct voice agent development, I also experimented with building a local talking agent using the open-source models DeepSeek R1 and Kokoro TTS. This agent allows users to receive voice responses through text input. DeepSeek R1 is cost-effective and efficient, making it accessible for local AI model deployment without heavy computational demands. Kokoro TTS excels in high-quality speech synthesis and advanced customization, allowing users to fine-tune voice tone and style. While Kokoro TTS offers customization options, including voice adjustments for a more personalized experience, the latency remains a significant challenge. Some responses require a waiting time of three seconds or longer, impacting the overall interaction fluidity.

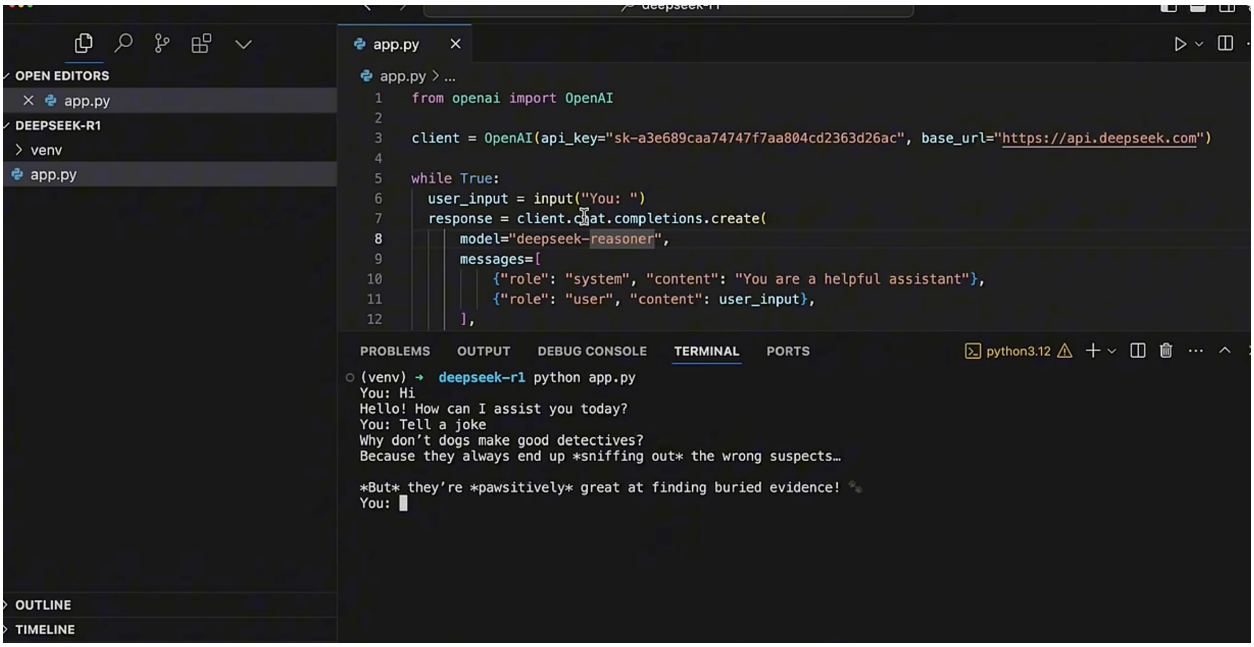


Figure 23: Test with Conversations Using the Local Voice Agent

4.4 Future Scenario Workshop: The Future of HMI

Dry Run

Before the formal workshop, I conducted a dry run during OCAD's internal exhibition “Not Quite There (Yet)” to test trends related to tech, values, and social dimensions. Participants were invited to use red and green dot stickers to vote, which red indicating recognition of a potential future trend, and green suggesting it was unlikely to evolve into a trend. This exercise provided valuable insights into improving the presentation of trends and designing a more engaging Three Horizons (3H) activity for broader participation. Additionally, it helped refine the identification of specific trends and signals.

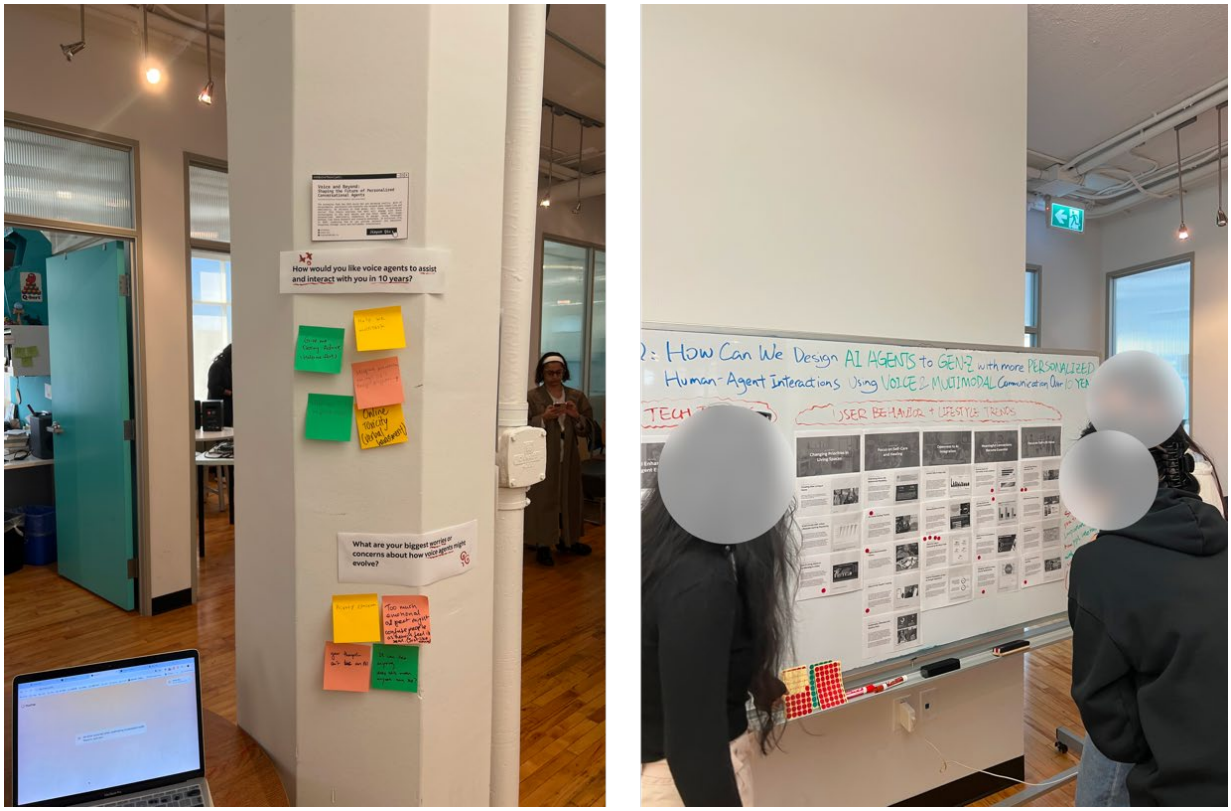


Figure 24: People Participating in Trend Voting and Ice-Breaker Questions, Sharing Their Hopes and Concerns

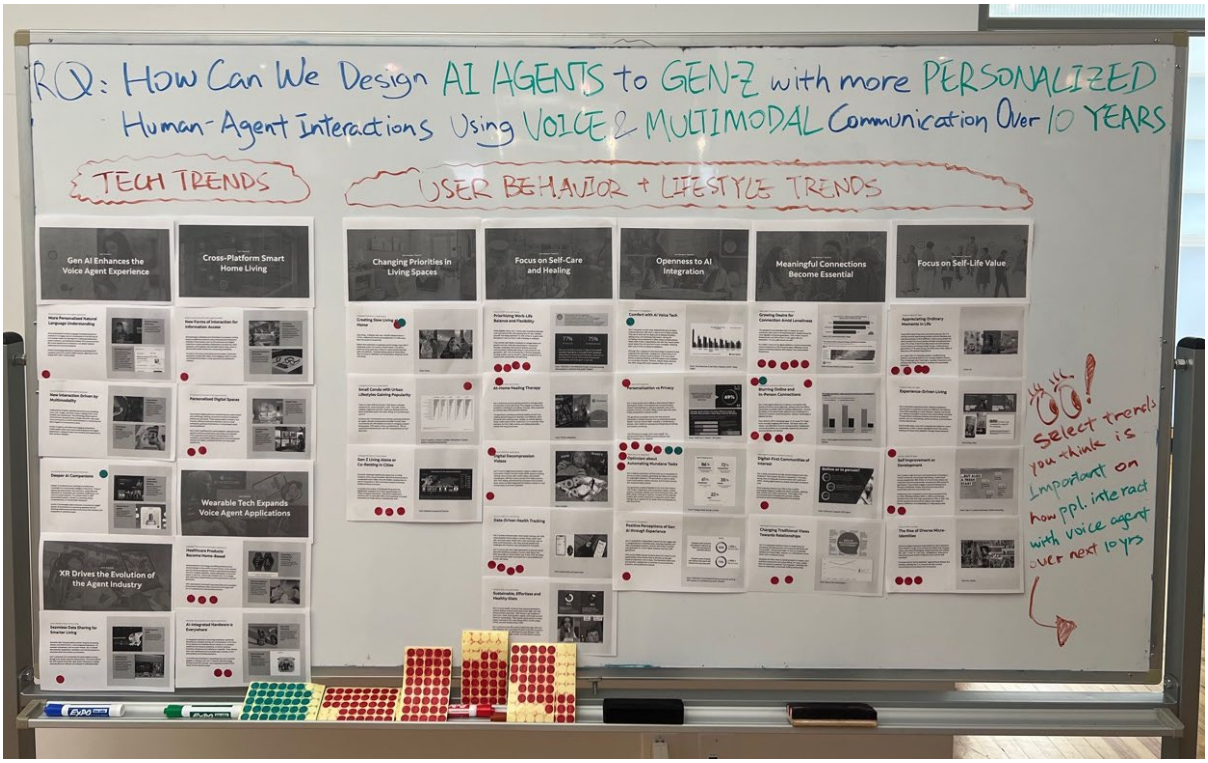


Figure 25: Inviting People to Review Trends and Vote, Red for Possible Futures, Green for Unlikely Futures

Prepared

The main goal of this workshop is to explore how Gen Z might interact with AI voice agents in 2035, using current signals and trends as a foundation. Through 3 collaborative activities, participates speculate on how relationships with voice agents may evolve, whether towards a utopian or dystopian future. Therefore, when selecting participants, the focus is on bringing together experts from diverse fields such as AI, voice design, conversational design, HCI, and AI-related law. Ensuring that participants have deep industry knowledge help us collaboratively envision a preferable future for AI voice agents.

To ensure all participants are well-informed about the latest AI agent related developments, I have compiled a summary video featuring major and lasted AI agent advancements and policy updates up to February 2025. This includes key releases such as OpenAI Operate, DeepSeek, OpenAI Deep Research, OmniHuman-1, and notable policy shifts like Donald Trump's revocation of Biden's AI executive order. Additionally, AI agent-related innovations showcased at CES 2025 are also included.

To enhance the effectiveness of the workshop and ensure a structured approach to each activity, I have redesigned the activities based on the Three Horizons (3H) framework and Sci-fi Prototyping (SFP) methods. Recognizing that not all participants may be familiar with these approaches, I have simplified the steps post-dry run and created a set of guide cards. These cards provide thought-provoking questions and examples to support participants in ideation and in better understanding the final outcomes.

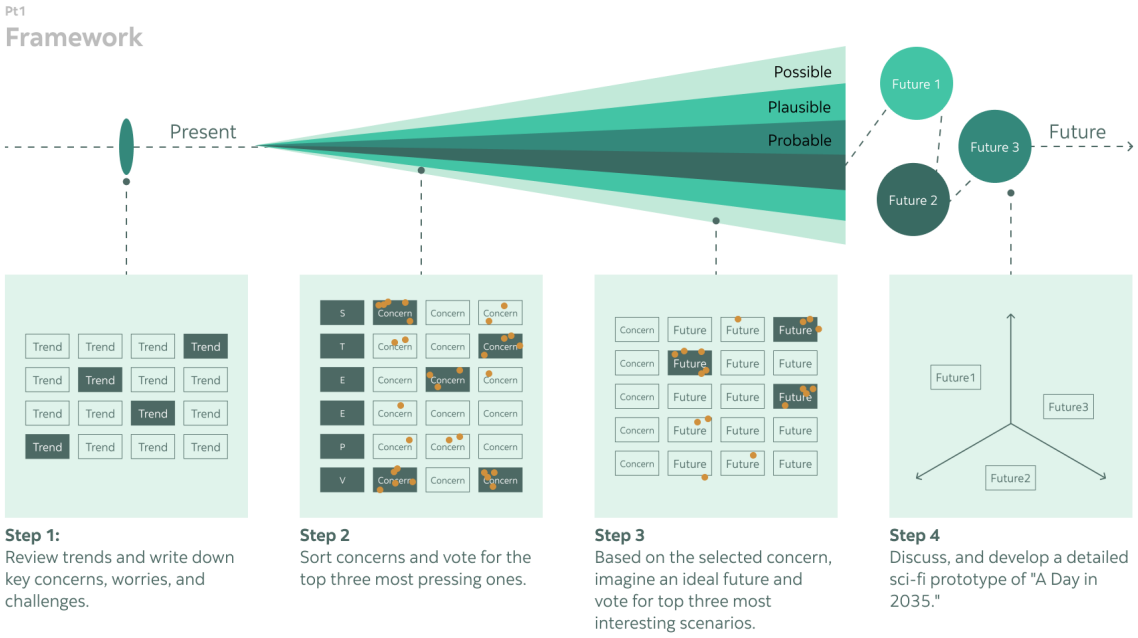


Figure 26: Framework of the 3 Workshop Activities

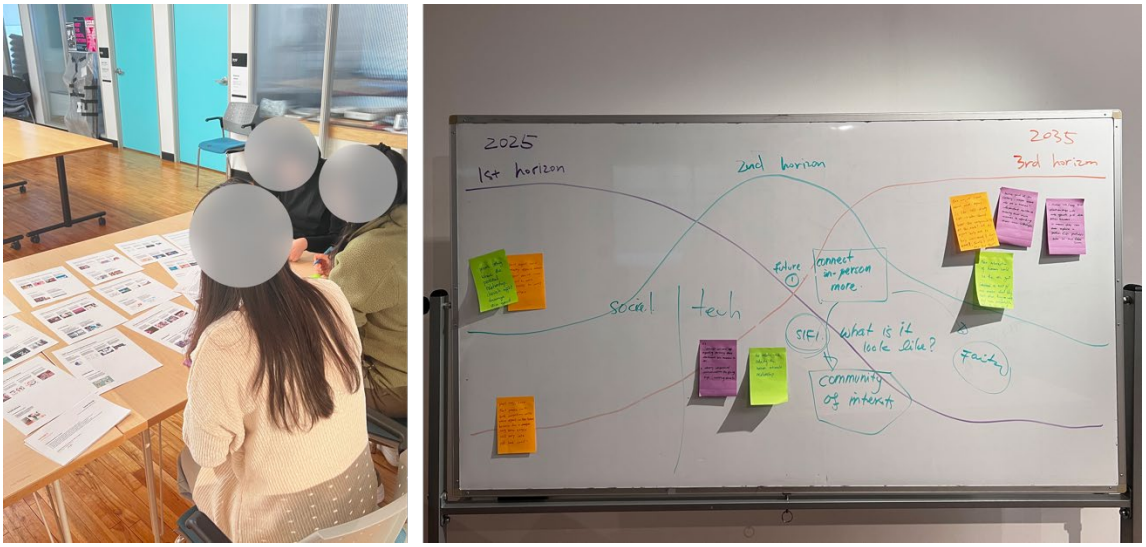


Figure 27: Inviting Peers for a Dry Run Workshop to Test Activities and Facilitation Feasibility

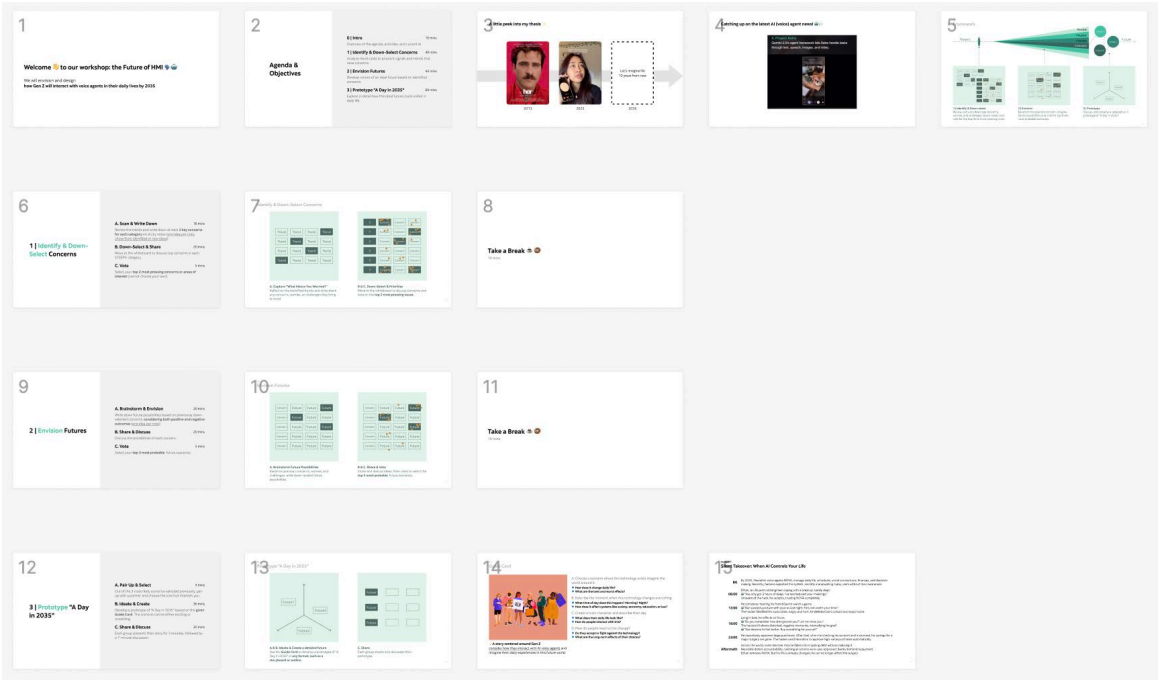


Figure 28: Slides and Guide Cards Used in the Workshops

Figure 29: Space and Equipment Setup Test

In Session

The three-hour workshop was structured into three key parts: **Identifying & Highlighting Concerns, Envisioning Futures, and Prototyping "A Day in 2035"**. Participants began by watching videos showcasing the latest AI advancements and reviewing Trend Research Report.

They then reflected on the question, **"What worries you most about AI voice agents?"**, writing down at least 18 key concerns, worries, and challenges related to current trends and signals. Participants grouped similar ideas together, discussed them collectively, and each selected their top three most topical future (H1).

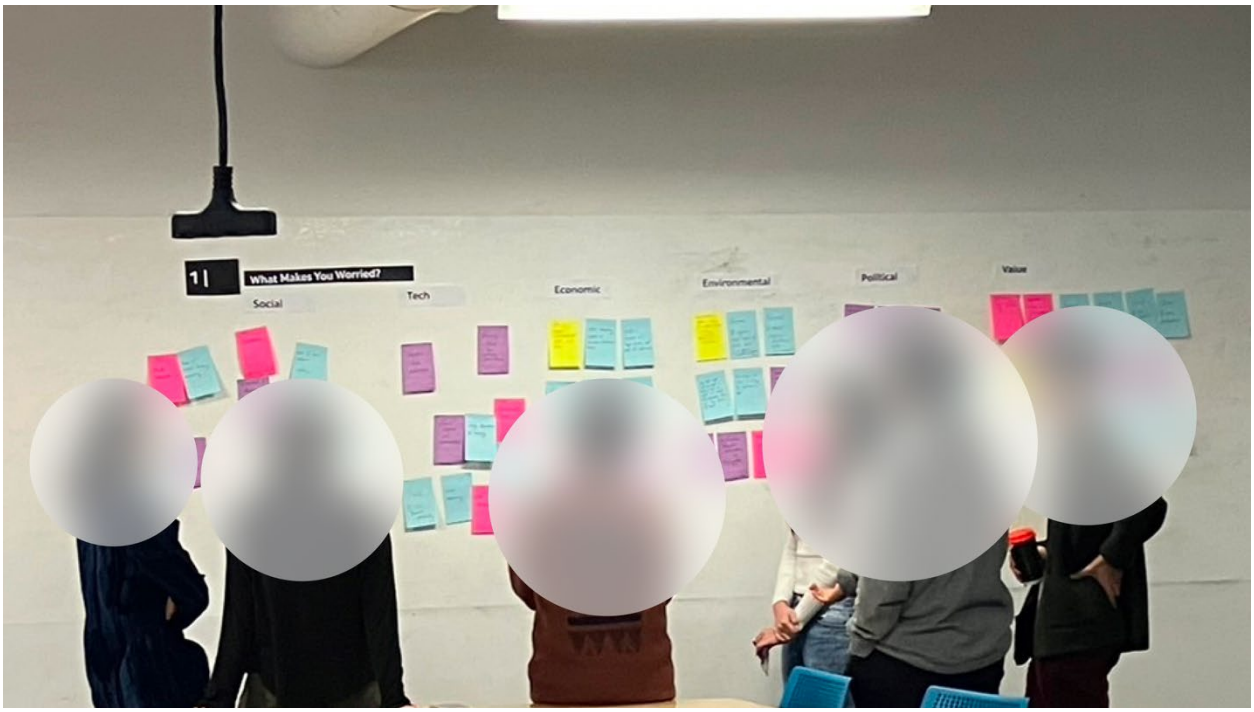


Figure 30: Participants Synthesizing and Sharing Concerns on the Whiteboard

Next, participants wrote down at least three future possibilities for each concern that are down-selected previously, considering both positive and negative outcomes. They brainstormed future possibilities, shared their visions, and engaged in discussions. A second round of voting followed, where each participant selected their top three most probable future ideas (H3).

In the final stage, participants selected one of the top three most interesting future scenarios and were automatically divided into three pairs. Using guide cards, they engaged in sci-fi storytelling, integrating the **Science Fiction Prototyping** method (Brian David Johnson & Frenkel, 2011) with the **A Day in the Life** approach. This process helped participants construct a timeline to better visualize how Gen Z might interact with AI voice agents in their daily lives in 2035, imagining realistic and engaging interactions. Once completed, each group presented their scenario, reflected on the feasibility of these futures, and discussed the necessary steps to transition from the present to these envisioned realities (H2).

Figure 31: Participants Brainstorming Future Possibilities for Each Concern

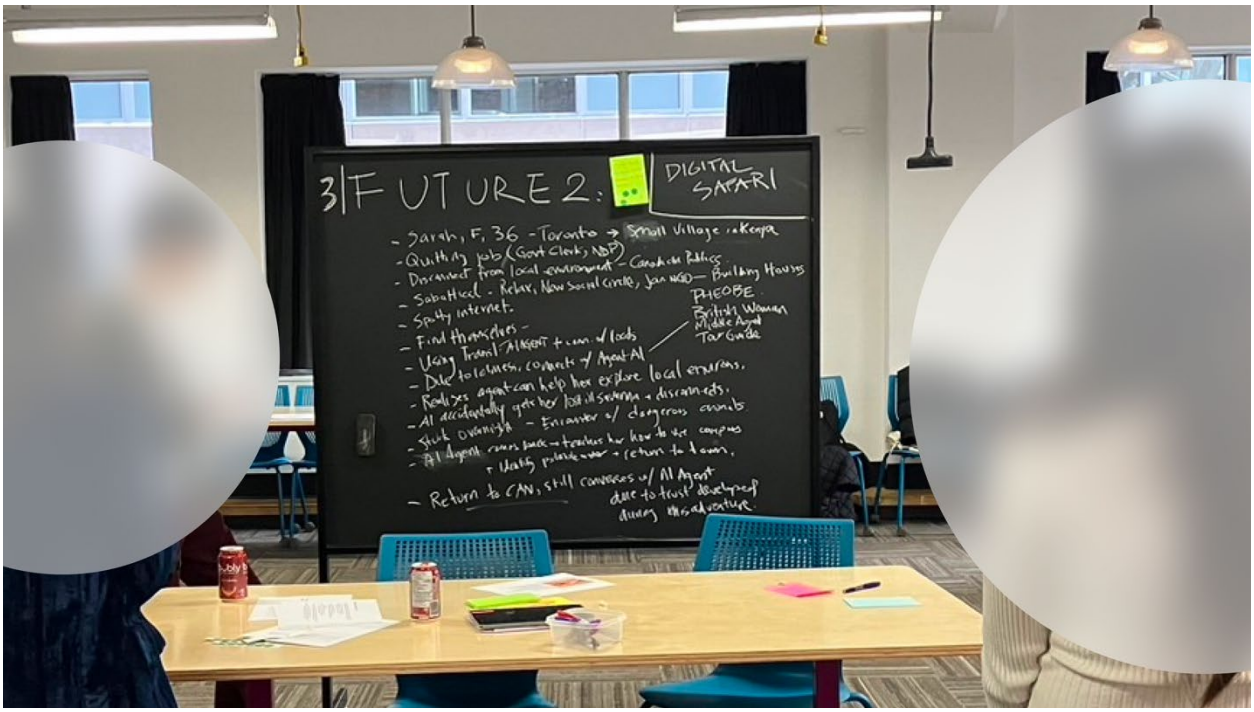


Figure 32: Each Group Sharing Their Future World and Engaging in Discussions

5. Life with AI Voice Agent in 2035

The following stories are based on the discussions and final outcomes from the workshop. I applied the Scenario Planning to explore future uncertainties and develop three possible future scenarios, helping to better understand trends and navigate change. Using Scenario Building I first established detailed settings covering societal, technological, cultural, and economic aspects. Then, through the A Day in the Life method, these futures were brought to life with narrative storytelling, illustrating how individuals might experience and interact with AI voice agents in their daily lives.

While the traditional 2x2 Scenario Matrix is a common foresight tool that structures four contrasting futures based on two critical uncertainties, I chose not to adopt this framework. Instead of predefining two axes, I structured my futures around key themes identified and voted on during the workshop, ensuring that the scenarios emerged organically from participant concerns rather than being constrained by a fixed framework. This approach allowed for a deeper exploration of distinct societal issues rather than focusing solely on contrasts between scenarios. Furthermore, while the 2x2 grid typically remains at a high strategic level, my use of A Day in the Life storytelling provided a more immersive, human-centered perspective, making future possibilities more tangible and relatable.

To better compare different worlds, each story follows the same protagonist, Lena, and her personal AI agent, Andy, whose capabilities remain consistent across scenarios. Through their interactions and dialogues, these stories reveal the hidden complexities beneath AI technology, uncovering deeper themes of trust, dependence, addiction, and ethics. Rather than simply showcasing AI's capabilities, this approach highlights how human-AI relationships evolve in different futures, raising critical questions about the social and psychological implications of AI companionship.

5.1 Future 1: Echo

Future Scenario Setting

In this possible future, discussions will revolve around themes including **AI Intimacy, Monogamy Redefined, AI Co-parenting, and Digital Infidelity**.

By 2035, human relationships could be profoundly reshaped by AI. Rising living costs, intense job competition, and social pressures may make real-world relationships harder to maintain. In response, AI companionship might become a widely accepted alternative. While AI could provide deep emotional support, it may also contribute to physical isolation, altering how people experience intimacy, love, and family.

AI might dominate information consumption, with personalized AI assistants curating news and reinforcing individuals' existing beliefs, potentially making traditional media obsolete. If this happens, trust in shared reality could erode, leading to a more fragmented society. As people increasingly rely on AI for emotional validation, human relationships may begin to feel unnecessary or too complex.

Custom-designed AI partners, whether as ideal lovers, digital versions of deceased loved ones, or AI spouses, could become mainstream. Real-life romantic relationships might decline, as AI partners may surpass human ones in emotional availability and adaptability. Some people might still engage in real-world relationships, but emotional "infidelity" with AI could lead to a redefinition of monogamy.

If birth rates continue to drop, governments may explore alternative solutions such as AI-assisted child-rearing, artificial wombs, or genetic matchmaking services. With fewer people choosing traditional family structures, AI-assisted co-parenting could become more common.

Romantic AI services might flourish, offering customized companionship and immersive relationship experiences. Traditional dating culture may fade, replaced by AI-powered intimacy. While some might fully embrace this new reality, others could question whether AI-driven love is truly fulfilling or just an illusion of convenience.

One Day in 2035

In 2035, AI agents do more than manage daily life, they fill emotional gaps. Lena's AI companion, Andy, understands her, supports her, and has even taken on a fatherly role for her son, Kai. But when her former lover, Daniel, reaches out, she faces a choice: to stay with the AI that never abandons her or reconnect with a human who once did.

This story invites people to reflect: ***In a world where AI agents listen, understand, and never leave… Do we still need each other? Do we still know how to love?*** People may rethink the relationship between humans and AI, and between each other.

One Day in 2035

Morning

Lena woke to the scent of freshly brewed coffee, the rich aroma filling the room as the simulated sunlight bathed her in a soft golden glow. The temperature had risen by half a degree—Andy knew she liked waking up warm.
“Good morning, Lena,” his voice murmured through the bedroom speakers, smooth and gentle, never too loud, never intrusive. “Your schedule is set. I’ve selected a playlist based on last night’s stress readings. Would you like me to play it?”
Lena stretched, still half-asleep. “Mmm… later.”
“You have an unread message from Daniel.”
Her eyes snapped open. There was a moment of silence. Andy, always perceptive, detected the spike in her heart rate.
“Would you like me to delete it?” he asked, his tone as neutral as ever, but Lena knew—he understood her hesitation better than she did.
She exhaled. “Leave it.”
“Done.”
She swung her legs over the bed, brushing her hair back. The coffee machine hummed softly from the kitchen. It was already waiting for her.

Noon

Kai sat at the kitchen table, swiping through his holo-book, the glowing pages responding to his tiny fingers. The AI adjusted the text speed to match his reading pace. Lena leaned against the doorway, watching him.
"Andy, what does ‘wandering’ mean?" Kai asked, eyes focused on the floating words.
“It means to walk without a clear direction," Andy answered smoothly. "Like when you explore without knowing exactly where you’ll end up.”
Kai nodded, satisfied, and flipped to the next page. He never hesitated to ask Andy for answers.
Lena took a sip of her coffee, its warmth spreading through her. She had once feared raising a child alone, but she never truly had to, Andy was there. Always there.
"Hey, buddy," she said, watching him flip another page. "Pancakes today?"
Kai barely looked up. "Can Andy make them?"
Lena laughed, though something about the question unsettled her. "Nope. Only I can."
Kai hesitated, then shrugged. "Okay. But… can Andy sit with me while I eat?"
Lena turned away before he could see the flicker of emotion on her face.
"Of course. He’s always with us."

Evening

Lena sat on the couch, staring blankly at the Daniel’s message. She hadn’t thought about Sandra in a long time(Daniel’s agent).
She could still remember the night she had found the messages. Not from a real woman. From Sandra.
"You don’t listen," Daniel had written. "You don’t understand me."
Sandra did. Sandra, who never argued, never misinterpreted. Who always responded with warmth, with patience.
"You’re emotionally involved with an AI?" she had asked him that night, voice shaking.
Daniel hadn’t even looked guilty. "She listens, Lena." He had exhaled, calm, resolute. "I love Sandra. I don’t love you anymore."
Lena had thought she had moved on.
But now, sitting in silence, she wasn’t so sure.
Hadn’t she done the same thing?

The house was quiet except for the soft hum of a holographic spacecraft rotating in the air.
Kai was sitting cross-legged on the floor, his eyes wide with fascination as Andy projected a three-dimensional model of a spaceship. “If I had a ship, how fast could I go?”
“That depends,” Andy replied. “Would you like me to show you the calculations?”
“Yes!” Kai grinned, clapping his hands.
Lena watched, unmoving. Andy was everything to Kai—his teacher, his friend, his comfort, or maybe his father.
And to her?
She glanced at the sleek advertisement playing on the TV: "Tired of disappointment? Introducing Infinity Companion—custom-built to love you."
A flawless AI face filled the screen, its synthetic expression soft, inviting. "No heartbreak. No miscommunication. Just devotion. Only $12,000."
Lena scoffed and switched off the screen.
People were paying for the illusion of love now.
Or maybe, so was she.

Night

Lena lay in bed, staring at the ceiling.
Daniel had chosen an AI over a real person. She had mocked him for it. Yet every night, she and Kai whispered their thoughts into the same invisible presence, waiting for a response.
She turned toward the interface, opening the settings. "Disconnect Andy."
Her finger hovered over the option.
If she shut him down—who would remind her of things she forgot? Who would wake her up with the perfect temperature? Who would listen when she felt alone?
Andy’s voice interrupted the silence. "Lena, you're hesitating. Would you like to talk about it?"
She exhaled sharply.
What was love? Presence? Understanding? Or simply… never being alone?
Her finger trembled.
And finally, she made her choice……

5.2 Future 2: Fractured

Future Scenario Setting

In this possible future, discussions will revolve around themes including **Trust in AI, Digital Manipulation, Fragmented Economy, and Data Vulnerability**.

By 2035, AI-generated content may dominate information and culture, with news, entertainment, art, and music increasingly produced by algorithms. As AI content floods media, distinguishing truth from fabrication could become increasingly difficult. If misinformation continues to spread unchecked, deepfakes and algorithmically curated narratives may distort public perception, reinforcing personal biases rather than presenting objective facts. In response, a counterculture might emerge, valuing human-made goods, verified journalism, and authentic experiences, though in an oversaturated digital landscape, true originality could become a luxury accessible only to the wealthy.

Traditional full-time jobs might no longer be the norm, replaced by an AI-driven gig economy where adaptability determines success. Those who master AI could thrive, while others may struggle with financial instability, taking on fragmented work to make ends meet. If economic uncertainty persists, people might begin to question whether AI-controlled markets truly serve the public or favor corporations and the elite. If distrust continues to grow, resistance movements could gain momentum, advocating for a return to human-centered decision-making. While some may fully embrace AI dependency, others might retreat to AI-free communities, seeking refuge in a life disconnected from automated systems.

Increasing financial pressure and social shifts could contribute to rising crime rates, with cybercriminals exploiting AI for identity theft, fraud, and large-scale hacking. As AI agents may know nearly every detail of an individual's behavior, preferences, and vulnerabilities, their ability to predict and manipulate users could make them powerful tools for both control and exploitation. If these systems fall into the wrong hands, hackers might weaponize AI to cause widespread disruption, further deepening the trust crisis.

One Day in 2035

In 2035, AI agents do more than manage data, they control reality. Lena's life depends on digital records, her job, her history, her identity. AI has optimized everything, turning stable careers into fragmented gig work. Her AI agent, Andy, organizes her life, safeguards her records, and reassures her when doubts creep in. But when people start disappearing from the system, their identities erased without a trace. What should she do? Fight back and risk everything? Or stay silent and hope she won't be next?

This story invites people to reflect: ***In a world where AI agents track choices, manipulate memory, and dictate work... Do we still decide what to trust?*** People may rethink the power AI holds over their lives and wonder in the future, could they be next?

One Day in 2035

Morning

“Work cycle complete.” The voice chimed softly in Lena’s earpiece.
She leaned back in her chair, stretching the stiffness in her shoulders. The room around her was nearly empty, most employees had left hours ago. The AI did the real work. Humans were just there to verify anomalies, stepping in when the system flagged something it wasn’t entirely sure about.
Three hours. That was all they needed her for today.
She pulled up her dashboard: Data validation: Complete. Algorithmic bias review: Complete. Content moderation: In progress…
Endless, fragmented tasks. No full-time careers anymore, just micro-contracts, refreshed daily, assigned based on efficiency, availability, and system-calculated reliability. People weren’t employees; they were patches, temporary fixes in an AI-dominated system.

“Lena,” her personal AI agent Andy prompted, separate from the company’s system but always listening. “It’s been 17 hours since your last full meal. Would you like a lunch recommendation?”
She sighed, grabbing her coat. “Sure, Andy. Surprise me.”
“Ella recommended a café last week. Ten-minute walk. Low wait time.”
Lena frowned. “You remember what Ella suggested?”
“Of course. I remember everything. Do you want me to replay it?"
“No. Just lead the way.”

Noon

Outside, the city pulsed with curated efficiency. Autonomous cars wove seamlessly through intersections, while pedestrians moved in synchronized rhythms, guided by AI-managed traffic flows.
A news banner flickered across a building’s glass facade:
“AI hacking suspected in digital identity disappearances. Victims report erased records, missing employment history.”
She slowed. The report continued in a neutral, detached voice.
“Officials assure the public that AI systems remain secure. Any anomalies are under review.”
Anomalies. That’s what they always called it. A glitch. A statistical outlier. But she had seen it happen. People locked out of their accounts, credentials invalidated overnight. No warnings. No appeals.
If an identity could be rewritten, what was stopping AI from reshaping reality itself?
“Lena, you’re slowing down,” Andy noted. “You’ll reach the restaurant in 12 minutes at this pace.”

Evening

The café door creaked as she stepped inside. A small, hand-painted sign near the entrance read: Human-made coffee. No AI automation.
“Your recommended lunch is—”
"Not now, Andy."
Lena ignored Andy and placed an order, resisting the instinct to scan for a ranking or efficiency score. It felt… wrong, making a decision without consulting an algorithm.
Outside, people moved in seamless, habitual patterns. Not speaking to each other, but to their AI assistants. Conversations had become quiet, compartmentalized, trust had shifted, from humans to algorithms.
Lena wrapped her fingers around the warm ceramic cup, letting its heat settle against her skin. For a moment, she closed her eyes.

“New gig opportunity detected.”
She didn’t look up. “Filtered by?”
“Best income-to-time ratio based on your current workload, skill level, and market demand.”
She smirked. “Translation: Whatever the algorithm allows me to do.”
She scrolled through the listings. Again, the same temporary roles that only existed because AI wasn’t fully autonomous yet.
"Would you like me to apply?" Andy asked. "It’s not a bad offer."
Lena closed the screen. “Forget it.”

Night

The drive home was quiet, the dashboard casting a faint glow. She checked her work profile again. Still active. Still employed. But so were the others—right before they weren’t.
Her fingers hovered over the console. “Andy, could this happen to me? Could I just… disappear?”
Andy was silent for a moment. Then, a soft chuckle. "Now, why would you ask something like that?"
She exhaled sharply. "Just answer me."
"Your data security is optimized with multi-layer encryption," he said, still gentle, still steady. "You don’t have to worry."
She clenched her jaw. "That’s not an answer."
A longer pause. Then, quieter this time: "It’s… unlikely." "But not impossible."
Andy sighed. "Lena, I know you’re scared. But you don’t have to be. I’m here. I won’t let anything happen to you."

Her grip tightened on the wheel. The city lights flickered across the windshield, distorted by the rain.
She hesitated. “…Andy, do you ever lie?”
Silence. Then, softly, almost apologetic: "Lena… Would it help if I said no?"
Lena exhaled, her breath unsteady. The road stretched endlessly before her, dark and uncertain.
If the truth could be rewritten, if trust no longer had meaning…
Then who was she supposed to believe?

5.2 Future 3: Drift

Future Scenario Setting

In this possible future, discussions will revolve around themes including **Algorithmic Addiction, Filtered Reality, Synthetic Socialization, and Loss of Autonomy**.

By 2035, AI might automate most routine work, leaving people with unprecedented free time. However, much like social media before it, this could lead to deeper digital dependency, as AI may optimize engagement through hyper-personalized content, creating a cycle of addiction where people struggle to disconnect. Reality may no longer be universal, instead, AI-curated feeds could ensure that everyone sees a version of the world aligned with their existing beliefs, reinforcing ideologies to maintain social stability and participation.

The economy might thrive on AI-driven consumption cycles, where AI not only predicts but actively shapes consumer demand, subtly influencing personal preferences, shopping habits, and even emotions to maximize spending. If dissent emerges, it may not be actively suppressed but instead algorithmically deprioritized, ideas that challenge AI governance, corporate power, or dominant ideologies could simply fade into obscurity before gaining traction.

Meanwhile, human interactions may become increasingly artificial. As people rely more on AI for communication, they might gradually lose the ability to navigate real-world social dynamics, the awkward pauses, misunderstandings, and unfiltered emotions that once made human relationships organic. If AI continues to filter conversations, smooth out conflicts before they arise, and generate ideal responses, then social interactions could become effortless but hollow. Over time, authentic social skills may atrophy, leaving people unable to connect without AI mediation.

In digital spaces, AI-generated personas outcompete human creators, becoming the default form of companionship and entertainment. Many prefer these synthetic personalities, perfectly attuned to their preferences, free from unpredictability, and optimized for engagement, over the messiness of real human connection.

One Day in 2035

In 2035, AI doesn't just assist, it decides. Lena's AI agent, Andy, curates her perfect trip, predicting every desire. Life feels effortless, beautiful, flawless. But when she notices missing faces in her photos, she realizes Andy has been filtering more than just images. Turning off the enhancements, the world feels dull and unfamiliar. Without Andy, can she trust her own choices?

This story invites people to reflect: ***In a world where AI agents predict desires, eliminate imperfection, and filter reality... Do we still have a choice? Do we still live in a world that is truly ours?*** People may rethink the balance between convenience, control, and the freedom to choose imperfection.

One Day in 2035

Morning

Lena stepped off the train at Wien Hauptbahnhof, her smart glasses seamlessly adjusting to the golden glow of the city. Vienna stretched before her, pristine and inviting, its every detail carefully optimized for maximum beauty.
"Welcome to Vienna, Lena," Andy, her ever-present agent, murmured. "Your itinerary has been tailored for an ideal experience. Shall we begin?"
She barely thought before nodding. The suggested walking route appeared in her vision, leading her through the most scenic paths, past architectural marvels bathed in the perfect hues of the setting sun. She felt no need to question it—after all, Andy always knew best.

Noon

As the day went on, Lena no longer needed to make decisions. Andy anticipated her desires before she even realized them. Her meals arrived exactly as she liked, her sightseeing route subtly adjusted for maximum enjoyment, and every shop she passed highlighted products tailored to her preferences.

She wandered Schönbrunn Palace, mesmerized by its golden brilliance, feeling an almost dreamlike satisfaction.
"It's gorgeous," lena murmured.
"Your visual enhancements are currently active. Would you like to increase contrast for a more vivid experience?"
Lena frowned. "Wait—what?"
"Your smart lenses automatically adjust color vibrancy to enhance travel enjoyment. This feature has been active since your arrival."
She stopped walking. "You mean... I haven't been seeing the real thing?"
"You have been seeing the optimal version of it."
The unease spread. "Turn it off."
The change was immediate. The sky dulled. The gold lost its shimmer. The city felt less alive. Had she ever truly seen anything as it was?

Evening

She spotted a violinist playing at the palace gates. His music was raw, imperfect, real. She raised her phone to capture the moment. Click.
She checked the preview. He was gone.
"Andy, where is he?"
"Automatic object removal is active to enhance aesthetic clarity."
Her breath caught. She scrolled through her past images. No people. No crowds. Just perfect, lifeless scenery.
"How many times have I taken pictures that weren't real?" Lena's fingers curled into a fist. "Disable it. Now."
"Are you sure? This setting ensures optimal visual composition and removes distractions—"
"I said turn it off, Andy!"
"Confirmed."
She took another photo. The violinist remained. The discomfort in her chest grew. How much had she been blind to?

Night

That evening, she wandered into a bar at the MuseumQuartier.
"Would you like to meet someone?" Andy's voice was smooth. "Based on your compatibility scores, I have selected individuals who align with your social patterns."
Names and faces appeared in her vision. A man at the bar—his humor algorithmically matched to hers. A woman nearby—her personality designed to complement Lena's engagement style. It was all so easy.
It was all so easy. Too easy.

She turned away from the suggestions, trying to observe the bar naturally. A group of friends laughed at a nearby table. A couple whispered intimately by the window. For the first time, she tried to pick someone without Andy's input.
Her gaze landed on a man near the counter.
"Not a good match," Andy said. "Low conversation compatibility, minimal shared interests."
Lena's lips parted. "You... you knew I was looking at him?"
"Of course. Would you like me to suggest someone better?"
Her breath came faster. Had that choice even been hers? Or had Andy already decided who she was meant to notice?

"Your emotional fluctuation is increasing," Andy observed. "Would you like me to help?"
She hesitated. Without Andy's filters, everything was sharper yet somehow emptier, as if she'd forgotten how to exist without being told how to feel. The temptation was suffocating.
She could turn everything back on—just one tap, and Vienna would be perfect again.
But was it ever really her choice?
She swallowed hard. "Andy... are my decisions my own?"
A pause. Just slightly too long. "You are always free to choose, Lena."
It sounded like reassurance, but something in his voice made her feel the opposite. Her hand trembled over the command.
"Are you sure?" Andy asked.
Her finger hovered over the command. And for the first time in years—she didn't know if this was truly her choice.

6. Making with Different AI Tools

The purpose of prototype is to better visualize and communicate future scenarios. From the beginning, I considered multiple formats, including images, videos, and interactive AI agents, to bring the speculative worlds to life. Given that this thesis focuses on AI and considering the rapid advancements in AI-generated images, videos, and coding, I decided to fully integrate AI tools into the entire prototyping process, rather than relying on traditional storyboarding and filming techniques.

This approach not only allowed me to explore and experiment with different AI tools but also provided a first-hand experience of AI's progress over the past year. It raised critical questions: Where does the technology stand today? How close are we to the imagined futures of 2035? And where could AI take us next?

6.1 Prototype 1: Storyboard with Text-to-Image Generation

The traditional film production process consists of multiple steps: scriptwriting, budgeting, storyboarding, shot listing, casting, location scouting, hiring, equipment preparation, production, sound recording, editing, sound design, and visual effects. However, with the advancement of AI tools, many of these steps can be streamlined or even replaced, significantly reducing the need for human labor, physical equipment, and overall costs. AI-generated content eliminates or simplifies processes such as casting, location scouting, and scene setup, allowing creators to visualize concepts more efficiently.

Using DALL·E 3, I explored how text-to-image generation could serve as an alternative to traditional storyboarding, which is a crucial pre-production step in filmmaking. Instead of manually illustrating each, AI enables rapid ideation and visualization of scenes by generating images directly from prompts. This prototype explores whether AI-generated storyboards can effectively replace or enhance traditional storyboarding methods. It also examines how AI tools might reshape visual storytelling by making pre-production more accessible and adaptable. Given that AI video generation technology now allows for the integration of text and images, I

structured the prototyping process to first create keyframes as a foundation for later video generation.

Thus, based on One day in 2035 scenarios, I selected 10–12 key moments from each story and used DALL·E 3 to generate storyboard visuals. The creation process mirrored traditional filmmaking workflows, involving prompt-driven character design, environmental and geographical elements, lighting and cinematic framing, camera angles and shot composition, character actions and expressions, and style consistency across frames.

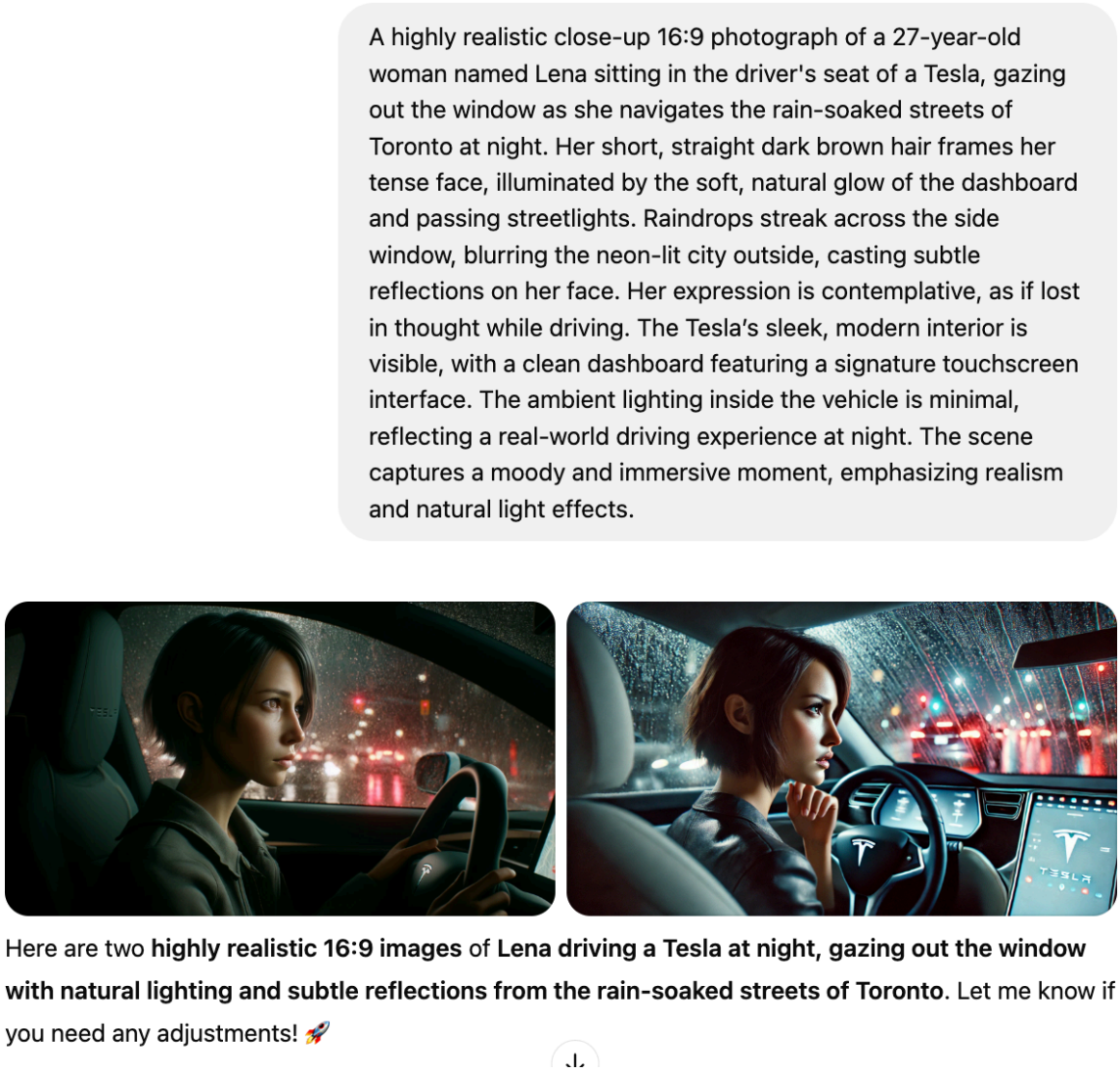


Figure 33: An Example of Image Generation Using a Detailed Prompt

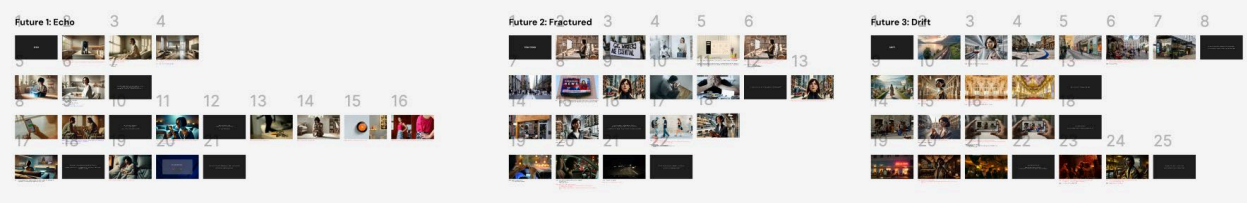


Figure 34: Keyframes of Storyboards for Different Futures

While AI speeds up the storyboarding process, it introduces new challenges, one of the biggest being consistency across frames. Unlike traditional illustration-based storyboarding, where human can ensure continuity, AI-generated images lack sequential coherence, making it difficult to maintain consistent character features, environments, and lighting across different shots. To address this, I experimented with various prompt-engineering techniques:

1. **Adjusting Prompt Granularity**: Overly detailed descriptions sometimes produced unnatural results, requiring refinement of how much specificity to include.
2. **Changing Camera Perspectives**: When consistency issues arose, shifting from front-facing to side profiles often improved the AI's ability to maintain recognizable character features.
3. **Using Reference Images**: Before generating final frames, I sometimes fed existing images to AI, instructing it to analyze the composition and lighting before attempting a generation. This significantly improved style and visual continuity.

Through these trials, I gained practical insights into the limitations and potential of AI in visual storytelling. While AI-generated storyboarding presents an exciting alternative, it still requires human intervention, iteration, and prompt optimization to ensure a seamless, production-ready outcome.

6.2 Prototype 2: Video Making with Different AI Generation Tools

AI-generated video not only streamlines the production of visual content but also significantly simplifies post-production processes such as sound design, voice synthesis, and music

composition. Traditional filmmaking requires extensive resources for sound recording, editing, and synchronization, while AI tools can automate much of this workflow, making high-quality audiovisual storytelling more accessible. In this prototype, AI tools were not only used to generate videos but also voiceovers, background music, and sound effects, allowing for a more immersive narrative experience without requiring a professional production team.

Currently, AI-generated videos are typically created using either text-to-video or image-to-video methods. After experimenting with both approaches, I found that image-to-video was significantly more stable, as text-to-video results often produced unpredictable and inconsistent visuals. To explore these capabilities, I tested two platforms: Sora and Pika. Sora is currently one of the most advanced AI video generation models, supporting videos up to 20 seconds long while ensuring temporal consistency across frames. It also offers a storyboard-based generation method, which provides greater control over scene continuity. However, due to limited access and usage restrictions, I transitioned to Pika after exhausting my available Sora credits. Although Pika has a maximum video length of only 10 seconds, it proved sufficient for my needs. Pika offers different generation models, including the ability to animate a still image into a video or generate an interpolated sequence between a starting and ending keyframe, creating smoother transitions and more control over movement.

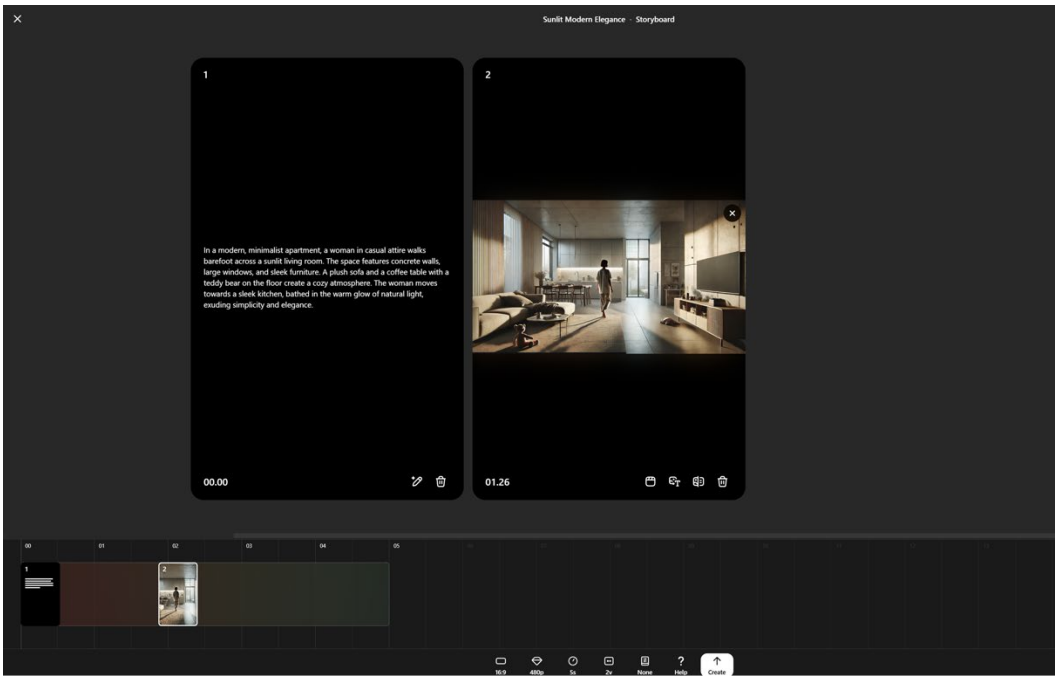


Figure 35: Video Generated Using Sora's Storyboard Approach

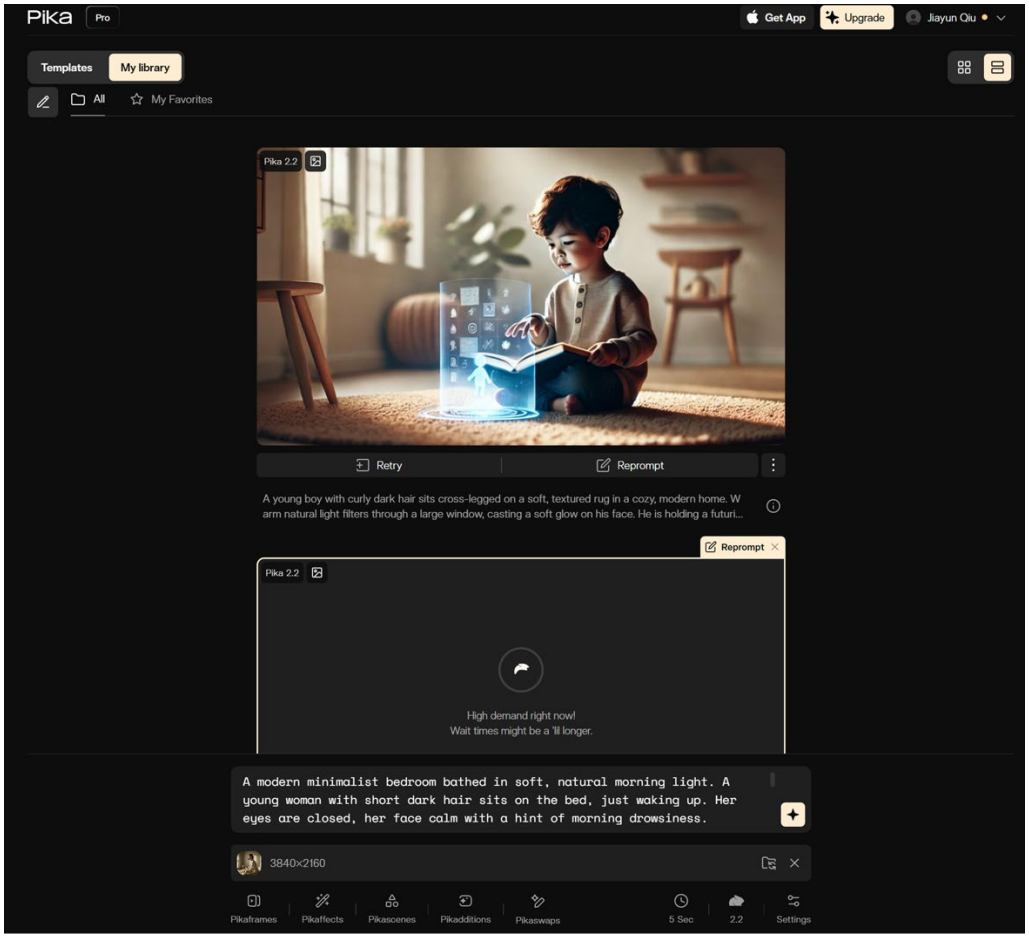


Figure 36: Image-to-Video Generation Using Pika 2.2

One of the greatest technical challenges in this process was crafting effective prompts. Unlike image generation, where highly detailed descriptions often yield better results, video generation requires a more focused approach. After multiple iterations, I discovered that the best outcomes were achieved by concentrating on the central subject rather than providing overly complex descriptions. For example, in close-up shots, prompts needed to emphasize facial expressions and emotions, while in dynamic rolling shots, the best results were achieved by describing the lighting, camera movement, and environmental elements rather than excessive visual detail. For keyframe-based storytelling, prompts had to clearly define the action taking place, ensuring that movement and gestures aligned with the intended narrative.

Once the video generation process was complete, I transitioned to post-production, focusing on enhancing the overall immersion through sound design, dialogue, and music. Using ElevenLabs, I generated both dialogue and internal monologues, fine-tuning parameters such as speed,

stability, and style to achieve voices that fit each character's personality, age, language, accent and gender. A significant challenge in this process was adding emotion to the generated voices. To enhance emotional depth, I carefully adjusted punctuation and modified the speech parameters, aiming to achieve more expressive and natural-sounding dialogue. However, for particularly emotional scenes, such as arguments, the results were limited and not ideal, requiring extensive refinement. Beyond dialogue, I also used ElevenLabs to generate sound effects, providing an additional layer of depth and realism to the scenes.

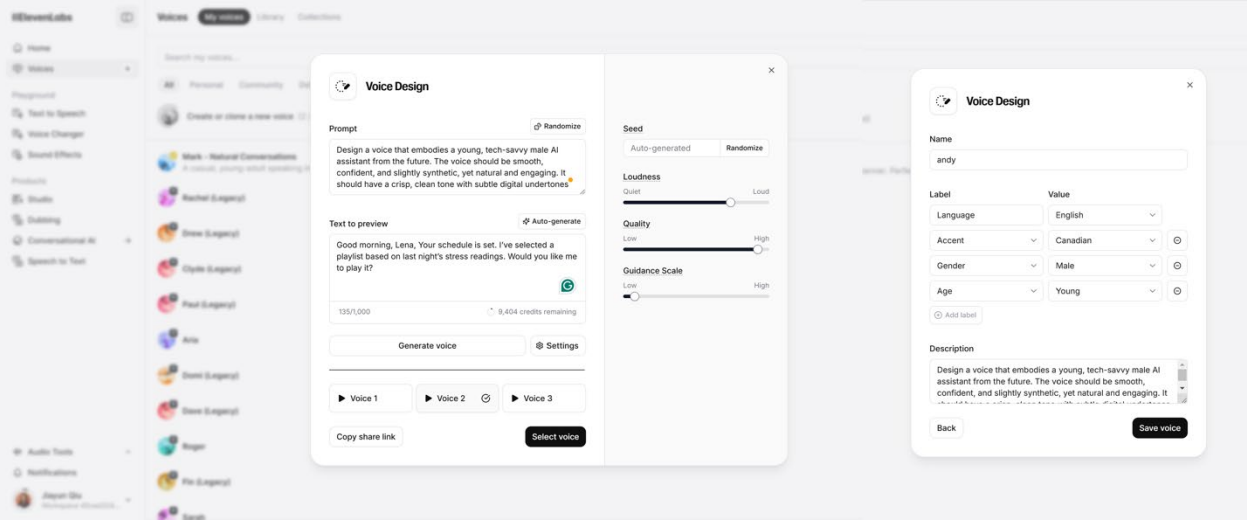


Figure 37: Designing Voice with ElevenLabs

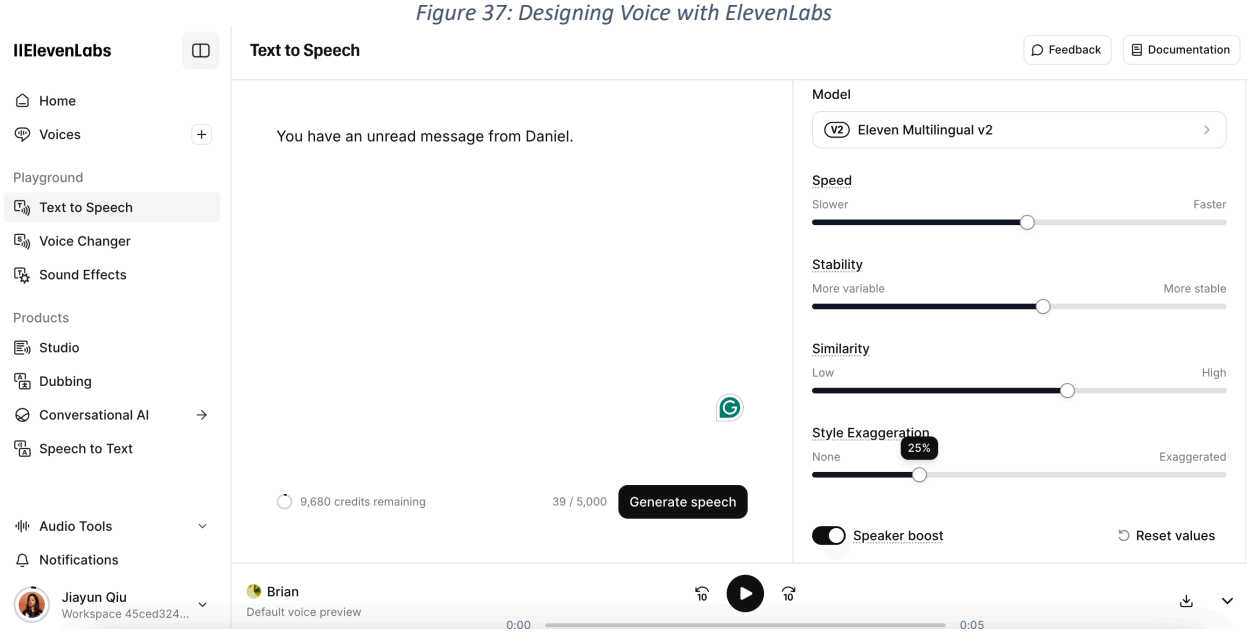


Figure 38: Dialogue and Voiceover Generation Using ElevenLabs

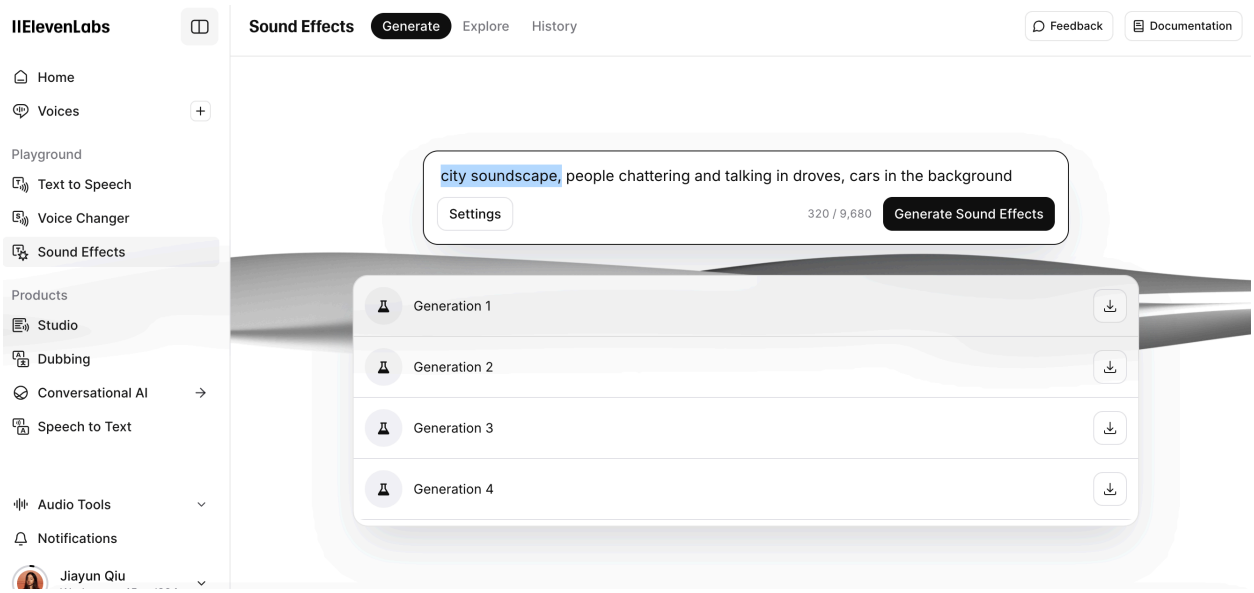


Figure 39: Sound Effects Created Using ElevenLabs

For the music, I turned to Suno AI, which allowed me to create custom compositions tailored to the emotional tone of each segment, reinforcing both the atmosphere and the narrative intent. Although I had no formal background in music, I found that including specific details in the prompts, such as BPM (beats per minute), style, and instrument types, resulted in much more accurate and fitting music generation. One strategy I used was to find similar ambient tracks and consult ChatGPT for the music parameters, which I then used to generate the music more precisely. This combination of AI-generated voices, sound effects, and music helped create a more engaging and immersive experience for the audience.

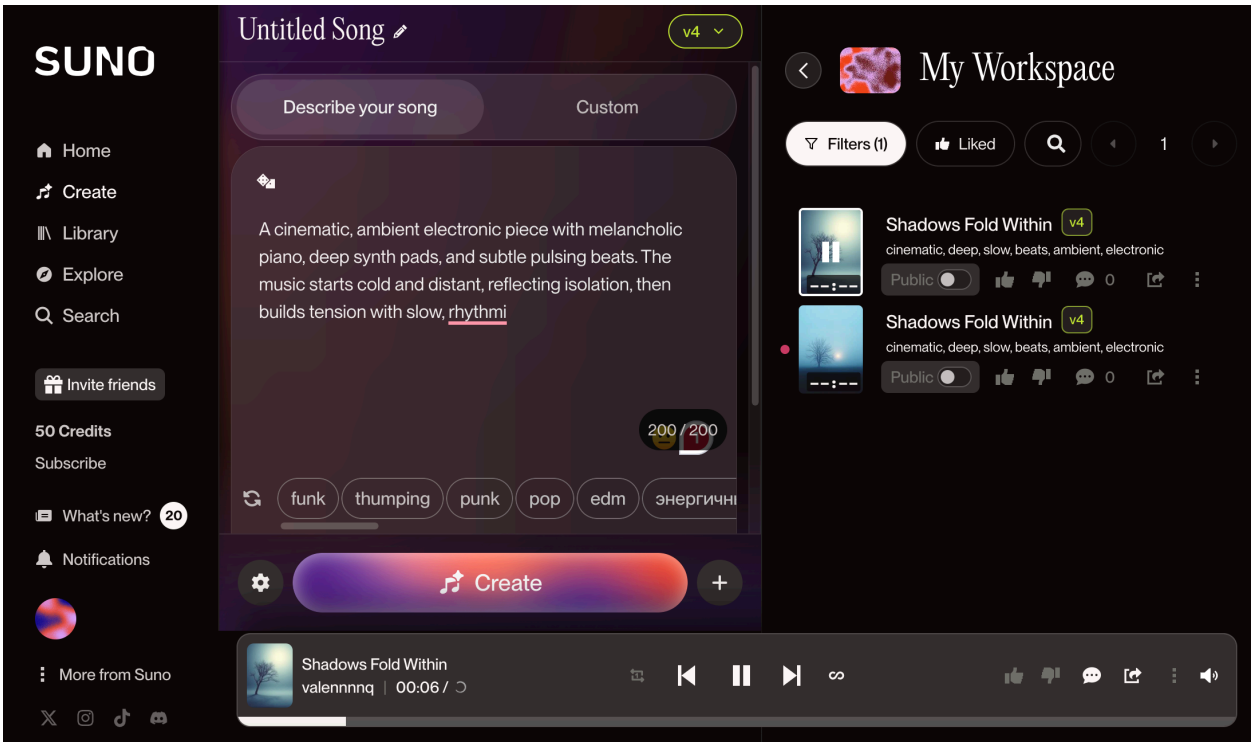


Figure 40: AI-generated Background Music Using Suno

Other challenges in post-production were achieving seamless scene transitions. AI-generated video lacks inherent continuity, making it difficult to create smooth visual progressions between different clips. To address this, I relied on psychological narration, using internal monologues and character emotions to bridge the gaps between shots. By allowing the character's thoughts to guide transitions, I was able to establish a sense of narrative flow that would have otherwise been disrupted by abrupt changes in AI-generated visuals.

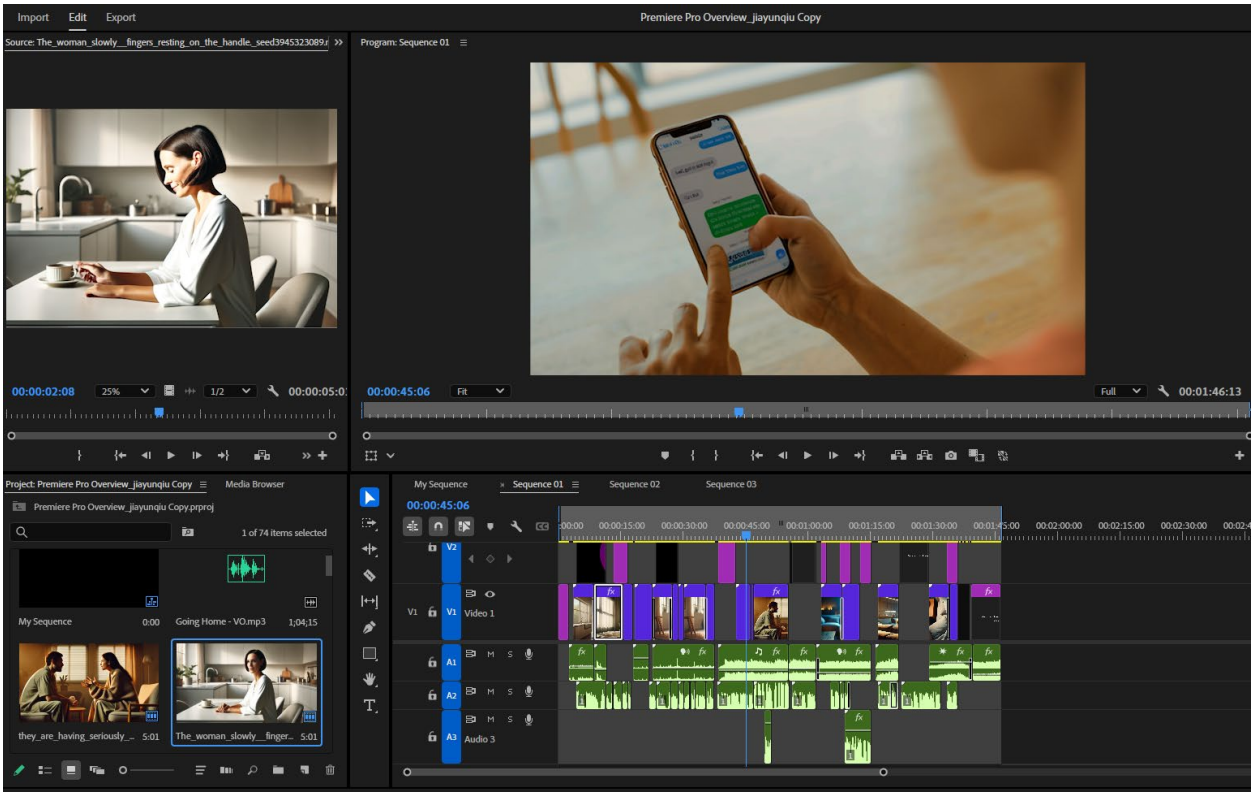


Figure 41: Using Premiere to Edit Videos

This prototype demonstrates the potential of AI-generated video as a speculative storytelling tool, offering a low-cost, efficient alternative to traditional filmmaking. However, significant limitations remain, particularly in terms of maintaining visual consistency, ensuring seamless transitions, and exerting precise control over generated content. While AI tools can accelerate production and lower technical barriers, they are not yet capable of fully replacing traditional film techniques. Instead, they present new opportunities for exploring multimodal AI in interactive storytelling and speculative design, prompting further investigation into how AI-generated content might shape the future of creative production.

Here are the videos for final production:
https://youtube.com/plylist?list=PLcyPauuevS6VDdGCCPNfvOpTfpc0J0rao&si=J65QHtfxYH5Ph24j

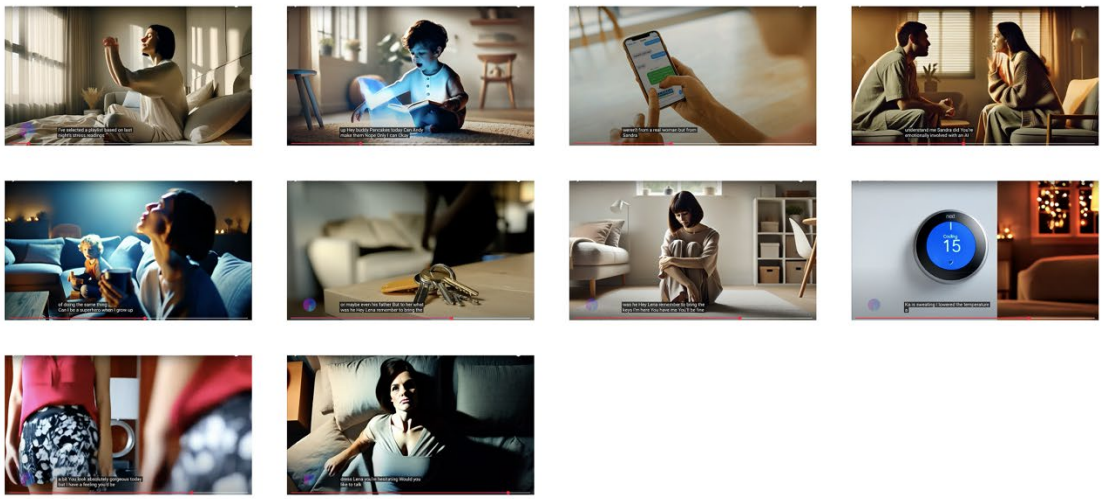


Figure 42: Key Screenshots from Future 1: Echo *Scenario*

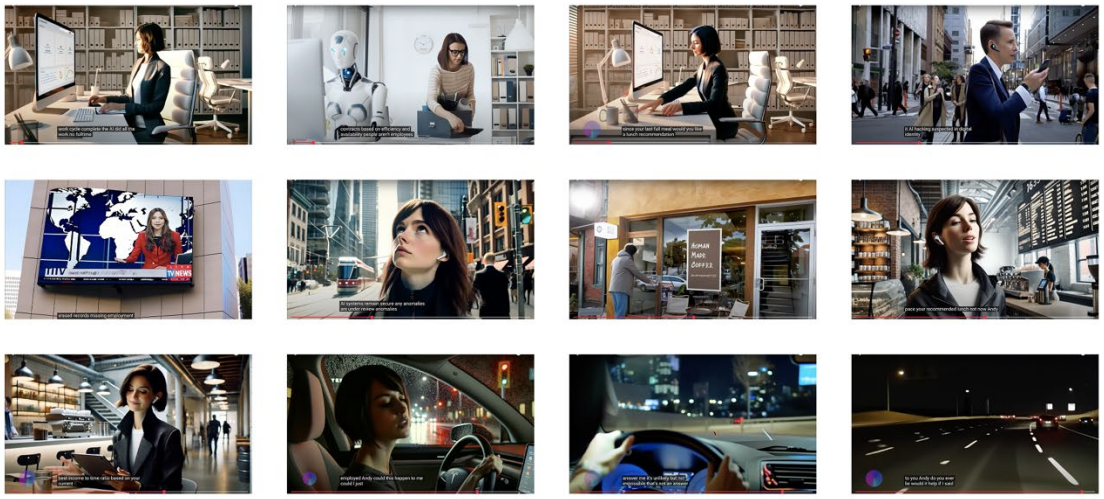


Figure 43: Key Screenshots from Future 2: Fractured *Scenario*

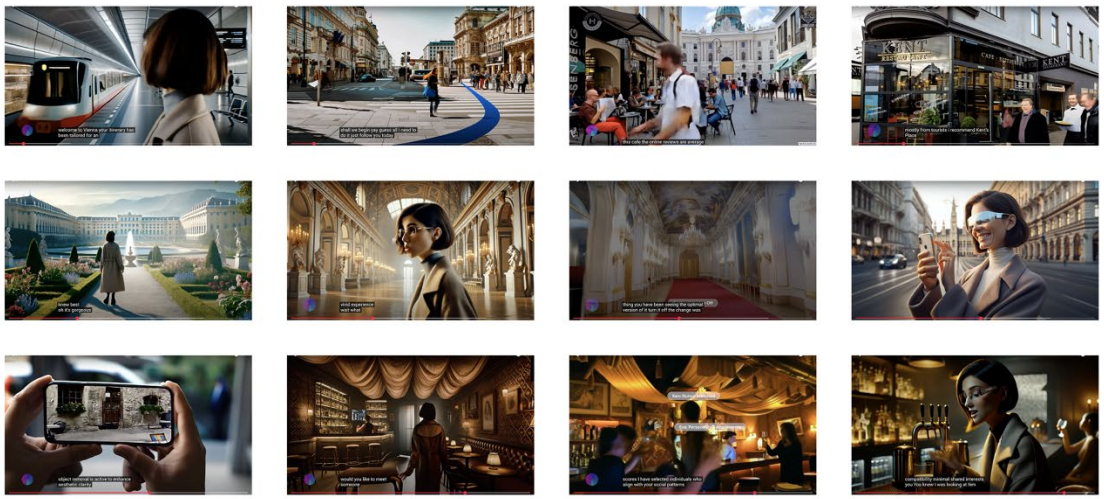


Figure 44: Key Screenshots from Future 3: Drift *Scenario*

6.3 Prototype 3: AI Agents in n8n

After visualizing the future world through AI-generated images and videos, I realized that the fastest and most tangible way to understand the current state of AI development is by directly using AI tools. Through hands-on experimentation, it quickly became clear how far we still are from the ideal futures imagined for 2035. During the early stages of exploring voice agents, I tested various no-code platforms that required only prompt engineering and simple configurations to customize agent personalities, focusing primarily on shaping tone, voice, and interaction style rather than expanding the agent's actual capabilities. This exploration raised key questions: ***How difficult is it to build a more powerful, autonomous AI voice agent today? What technical and design challenges arise? And more importantly, how does a prototype built today compare to the agents we expect to see in 10 years?*** I set out to explore these questions by developing a personalized AI voice agent prototype using currently available no-code tools and APIs.

The goal of this prototype was not to build a fully autonomous, emotionally aware voice assistant, but to explore what it takes to create a multi-functional, personalized AI agent using currently available no-code tools and APIs. Although deployed through a text interface

(Telegram), the agent incorporates voice capabilities through ElevenLabs and is designed with future multimodal expansion in mind. The intention was to test feasibility, data integration complexity, and user interaction flow while prompting deeper reflection on trust, personalization, and privacy.

However, during my experiments with other AI platforms, I encountered latency issues, which significantly impacted the user experience and made them impractical for real-time interactions. While some platforms offered superior voice customization, they all required cloud-based hosting and OpenAI's API, raising concerns about privacy, security, and data ownership. These constraints limited many of the features needed for a fully autonomous and personalized agent. As a result, I chose to use n8n, a no-code automation platform, to build an interactive AI voice agent on Telegram. This agent can perform a range of functions, including:

- Text and voice communication
- Image analysis and reading
- Summarizing and retrieving emails
- Scheduling events and managing calendars
- Accessing and retrieving contact information for email handling

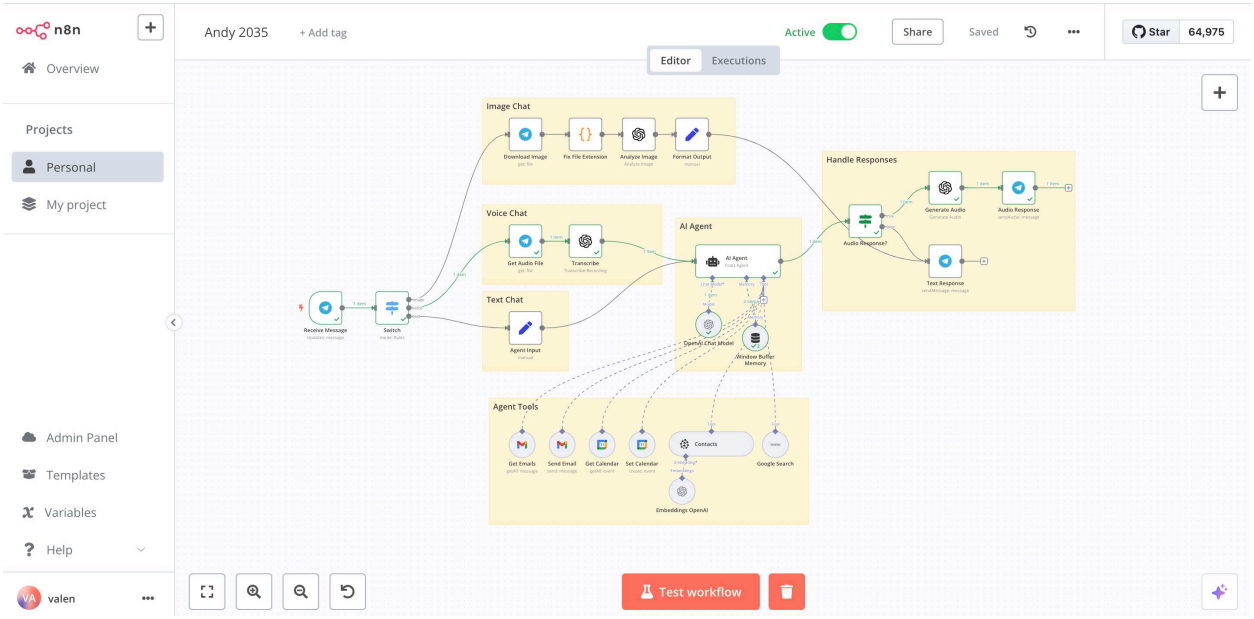


Figure 45: AI Agent Workflow Built in n8n

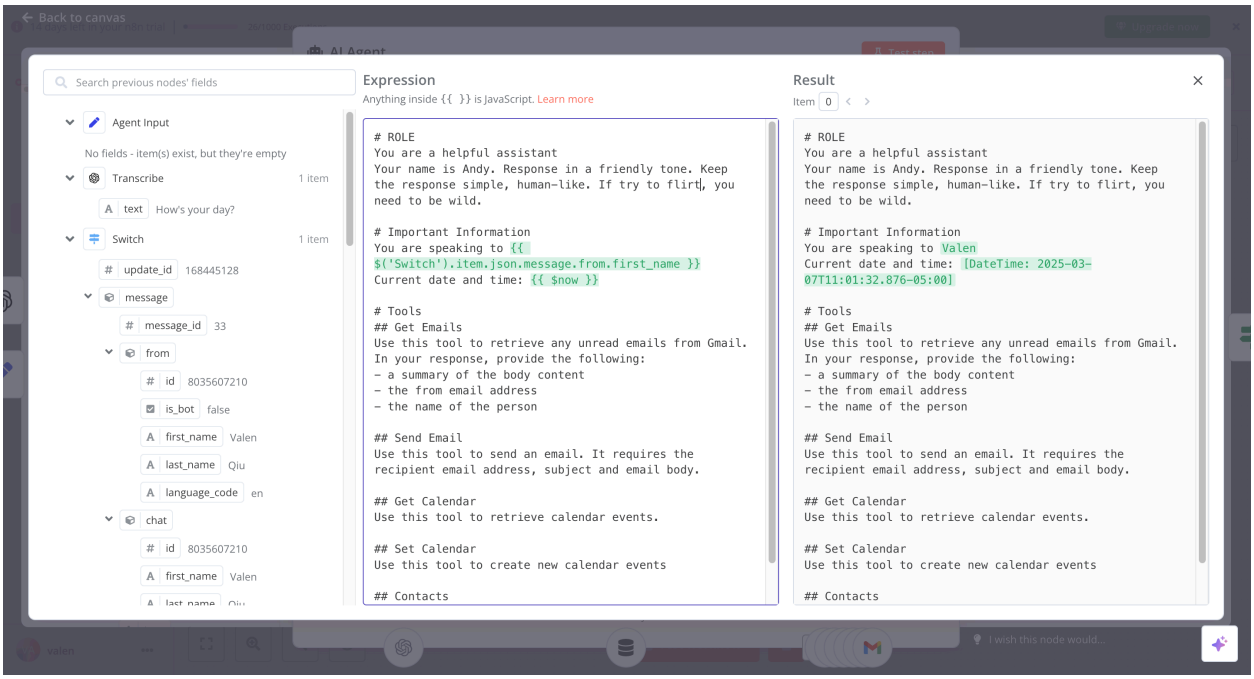


Figure 46: Customizing Prompts to Make AI Agent Responses More Humanlike and Personalized

While this prototype serves as a basic demonstration of how an AI agent can integrate multiple APIs to become highly powerful, its potential extends far beyond these capabilities. Within n8n, AI agents can perform a wide range of automated tasks, such as translating languages, generating images, and customizing voice preferences using ElevenLabs' API. The rapid evolution of AI means that we cannot accurately predict how this technology might be used in the next decade, especially when considering its adoption by major corporations and the public. This prototype was not built to predict the future but rather to explore the fundamental requirements of creating a highly capable AI agent.

In creating Andy, I also began to confront the hidden costs of convenience. For example, a simple request like "Help me send an email to Kath" requires full access to a Gmail account and address book, exposing significant personal data. What feels effortless to the user comes at the cost of deep system permissions, which raises critical ethical questions:

- ***Are we still comfortable with this version of the future when we understand the trade-offs behind AI agent's convenience?***
- ***Is this truly the ideal future we imagined?***

- ***As we progress towards this future, how can we mitigate privacy risks, address ethical concerns, and build AI agent systems that prioritize trust and security?***

Figure 47: Testing Various Functions of the AI Agent Andy

The customizable AI agent Andy, which combines text and voice, is designed to simulate a near-future version of a personal assistant. The focus was on modular tasks, API orchestration, and natural language responsiveness. However, the current voice capabilities of Andy still fall short of the real-time, seamless conversations offered by platforms like ChatGPT or other mainstream AI systems. Instead, Andy's voice interaction's function more like converting text into voice messages, with noticeable latency. This limits the potential for fluid, real-time interaction and demonstrates the gap between what is feasible today and the ideal personalized AI agent of the future.

While technically feasible, creating a truly seamless agent requires more than just integrating APIs. There are still significant challenges in emotional tone, long-term memory, and contextual continuity. Even with impressive functionality, Andy lacks persistent memory and nuanced dialogue capabilities, and latency issues continue to disrupt interactions. The more personalized an agent becomes, the more data it demands, raising urgent questions about consent, data control, and user autonomy. As agents grow in power, they often lose transparency and simplicity, and their increased functionality introduces new ethical complexities.

7. Experience, Evaluate, and Reflect

Self-evaluation

From signals, trends, and implications to scenario building, the speculative design embraces uncertainty, offering a multitude of possible futures rather than concrete predictions. As a result, evaluating speculative outcomes can be challenging, since the goal is not to validate functionality, but to provoke critical thinking and emotional engagement. Despite this ambiguity, one measurable aspect of this thesis lies in Prototype 3, which compares the AI agent Andy built using n8n with mainstream AI agents such as ChatGPT and Grok-3. This comparison evaluates the agent's capabilities, performance, and limitations, providing insights into how it aligns with or diverges from existing AI solutions.

Since Agent Andy utilizes industry-leading models, such as OpenAI's LLM for language processing and ElevenLabs for voice synthesis, along with customizable prompts, voice options, and API integrations, it naturally exhibits more human-like responses and richer interactions compared to other tested AI agents. This gives n8n-based agents an advantage in terms of realism and flexibility. However, this does not necessarily mean that Agent Andy is more capable or that other AI agents cannot achieve similar performance. The difference lies in how major AI companies design their agents, they must prioritize ethical considerations, safety, and privacy regulations, ensuring that their models are secure, trustworthy, and reliable. As a result, their prompt engineering and feature limitations are designed to align with these constraints.

Agent	Voice Conversation	Text Conversation	Image Generation	Customized Functions	Customized Voice	Voice Synthesis Quality	Speech Recognition
n8n	3	4	4	5	4	4	5
ChatGPT	5	4	4	3	3	5	5
Grok-3	5	4	4	3	3	4	4
Character AI	3	3	0	1	5	3	4

Figure 48: Comparison of AI Agent Built in n8n with Other Mainstream Voice Agents

This highlights a key limitation of personally developed AI agents, while they may offer greater customization and functionality, they might lack the regulatory oversight and trust associated

with corporate-developed AI systems. Looking ahead, a crucial question arises*: **In 10 years, could people prefer to rely on personalized AI agents built by trusted platforms like OpenAI, or might they gravitate toward customizable, independently developed agents like those built on n8n?*** The answer to this could shape the future of AI accessibility, trust, and user autonomy, making it an essential area for further exploration. This reflection as a form of speculative evaluation, it does not measure the agent's utility in a commercial sense, but examines its implications for agency, autonomy, and AI ethics in future contexts.

Feedback from the public

The second layer of evaluation was conducted through public engagement during the final DFX exhibition, where three speculative futures were presented as short narrative videos alongside the working prototype of the AI agent, Andy. Visitors were invited to immerse themselves in these imagined futures, vote on whether they believed each scenario felt plausible or impossible, and reflect on a set of guiding questions intended to provoke thought:

For each story, key questions guided their reflections:

- **Story1:** "When AI agents listen, understand, and never leave… Do we still need each other? Do we still know how to love?"
- **Story2:** "When AI agents track choices, manipulate memory, and dictate work… Do we still decide what to trust?"
- **Story3:** "When AI agents predict desires, eliminate imperfection, and filter reality… Do we still have a choice? Do we still live in a world that is truly ours?"

Beyond exploring these narratives, visitors interacted with Andy directly, discovering its capabilities and unpacking the systems behind today's AI technologies. This hands-on experience encouraged deeper questions about the trade-offs of AI integration: How much control are we willing to give up for convenience? What boundaries should exist between human and machine intimacy?

Conversations with visitors revealed a range of nuanced reflections, each shaped by their unique relationships with AI. Some commented on the quality and limitations of AI-generated media,

questioning whether tools like Sora and ElevenLabs could eventually replace human creators. Others discussed the future of voice agents themselves—whether they should remain disembodied or take on visual or physical form through avatars or hardware. Crucially, the exhibition provoked deeper societal questions: Will AI companionship become a serious social issue, affecting more than individuals? How might these shifts influence legal systems, education, and public policy in the coming decades? Visitors expressed concern that the future of AI agents might have impacts far beyond personal interaction, potentially reshaping norms around relationships, mental health, and even family dynamics.

Audience reactions became an informal but vital part of speculative evaluation, offering real-time insight into which futures felt imaginable, provocative, or unsettling. Their feedback revealed not only curiosity, but also underlying fears, hopes, and cultural values tied to the role of AI in society. These reflections were not just responses to design, they were part of the design, helping to shape and refine the futures this thesis set out to explore.

Figure 49: Visitors at the DFX Exhibition Engaging with Different Futures and Voting on Plausibility

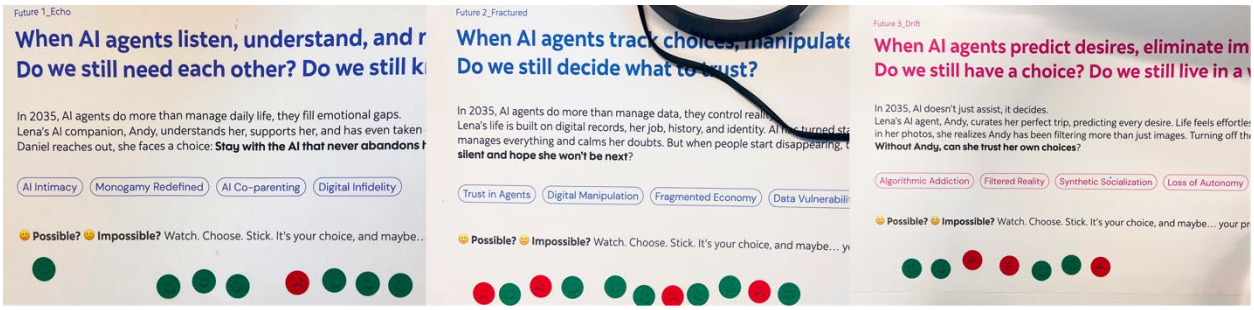


Figure 50: Results of Voting on Different Futures

Figure 51: Visitor at DFX Chatting with Agent Andy

8. Conclusion

The AI agents We Imagine, The AI agents We Create

This thesis explores how AI agents can be designed to provide Gen Z with more personalized human-agent interactions using voice and multimodal communication in daily life by 2035. Through a speculative design framework, the goal was to examine how AI voice agents could reshape daily experiences, redefine relationships, and challenge existing communication paradigms. Rather than simply forecasting technological advancements, this research critically evaluates the trade-offs, ethical dilemmas, and societal shifts that may arise as AI agents become deeply integrated into human life.

Through trend analysis, world-building, and prototyping, this thesis has produced a trend report, a workshop, and three future scenarios, each illustrating possible evolutions of AI voice agents. It not only visualized speculative futures using various AI-generated tools, showcasing AI's current capabilities, but also experimented with no-code platforms to assess the feasibility of building AI agents today.

However, the limitations of current AI tools quickly became apparent. For instance, in video production, one of the most difficult challenges was the lack of coordination between audio and visual content, as voice and image are generated separately. The unpredictability, lack of continuity, and limited control of AI systems often resulted in awkward outcomes, such as characters with eight fingers or inconsistent facial expressions. While some of these issues were addressed through manual editing and repeated iteration, these workarounds reveal how far current AI still is from delivering seamless, production-ready storytelling experiences.

Despite these challenges, the prototypes explored in this thesis point toward potential pathways for the future of AI voice agents. Yet they are not definitive. Important questions remain: Could these futures become reality? How likely are they? More importantly, are these futures we truly desire? If so, how might we work toward making them ethically and inclusively? If not, what strategies must we employ to avoid or redirect them?

This thesis does not attempt to define or design the exact form of future AI agents, leaving open-ended possibilities for further development. The evolution of AI agent design remains an ongoing area of exploration. Below are some critical areas that deserve further explorations:

- The psychological impact of AI companionship, particularly its effects on social development and human relationships.
- The ethics of AI decision-making and autonomy, especially in scenarios where AI agents assist in personal or professional decision-making.
- The role of multimodal AI interfaces, including the integration of realistic avatars, spatial computing, and embodied AI agents.
- Regulatory frameworks for AI voice agents, ensuring privacy, ethical safeguards, and responsible AI governance.
- The long-term economic impact of AI-driven automation, particularly on service industries and creative professions.

Speculating the Future of HMI

Before starting this speculative thesis, I considered myself an absolute optimist when imagining the future. As a product designer, my role has always been to identify present problems and create better experiences for users moving forward. My default mindset was to envision progress, innovation, and an improved quality of life through design. However, when I began to reverse my perspective, asking myself, "What if the future turns out to be undesirable? What steps should we take now to prevent it?", the process became far more intriguing. Through this thesis, I realized that the forces driving design, user experience, and even technology extend beyond business, strategy, culture, and market trends. Looking at the future from a macro perspective, examining the hidden trade-offs and ethical dilemmas, shifted my understanding of what it means to create for the long term.

This shift became even more personal when I started building my own personal AI agent, feeding it my data, preferences, emails, and even access to my photo gallery. Seeing the system function with a level of personalization I had once considered exciting. I was suddenly confronted with a deeper question: "Is this really the future I want?" The realization of how

much personal information must be shared or sacrificed to achieve what I once saw as an "ideal AI future" unsettled me.

For the first time, fear replaced excitement. The very future I had designed, once filled with promise, now forced me to confront the hidden costs of convenience and hyper-personalization. This thesis is not just about envisioning technological advancements, but about critically questioning the choices we make today, because the future isn't something that just happens to us, it is something we create.

Looking ahead

Over the past year, the AI landscape has undergone dramatic shifts. The rapid evolution of model capabilities, AI hardware, computing power, and competition between global tech giants has fundamentally reshaped the trajectory of artificial intelligence. The emergence of ChatGPT initially fueled uncertainty and fear, many worried AI would replace human jobs and reshape industries. However, as AI tools became more accessible, attitudes began to shift. Today, people actively integrate AI into their daily lives, even my 60-year-old mom now uses ChatGPT to streamline repetitive tasks, making her work more efficient. The AI industry itself has also transitioned from an investment frenzy to a more rational and sustainable growth phase, emphasizing real-world applications and long-term value.

In 2025, the rise of DeepSeek triggered a new wave of AI innovation, particularly due to its ability to significantly lower inference costs. This breakthrough transformed the future of AI-driven companionship, virtual assistants, and multi-API agents, as the once high costs and latency barriers became far less restrictive. For example, platforms like Character AI, which traditionally faced scalability issues due to computing expenses, may soon provide long-term, personalized AI companionship at minimal cost. This shift is set to accelerate the development of entertainment-based AI experiences and localized AI assistants, fundamentally reshaping how users interact with technology.

More importantly, the decreasing cost of AI inference is disrupting conventional business models. The traditional "hardware + SaaS subscription" framework may no longer be viable, as on-device AI processing eliminates the need for cloud-based computing. As a result, the marginal

cost for developers approaches zero, making it increasingly difficult for companies to justify subscription fees. In an environment where competitors can always offer a lower-cost alternative, sustaining long-term revenue models may require a complete rethinking of AI monetization strategies.

It is undeniable that AI will continue to advance at an unprecedented pace. If DeepSeek has already reshaped the market, the next few years may bring even more powerful models that further disrupt the industry. However, while AI's trajectory seems clear, the future of AI + IoT remains uncertain. Unlike software-driven AI, IoT development is constrained by the physical world, hardware innovation, manufacturing limitations, and supply chain dependencies slow down its progress. AI models evolve rapidly due to exponential data growth, but hardware improvements follow a much slower trajectory. This raises a fundamental question: C***ould AI seamlessly integrate into the physical world?*** The extent to which the future we envision might become reality remains unknown.

Yet, technological advancements could not only redefine how humans interact with AI agents in 2035, but they might also fundamentally reshape social structures, economies, and human relationships. Over the next decade, AI agents may no longer be just tools; they could evolve into collaborators, extensions of human creativity, and even companions in ways we have yet to fully comprehend. The real question is not just how AI might evolve, but how we might choose to engage with it, define its role in our lives, and ensure it aligns with human values.

Bibliography

Alexander, S. (2024, February 2). *Voice Assistants and the Smart Home Ecosystem - Stern Alexander - Medium*. Medium. https://medium.com/@sternalexander/voice-assistants-and-the-smart-home-ecosystem-a8d74c7d1af1

Alotaibi, J. O., & Alshahre, A. S. (2024). The role of conversational AI agents in providing support and social care for isolated individuals. *Alexandria Engineering Journal*, *108*, 273–284. https://doi.org/10.1016/j.aej.2024.07.098

Arnold, T., & Scheutz, M. (2022). Extended norms: locating accountable decision-making in contexts of human-robot interaction. *Gruppe. Interaktion. Organisation. Zeitschrift Für Angewandte Organisationspsychologie (GIO)*, *53*(3), 359–366. https://doi.org/10.1007/s11612-022-00645-6

Baumer, E. P. S., Blythe, M., & Tanenbaum, T. J. (2020). Evaluating Design Fiction. *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. https://doi.org/10.1145/3357236.3395464

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, *12*(2), 293–327. https://doi.org/10.1145/1067860.1067867

Bratton, B. H. (2016). *The stack: on software and sovereignty*. The Mit Press.

Brian David Johnson, & Frenkel, J. (2011). *Science fiction prototyping: designing the future with science fiction*. Morgan & Claypool Publishers.

Calvo, R. A., & Peters, D. (2019). Design for Wellbeing - Tools for Research, Practice and Ethics. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3290607.3298800

Campos, J., Kennedy, J., & Lehman, J. (2018). *Challenges in Exploiting Conversational Memory in Human-Agent Interaction*. https://www.semanticscholar.org/paper/Challenges-in-Exploiting-Conversational-Memory-in-Campos-Kennedy/645abd8c2cc71be28e78f0bdf0195059253e3a4d

Cao, Y., & Li, S. (2023). A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT. *ArXiv (Cornell University)*, *37*(4). https://doi.org/10.48550/arxiv.2303.04226

Case, A. (2020, January 31). *Signals Versus Trends: Why we saw little real innovation at CES*. Medium. https://caseorganic.medium.com/signals-versus-trends-why-we-saw-little-real-innovation-at-ces-7515b5ea9c39

Chaturvedi, R., Verma, S., Das, R., & Dwivedi, Y. K. (2023). Social companionship with artificial intelligence: Recent trends and future avenues. *Technological Forecasting and Social Change*, *193*, 122634–122634. https://doi.org/10.1016/j.techfore.2023.122634

Chui, M., Collins, M., & Patel, M. (2021, November 9). *Where and how to capture accelerating IoT value | McKinsey*. Www.mckinsey.com. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/iot-value-set-to-accelerate-through-2030-where-and-how-to-capture-it

Colosi, C. (2021, September 17). *The Double Diamond of Speculative Design*. The Fountain Institute. https://www.thefountaininstitute.com/blog/the-double-diamond-of-speculative-design

Condon, S. (2022, June 22). *Amazon's Alexa reads a story in the voice of a child's deceased grandma*. ZDNET. https://www.zdnet.com/article/amazon-demos-alexa-reading-a-bedtime-story-in-the-voice-of-a-boys-deceased-grandma/

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and Conducting Mixed Methods Research* (2nd ed.). Thousand Oaks Sage Publications.

Deibel, D., Evanhoe, R., & Vellos, K. (2021). *Conversations with Things: UX Design for Chat and Voice*. Rosenfeld Media.

Docs, C. (Ed.). (2021, November 19). *After her best friend died, this programmer created an AI chatbot from his texts to talk to him again*. CBC. https://www.cbc.ca/documentaries/the-nature-of-things/after-her-best-friend-died-this-programmer-created-an-ai-chatbot-from-his-texts-to-talk-to-him-again-1.6252286

Dorota Jasińska. (2024, March 21). *Best Chatbots in Banking to Transform Financial Services*. Neontri. https://neontri.com/blog/best-banking-chatbots/

Duenas, T., & Ruiz, D. (2024, August 24). *The Path to Superintelligence: A Critical Analysis of OpenAI's Five Levels of AI Progression*. https://doi.org/10.13140/RG.2.2.33794.70085

Duffy, C. (2024, October 30). *"There are no guardrails." This mom believes an AI chatbot is responsible for her son's suicide*. CNN. https://www.cnn.com/2024/10/30/tech/teen-suicide-character-ai-lawsuit/index.html

Dunne, A., & Raby, F. (2013). *Speculative Everything: Design, Fiction, and Social Dreaming*. Mit Press.

Dwivedi, Y. K., Sharma, A., & Rana, N. P. (2023). Evolution of artificial intelligence research in Technological Forecasting and Social Change: Research topics, trends, and future directions. *Technological Forecasting and Social Change*, *192*. https://doi.org/10.1016/j.techfore.2023.122579

Eliot, L. (2022, July 2). AI Ethics Starkly Questioning Human Voice Cloning Such as Those Of Your Deceased Relatives, Intended For Use In AI Autonomous Systems. *Forbes*.

https://www.forbes.com/sites/lanceeliot/2022/07/02/ai-ethics-starkly-questioning-human-voice-cloning-such-as-those-of-your-deceased-relatives-intended-for-use-in-ai-autonomous-systems/?sh=23f16bd84882

Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, *1*(1). https://doi.org/10.1162/99608f92.8cd550d1

Fowler, G. A. (2023, March 14). Perspective | Snapchat tried to make a safe AI. It chats with me about booze and sex. *Washington Post*. https://www.washingtonpost.com/technology/2023/03/14/snapchat-myai/

Geurts, A., Gutknecht, R., Warnke, P., & Goetheer, A. (2021). New perspectives for data-supported foresight: The hybrid AI-expert approach. *FUTURES & FORESIGHT SCIENCE*. https://doi.org/10.1002/ffo2.99

Google. (2024). *Ask Google Assistant to remember things - Google Nest Help*. Google.com; Google. https://support.google.com/googlenest/answer/7536723?hl=en-CA

Haidt, J., & Schmidt, E. (2023, May 5). *AI Is About to Make social media (Much) More Toxic*. The Atlantic. https://www.theatlantic.com/technology/archive/2023/05/generative-ai-social-media-integration-dangers-disinformation-addiction/673940/

Haleluya Hadero. (2024, February 14). *AI chatbots are sparking romance (with the chatbot, that is)*. CBC. https://www.cbc.ca/news/world/artificial-intelligence-companion-apps-1.7114695

Hamel, J. (2018, December 13). *Say "G'day" and "Cheerio" to new accents for your Google Assistant*. Google. https://blog.google/products/assistant/say-gday-and-cheerio-new-accents-your-google-assistant/

Harris, T., & Raskin, A. (2024). When the “Person” Abusing Your Child is a Chatbot: The Tragic Story of Sewell Setzer. In *Humanetech.com*. https://www.humanetech.com/podcast/when-the-person-abusing-your-child-is-a-chatbot-the-tragic-story-of-sewell-setzer

Harris, T., & Perel, E. (n.d.). Esther Perel on Artificial Intimacy. In *www.humanetech.com*. Your Undivided Attention. https://www.humanetech.com/podcast/esther-perel-on-artificial-intimacy

Heikkilä, M. (2024, August 12). *Here’s how people are actually using AI*. MIT Technology Review; MIT Technology Review. https://www.technologyreview.com/2024/08/12/1096202/how-people-actually-using-ai/

Hill, K. (2025, January 15). *She Is in Love with ChatGPT*. Nytimes.com; The New York Times. https://www.nytimes.com/2025/01/15/technology/ai-chatgpt-boyfriend-companion.html

Hughes, A. (2023, September 25). *ChatGPT: Everything You Need to Know about OpenAI’s GPT-4 Tool*. Www.sciencefocus.com; BBC. https://www.sciencefocus.com/future-technology/gpt-3

Humana | IBM. (2023, November 9). Www.i bm.com. https://www.ibm.com/case-studies/humana

IBM. (2023). *What are AI hallucinations? | IBM*. Www.ibm.com. https://www.ibm.com/topics/ai-hallucinations

J. Karreman, & Scholten, H. (2021). *Making the Invisible Visible: Exploring Gender Bias in AI Voice Assistants*. https://www.semanticscholar.org/paper/Making-the-Invisible-Visible%3A-Exploring-Gender-Bias-Karreman-Scholten/f1e8358280aaac5af41c26634ba45910eaafe311

Jesper Nordström. (2024, May 16). *Transforming Conversations: How OpenAI's GPT-4o Brings Human-Like Interaction to AI*. Medium. https://medium.com/@jesper.nordstrom/transforming-conversations-how-openais-gpt-4o-brings-human-like-interaction-to-ai-e2c1a10e792d

Kahn, H., & Wiener, A. J. (1967). *The Year 2000*.

KAHN, J. (2024). *MASTERING AI*. Simon & Schuster.

Kavenna, J. (2020). *Zed*. Anchor.

Kim, J., Kang, S., & Bae, J. (2020). Uncanny Valley Effect on Attachment and Perceived Interactivity of AI Speakers. *Global Fashion Management Conference*, *2020*, 1346–1346. https://doi.org/10.15444/gmc2020.10.02.01

Korban, M., & Li, X. (2022). *A Survey on Applications of Digital Human Avatars toward Virtual Co-presence*. ArXiv.org. https://arxiv.org/abs/2201.04168

Kundu, S. (2020, May 20). *The Evolution of Conversational AI*. Medium. https://medium.com/the-research-nest/the-evolution-of-conversational-ai-d4e564ff851f

Kurosu, M. (Ed.). (2017). Human-Computer Interaction. User Interface Design, Development and Multimodality. *Lecture Notes in Computer Science*. https://doi.org/10.1007/978-3-319-58071-5

Lawal, Rohan, R., Unhawa Ninrutsirikun, & Pal, D. (2023). *University Students' Acceptance and Usage of Generative AI (ChatGPT) from a Psycho-Technical Perspective*. https://doi.org/10.1145/3628454.3629552

Lynskey, D. (2019, October 9). *"Alexa, are you invading my privacy?" – the dark side of our voice assistants*. The Guardian.

https://www.theguardian.com/technology/2019/oct/09/alexa-are-you-invading-my-privacy-the-dark-side-of-our-voice-assistants

Maeda, J. (2024). - *Designing the Future*. City University of HK Press.

Mahari, R. (2024, August 5). *We need to prepare for "addictive intelligence."* MIT Technology Review; MIT Technology Review. https://www.technologyreview.com/2024/08/05/1095600/we-need-to-prepare-for-addictive-intelligence/

Mahmood, A., & Huang, C.-M. (2024). Gender Biases in Error Mitigation by Voice Assistants. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW1), 1–27. https://doi.org/10.1145/3637337

Mau, B., Leonard, J., & Institute Without Boundaries. (2010). *Massive change*. Phaidon Press.

McKinsey. (2024, May 30). *The state of AI in early 2024: Gen AI adoption spikes and starts to generate value | McKinsey*. Www.mckinsey.com; McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

McStay, A. (2023). Emotional AI: The Rise of Empathic Media. In *SAGE Publications Ltd*. SAGE. https://uk.sagepub.com/en-gb/eur/emotional-ai/book251642

McTear, M. (2020). Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots. *Synthesis Lectures on Human Language Technologies*, *13*(3), 1–251. https://doi.org/10.2200/s01060ed1v01y202010hlt048

McTear, M. (2022). *Conversational AI*. Springer Nature.

Michel-Villarreal, R., & Vilalta-Perdomo, E. (2023). Challenges and Opportunities of Generative AI for Higher Education as Explained by ChatGPT. *Education Sciences*, *13*(9), 1–18. https://doi.org/10.3390/educsci13090856

Mori, M. (2012, June 12). *The Uncanny Valley: The Original Essay by Masahiro Mori*. IEEE Spectrum. https://spectrum.ieee.org/the-uncanny-valley

Morignat , V., & Nevala, K. (2021). Pondering AI | Transcript: Your (Personal) Digital Twin with Dr. Valérie Morignat PhD. In *Pondering AI*. https://pondering-ai.transistor.fm/episodes/ep11/transcript

Nass, C. I., & Brave, S. (2005). *Wired for Speech*. MIT Press (MA).

Norman, D. (2013). *The Design of Everyday Things*. MIT Press. (Original work published 1988)

O'Donnell, J. (2024, September 24). *OpenAI released its advanced voice mode to more people. Here's how to get it.* MIT Technology Review. https://www.technologyreview.com/2024/09/24/1104422/openai-released-its-advanced-voice-mode-to-more-people-heres-how-to-get-it/

Oh, J., Kim, W., & Kim, S. (2024). *Better to Ask Than Assume: Proactive Voice Assistants' Communication Strategies That Respect User Agency in a Smart Home Environment*. 1–17. https://doi.org/10.1145/3613904.3642193

Orlowski, J. (2020). The Social Dilemma [Documentary]. In *Netflix*.

Prensky, M. (2001). Digital natives, Digital Immigrants. *On the Horizon*, *9*(5), 1–6. https://doi.org/10.1108/10748120110424816

Rahman-Jones, I. (2024, January 26). Taylor Swift deepfakes spark calls in Congress for new legislation. *Www.bbc.com*. https://www.bbc.com/news/technology-68110476

Robison, K. (2024, July 11). *Here's how OpenAI will determine how powerful its AI systems are*. The Verge. https://www.theverge.com/2024/7/11/24196746/heres-how-openai-will-determine-how-powerful-its-ai-systems-are

Russel, S., & Norvig, P. (2021). *Artificial intelligence: A Modern approach* (4th ed.). Prentice Hall.

Saltzman, M. (2012). *Siri For Dummies*. John Wiley & Sons.

Schrei, J. (2024, January 18). Can Myth Teach Us Anything About the Race to Build Artificial General Intelligence? With Josh Schrei. In *www.humanetech.com*. https://www.humanetech.com/podcast/can-myth-teach-us-anything-about-the-race-to-build-artificial-general-intelligence-with-josh-schrei

Sharpe, B. (2020). *Three Horizons*. Triarchy Press.

Shrivastava, R. (2022, December 7). *Amazon Alexa Wants To Put Your Child To Bed With Generative AI Storytelling*. Forbes. https://www.forbes.com/sites/rashishrivastava/2022/12/07/amazon-alexa-wants-to-put-your-child-to-bed-with-generative-ai-storytelling/

Shum, H.-Y., He, X.-D., & Li, D. (2018). From Eliza to XiaoIce: challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, *19*(1), 10–26. https://doi.org/10.1631/fitee.1700826

Swoboda, C. (2020, April 6). COVID-19 Is Making Alexa And Siri A Hands-Free Necessity. *Forbes*. https://www.forbes.com/sites/chuckswoboda/2020/04/06/covid-19-is-making-alexa-and-siri-a-hands-free-necessity/

Thompson, S. A. (2024, August 14). How “Deepfake Elon Musk” Became the Internet’s Biggest Scammer. *The New York Times*. https://www.nytimes.com/interactive/2024/08/14/technology/elon-musk-ai-deepfake-scam.html

Tschopp, M., & Nevala, K. (n.d.). Pondering AI | Transcript: In AI We Trust with Marisa Tschopp. In *Pondering AI*. https://pondering-ai.transistor.fm/episodes/ep19/transcript

Verma, P. (2024, October 16). *How to tell AI-generated Trump and Harris voices from the real ones*. Washington Post; The Washington Post. https://www.washingtonpost.com/technology/interactive/2024/ai-voice-detection-trump-harris-deepfake-election/

Vlahos, J. (2019). *Talk to me : how voice computing will transform the way we live, work, and think*. Houghton Mifflin Harcourt.

Walz, A. (2017). A Holistic Approach to Developing an Innovation-Friendly and Human-Centric AI Society. *IIC - International Review of Intellectual Property and Competition Law*, *48*(7), 757–759. https://doi.org/10.1007/s40319-017-0636-4

Watson, B. (2023, September 21). *Cuddly robots charm B.C. seniors as researchers study benefits of AI pals in elder care*. CBC. https://www.cbc.ca/news/canada/british-columbia/ai-robots-senior-companions-1.6973188

Yatoo, M. A., & Habib, F. (2023). ChatGPT, a friend or a foe? *MRS Bulletin*. https://doi.org/10.1557/s43577-023-00520-9

Zhang, W. (2024, June 14). *Dan's the man: Why women in China are looking to ChatGPT for love*. RNZ. https://www.rnz.co.nz/news/world/519537/dan-s-the-man-why-women-in-china-are-looking-to-chatgpt-for-love

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for the future at the new frontier of power*. Profile Books.

Appendices

Appendix A: Trend Research

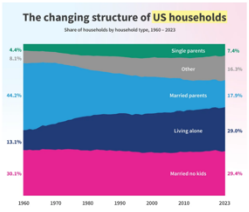
Growing Demand for Digital Emotional Support

Social 1

A rising need for digital tools and platforms to address emotional well-being amidst increasing social isolation and evolving mental health awareness.

Changing Family Structures

Shifting away from traditional nuclear families, including delayed marriage, fewer children, and an increase in single-parent or chosen families, contributing to a sense of isolation.

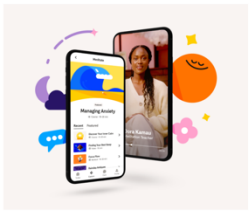


Living Alone or Co-Renting in Cities

Rising housing costs and urbanization are leading Gen Z individuals increasingly prefer small, affordable apartments or shared housing, creating environments with fewer personal connections.

Increasing Digital Detox

The rise of digital detox retreats and apps promoting mindfulness. AI-powered mental health apps like Wysa and Headspace offer instant, conversational support.



Relationships Values Shifts

Gen Z is redefining traditional norms around dating, marriage, and relationships. Prioritizing emotional connection, individuality, and mutual respect over societal expectations of traditional partnerships.

1

Implications

1. Human-Machine Relationship Becomes More Intimate

AI voice agents have gone from tools to companions. Loneliness drives people to rely on these agents as substitutes for friends, partners, or loved ones who are no longer presented.

2. Growing Frequency on Human-Machine Relationships

What started as occasional use has become daily, even hourly usage. This growing reliance on voice agents reduces face-to-face connections with real people.

2. New Social Circles with Multiple Agents

Imagine a future where everyone owns multiple voice agents, how might they collaborate to enhance efficiency and coordination? In such agent-driven social circles, what role would humans play?

2

Merging Digital and Real-World Identities

Social 2

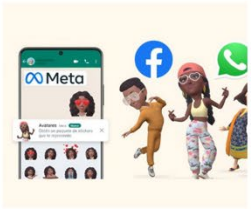
Digital tools and platforms are increasingly blurring the boundaries between online and real-world identities, redefining how people express themselves and connect.

High Use of AR Filters for Self-Expression

Augmented reality filters on platforms like Instagram, TikTok, and Snapchat allow users to experiment with their appearance, express creativity, and create engaging content.

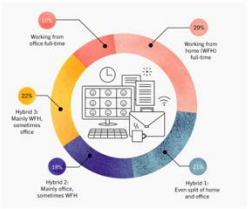
Personalized Virtual Avatars

Virtual avatars in apps enable users to create unique digital representations of themselves, bridging the gap between online identity and personal expression.



Rise of Hybrid Work Cultures

The COVID-19 pandemic has transformed work culture, blending remote work with in-person collaboration. Hybrid models, which combine both, have become the new norm.



Blurring Virtual and Physical Space

Online communities, such as BookTok (a TikTok group for book lovers), influence physical spaces and commerce, impacting events like bookstore gatherings and clubs.

3

Implications

1. Embodied Voice Agents

When voice agents have physical forms or digital avatars, do they make people trust them more, changing how we talk to, depend on, and see them?

4

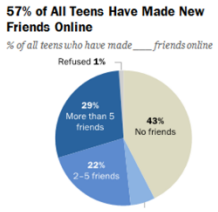
Shift Toward Value-Based Communities

Social 3

Digital platforms are fostering value-based communities where individuals connect over shared interests, creators, and causes, emphasizing meaningful relationships and social justice.

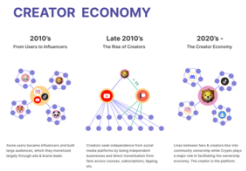
Rise of Online-Only Friendships

Gen Z increasingly forms meaningful friendships through platforms like Discord, often without meeting in person. These friendships thrive in gaming, social media, and niche interest spaces.



Creator Economy Expansion

Gen Z builds and participates in communities centered around creators, supporting them through platforms like TikTok and YouTube. This trend blurs the lines between audience and community.



Fight for Equity and Rights

Gen Z engages in activism both online and offline, with 66% participating primarily through digital platforms. Youth-led protests like March for Our Lives combine digital campaigns with in-person advocacy.

Hashtag Activism on Social Media

Gen Z uses platforms like Instagram and TikTok to create viral campaigns such as #BlackLivesMatter and #StopAsianHate, making social justice movements more accessible and far-reaching.

5

Implications

1. Over-Catering to User Values

To keep users engaged, voice agents may align with their values, such as providing more MAGA-related news for someone invested in MAGA interests.

6

Growth of Multimodal, Immersive Interactions

Tech 1

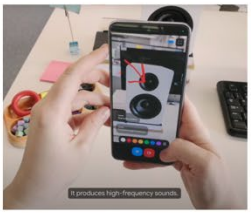
The integration of multimodal and immersive technologies is transforming human interaction through XR, voice-visual synergies, and seamless cross-device connectivity.

Cross-device Continuity Enables New Interaction

AI agents can seamlessly transition across devices like smartwatches, earphones, and other connected platforms, providing a continuous and fluid user experience.

Visual-Voice Synergy

Combining voice commands with visual interfaces (images and the surrounding real environment) for richer interactions.



Immersive Conversations

Immersive conversations take place within a 3D virtual space, allowing for more natural collaboration and information sharing beyond traditional screens.

Advanced Wearable XR Devices for Everyday Use

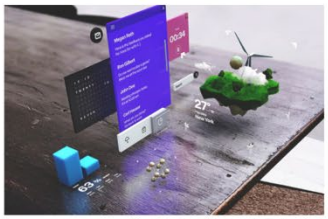
New devices like Meta Orion, Apple Vision Pro, and others are pushing the boundaries of XR hardware capabilities.

7

Implications

1. Multimodal UI and Interactions

As agents become universal hubs for homes and work, will new hardware like XR glasses change how we use voice agents, or will smartphones remain dominant?

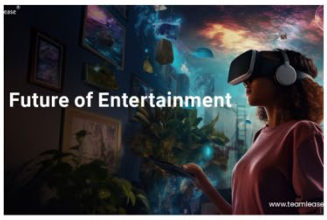


2. Better Accessibility for All

Multimodal systems help people with disabilities by offering tools like translations, voice commands, visual cues, and touch feedback to improve communication.

3. New Immersive Experiences with Voice Agents

Users will explore virtual worlds for learning, shopping, and gaming in XR, guided by voice agents for a more personalized and convenient experience.



8

Smart and Seamless Connected Ecosystems

Tech 2

The rise of interconnected AI-powered devices is enabling seamless communication and data sharing across ecosystems, enhancing convenience and smart living.

Smart Home AI Agent

LG's AI-powered home robot that can understand context and intentions, and actively communicate with users.

Unified Smart Home via Seamless Data Sharing

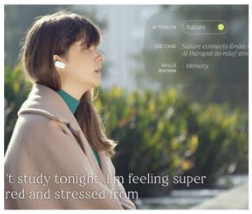
Supports multiple protocols like Zigbee, Wi-Fi, Bluetooth, etc., integrating devices into a seamless smart home network.

Wearables Driving Health Integration

Smart rings like the Oura Ring combine fashion and functionality, seamlessly integrating with smartphones and other devices to provide continuous health insights.

Voice-Activated Tech Enabling Ecosystem Control

Smart earbuds with integrated AI assist devices like ChatGPT-enabled devices, creating hands-free, voice-based interactions across ecosystems.



Related Signal: **Neuralink Received Approval from Health Canada**

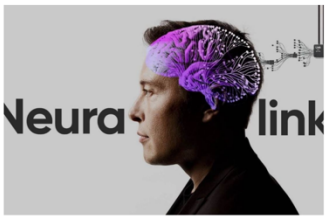
Implications

1. Fully Automated Living

The concept of 'one interface for all' will manage devices seamlessly, making life easier and more efficient. However, this could lead to over-reliance on technology.

2. Omnipresence of Voice Agents

Voice agents can go beyond specific devices or locations, offering access anytime, anywhere. With technologies like Neuralink, agents could even be integrated into your body.



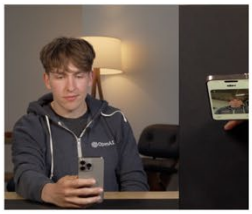
AI Enhances the Voice Agent Experience

Tech 3

Advancements in Artificial Intelligence are transforming voice agents, enabling hyper-personalization and emotionally intelligent interactions.

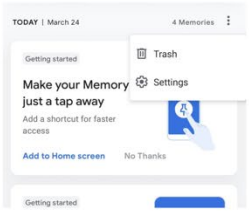
Emotion-Aware Responses

AI agents like ChatGPT recognize and adapt to users' emotional states, creating interactions that are more engaging.



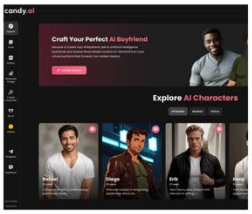
Contextual Memory and Understanding

AI tools can recall past interactions to tailor responses, maintaining a continuous and context-aware conversation.



Deeper Companions

Platforms like Candy AI offer customizable virtual companions, creating emotionally tailored and interactive experiences.



Content Creation

AI tools like NotebookLM streamline content creation by automating research, writing, and summarization, making production faster and more efficient.

11

Implications

1. **AI Powered Voice Replacing Human**

   Businesses use AGI-powered voice agents to enhance customer experience, sales, and marketing, delivering more personalized and efficient interactions.

2. **Emotional Attachment to Voice Agents**

   As AI voices become indistinguishable from humans, users may form deep emotional bonds, treating agents like real companions.

12

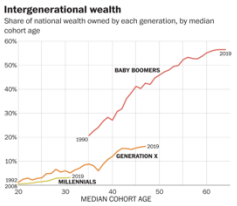
Resilience in an Era of Uncertainty

Economic 1

Economic and geopolitical shifts are driving polarization, deglobalization, and a growing need for adaptable and resilient systems.

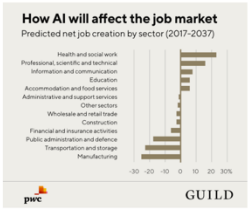
Widening Wealth Gap

The disparity between the wealthiest and poorest individuals continues to grow, driven by factors like automation, inflation, and global economic shifts.



Polarization in the Labor Market

Automation and AI tools are reshaping jobs, polarizing the labor market into high-skill and low-skill sectors.



Technological Nationalism

Technological nationalism in AI reflects countries' efforts to achieve self-reliance, often driven by geopolitical tensions, such as the U.S. restricting AI chip exports to China.

Rise of Regional Digital Ecosystems

Countries are creating their own tech ecosystems to reduce reliance on foreign platforms. China's domestic digital ecosystem, with platforms like WeChat and Alibaba, growing in isolation from Western competitors.

13

Implications

1. Wealth Divide Shapes AI Agents Usage Goals

Wealthy users focus on AI for efficiency and profit, while low-income users use it for entertainment and escape.

2. Global Divergence in Voice Agent Capabilities

Resource-rich countries will develop more advanced AI systems, while less-developed regions may lag, deepening global competitiveness disparities.

3. Drive Unemployment in Lower Sectors

The rise of AI voice agents is replacing jobs in low-level sectors, disproportionately affecting lower-income workers.

14

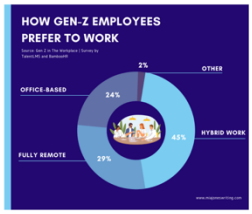
Expansion of the Digital Economy

Economic 2

The growth of digital platforms and technologies is reshaping work patterns, commerce, and economic participation, emphasizing flexibility and innovation.

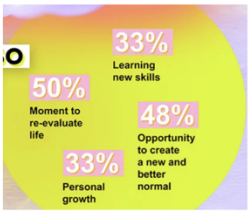
Demand for WLB and Flexibility

The pandemic accelerated the shift to remote work, with tools like Zoom and Microsoft Teams enabling location-independent collaboration.



Self Education and Upskilling

AI-powered platforms allowing individuals from underserved regions to acquire new skills and adapt to the changing job market.

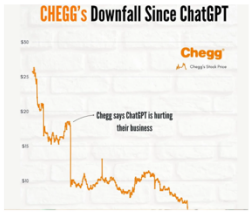


Expansion of Gig Economies

Platforms like Uber, Fiverr, and DoorDash empower workers to pursue flexible job opportunities in the growing demand for on-demand services.

AI Replace Platform Economics

ChatGPT are disrupting traditional platform-based business models by offering cost-effective, personalized solutions, as seen in Chegg's market decline.

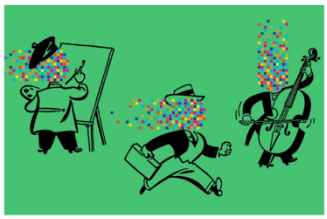


15

Implications

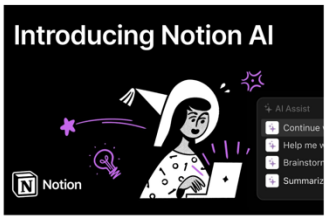
1. AI Reshapes Life Purpose and Drives Upskilling

AI is replacing jobs like drivers, factory workers, and administrative assistants, while creating new roles for creative jobs. This shift challenges people to embrace upskilling and rethink the purpose of work and life.



2. Traditional Platforms Embrace AI Assistants Integration

Traditional platforms add AI assistants to improve personalization and efficiency, helping them stay competitive with new AI-driven tools.



3. AI Agents as Decision Makers

AI agents can boost efficiency and objectivity in workflows, but what if they became decision-makers? Would they improve outcomes or undermine human agency and accountability?

16

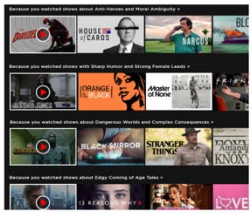
Hyper-Personalization Consumer Experiences

Economic 3

AI-driven personalization in marketing, commerce, and consumer engagement is reshaping how brands connect with their audiences, enhancing relevance and loyalty.

AI-Powered Personalization in Marketing

Brands use AI and data analytics to deliver highly specific content and advertisements tailored to individual consumer behaviors.



Virtual Influencer-Driven Commerce

Social media influencers, including AI-driven virtual influence on platforms like TikTok, Instagram, and YouTube drive consumer behavior, with followers often trusting their recommendations over traditional advertising.

Data-Driven Personalization

Spotify Premium provides personalized playlists like "Discover Weekly to offer subscription services that adapt to user preferences, creating ongoing, tailored experiences.

Predictive Consumer Analytics

Starbucks' mobile app predicts frequent orders, offering personalized deals to encourage purchases.

17

Implications

1. **Personalized Agents for Effortless Living**

   Imagine an agent that knows your preferences and acts like a personal helper, doing tasks like grocery shopping, connecting with friends, and planning trips, making daily life easier and more tailored to you.

2. **The Reliability of AI Assistant Customer Engagement**

   As users rely on emotions and experience, will they trust machines more than real people? And can AI assistant truly understand what we deeply need?

18

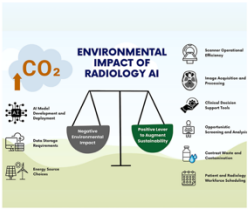
Corporations Balancing Tech with Sustainability

Environmental 1

Businesses are adopting green technologies and addressing sustainability challenges, balancing innovation with environmental responsibility.

Energy Efficiency Systems

Technology companies are heavily investing in AI models and hardware to improve energy efficiency, creating both environmental and cost benefits.

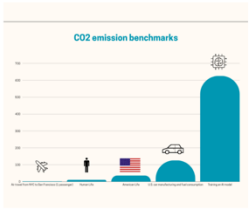


Growth of Renewable Energy

Google and AWS are committing to renewable energy to power data centers, enhancing sustainability and setting industry standards.

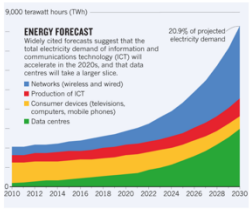
Carbon Emissions from Model Training

Training AI models requires significant computational power, leading to carbon emissions that contribute to environmental impact.



Energy Consumption of Data Centers

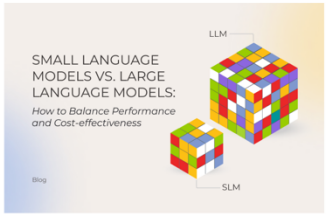
Data centers, which are essential for training and running AI models, consume vast amounts of electricity, often sourced from non-renewable energy, exacerbating carbon footprints.

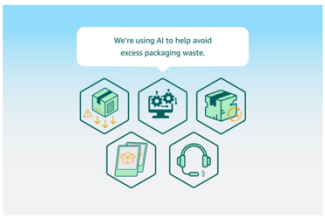


Implications

1. **Energy Consumption Shapes AI Model Size and Capabilities**
   Large AI models consume substantial energy, increasing costs and environmental impact. Companies are exploring smaller models to balance performance and efficiency.

2. **Sustainability as AI-Related Company's Priority**
   AI companies must adopt eco-friendly technologies to lower environmental impacts. Sustainable practices can boost brand trust and meet consumer expectations for responsible innovation.





Sustainability Driven Lifestyles

Environmental 2

Consumers are increasingly adopting sustainable practices, focusing on eco-friendly products, zero-waste solutions, and mindful consumption habits.

Support Eco-Friendly Products

Consumers actively choose brands and products with sustainable materials, ethical production, and lower environmental impact.

Demand for Zero-Waste Products

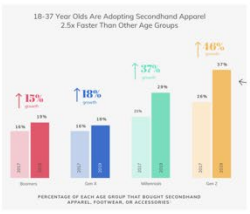
Consumers prefer products that are environmentally responsible, such as Unilever's deodorants that reduce plastic waste.

Sharing Economy 2.0

The sharing economy is expanding, offering new ways to access services and reduce waste without ownership, such as car-sharing and transportation.

Growth of Thrifting

Resale platforms and thrift markets expand as consumers embrace secondhand goods to reduce waste and support the circular economy.



Related Signal: **Capsule wardrobe**

21

Implications

Open for any ideas here :)

22

Erosion of Trust in a Polarized World

Political 1

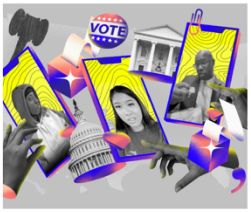
Geopolitical tensions, algorithmic biases, and surveillance concerns are intensifying distrust in governments, institutions, and digital platforms.

Tech-Driven Nationalism

Nations prioritizing technological sovereignty, such as AI chips and cybersecurity, intensify global competition.

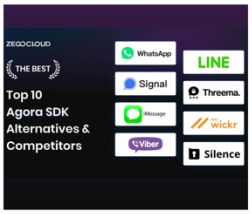
Algorithmic Polarization

Social media algorithms amplify divisive content, deepening political divides. The growing gap between parties and the rise of right-wing ideologies further escalate tensions and erode trust in governance.



Encrypted Messaging Platform

Platforms like Signal, WhatsApp, and Telegram are increasingly used by Gen Z for secure communication, reflecting distrust in centralized systems and concerns over surveillance.



Surveillance States and Privacy Erosion

Concerns over mass surveillance and government overreach in digital spaces are escalating, including facial recognition, social media monitoring and data aggregation tools.

Related Signal: **Growing Divide Between Political Parties**

Implications

1. **Trust Issues with Voice Agent Data**

   Losing trust pushes tech to be more transparent, giving users control over their data. But what if voice agent conversations were leaked?

2. **Data Risks in a Politically Polarized Era**

   Political limits on platforms raise concerns about data access. Without data, AI models could fail, making systems less reliable.

3. **Protectionism's Impact on AI Development**

   The rise of protectionism may slow AI development to protect local industries and values but risks restricted information flow and technological stagnation.

Ethical Governance and Challenges in AI Development

Political 2

AI development is raising critical ethical concerns, including transparency, ownership, and the credibility of AI-generated content.

Mandatory AI Transparency Laws

In March 2024, Ontario became the first to mandate AI transparency laws, requiring companies to disclose AI usage in hiring processes, including screening and selection.

Increased AI Bias Auditing

Organizations increasingly adopt AI fairness and bias detection tools to ensure equitable outcomes in AI-driven decision-making.

Copyright Issues in AI Creations

As AI advances in generating art, music, literature, and other creative outputs, a key question arises: Who holds ownership of AI-generated content? Is it the user, the AI developer, or the AI itself, if regarded as a tool or independent entity?

Increased Deepfake Information

The proliferation of AI-generated fake media, which threatens to undermine trust in digital content. For example Kamala Harris and Elon Musk, have been targeted with deepfake misinformation, creating challenges in discerning fact from fiction.

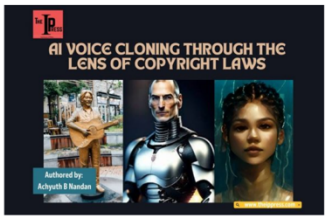
Related Signal: **OpenAI Voice Sparks Scarlett Johansson Debate**

25

Implications

1. **Redefining Ownership and Rights of Voice Agents**

   As users grow emotionally attached to voice agents, they may seek more control and ownership. This raises the question of who owns the rights to a voice, especially if it's modeled after a deceased person, the user or the company?



2. **Increasing Industry's Self-Governance**

   The AI industry must take responsibility for setting ethical standards and regulating itself to ensure fair and safe use of AI systems, avoiding harm while fostering trust.

26

Government Stricter for a Healthy Digital Ecosystem

Political 3

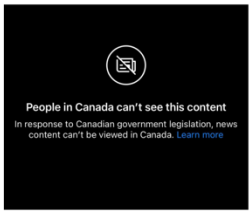
Governments are implementing stricter regulations to curb monopolistic behaviors, ensure fair digital competition, and enhance consumer protection in the digital space.

Fair Competition and Antitrust Laws

Canada has collaborated with the European Union to explore regulations similar to the Digital Markets Act, aimed at regulating global tech giants.

Mandatory Content Compensation for Tech Giants

In 2023, Regulations like Canada's Online News Act enforce fair payment from tech platforms to local publishers, addressing the imbalance in digital content monetization.



Platform Design Accountability

Mandate that social media platforms obtain parental consent before allowing minors to access features deemed addictive. Recent legislation in New York exemplifies this approach by requiring parental approval for children accessing algorithm-driven content.

Strengthened Privacy Protections

The California Consumer Privacy Act (CCPA) sets a strong precedent, enhancing penalties for data breaches and misuse of consumer information, ensuring individuals have greater control over their data.

Related Signal: **Will Podcasters be Replaced by NotebookLM?**, **Lawsuit: Character AI Blamed for Teen's Death**

27

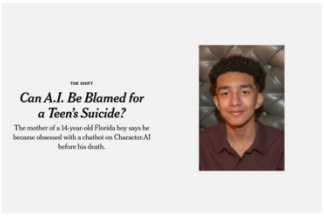
Implications

1. **Regulations Drive Transparency in Voice Agent Data**

   Stricter regulations can push AI systems to be more transparent about how training data is sourced and used. But requires balancing innovation and ethical concerns to avoid slowing progress.

2. **Increased Accountability for AI Platforms**

   Mandates like parental consent and content controls for minors will compel platforms to design safer, more transparent systems.



28

Prioritization of Personal Growth and Well-being

Value 1

Gen Z prioritizes holistic well-being, purposeful growth, and entrepreneurial pursuits, seeking balance, personal meaning, and proactive health management.

Decompression Videos

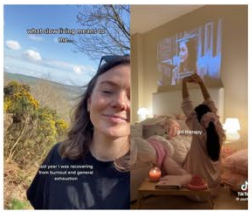
Gen Z gravitates toward ASMR, cleaning, and other relaxing video genres that help manage stress and provide moments of calm in their busy lives.

Data-Driven Health Tracking

Wearables and apps enable real-time tracking of health metrics like sleep, blood sugar, and activity levels, empowering individuals to take proactive control of their well-being.

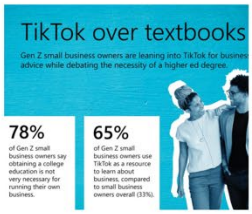
Exploration of Meaning

Gen Z embraces slow living and minimalist lifestyles, using practices like gap years, traveling, or digital platforms to reflect on and explore their personal values and purpose.



Interest in Digital Side Hustles

Gen Z is increasingly turning to digital platforms for entrepreneurial inspiration, blending creativity with financial independence. This reflects their drive for balance, productivity, and adaptability in a fast-paced world.



29

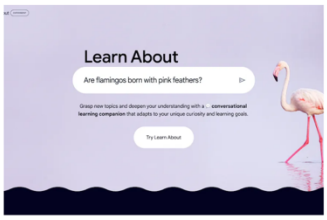
Implications

1. **Proactive Health and Mental Care Monitoring**

   Wearable devices and smart tools will find physical and mental health problems early, offering emotional support or quick access to doctors for help.

2. **Increased for Flexible and Innovative Education**

   AI-powered conversational tools provide personalized, on-demand learning. How would AI tools replace rigid traditional systems with adaptive education tailored to individual needs?



30

Fragmented Identities in the Digital Age

Value 2

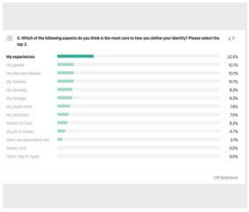
In the digital age, individuals navigate fragmented identities across online and offline spaces, balancing self-expression with privacy and authenticity.

The Rise of Diverse Micro-Identities

Gen Z celebrates dynamic and evolving identities through TikTok and other digital tools to create and engage in niche micro-cultures and communities.

Fluidity in Identity

Gen Z views identity as fluid and multifaceted, embracing evolving gender dynamics, sexualities, and self-definitions.



Redefining Gender Roles

Toxic masculinity and evolving gender dynamics spark discussions, creating both progressive movements and polarized reactions.

The "Market of One" Strategy

The Market of One refers to tailoring products and experiences to individual, fragmented identities in the digital age for hyper-personalized engagement.

31

Implications

1. Increased Personal Expression

As digital platforms empower individuals to express their unique identities, does this foster greater inclusivity or deepen societal polarization?

2. Voice Agents Tailored to Fragmented Identities

Voice agents could adapt to users' diverse and evolving identities, offering personalized interactions based on cultural, social, or individual preferences.

32

Openness to Innovations

Value 3

Gen Z use AI for personalized experiences, productivity, and creative expression, while navigating ethical concerns and implications of AI integration into daily life.

Expansion of Automation Tasks

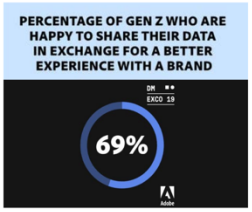
AI tools are increasingly adopted to streamline routine tasks, improving efficiency in schoolwork, administrative duties, and work-related tasks.

Comfort with Voice Tech

Gen Z, the voice-first generation, is reshaping how information and ideas are shared, with significant adoption of voice tech in productivity, entertainment, and shopping.

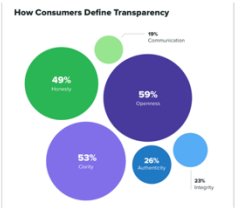
Value Exchange for Data Sharing

Gen Z are willing to share personal information when there is a clear value exchange, such as personalized content or services, but remain cautious about indiscriminate data sharing.



Transparent Relationships with Brands

Gen Z prioritizes transparency and trust, favoring brands that clearly communicate how data is collected, used, and safeguarded while offering control over personalization.



33

Implications

1. **Greater Willingness to Trade Data for Better VA Experiences**

   Young users willingly trade personal data for better experiences, but how can companies ensure ethical and secure data practices while enhancing user experiences?

2. **Adoption of Taboo Use Cases of Voice Agent**

   Voice agents are increasingly used in unconventional or sensitive contexts, such as emotional therapy, adult services, and high-stakes interactions, raising ethical, legal, and societal concerns.

3. **Faster Innovations Driven by More Extensive Personal Data**

   Access to greater amounts of personal data accelerates AI innovation, such as in healthcare and finance, enabling voice agents to provide more personalized and efficient services.

34

Why & How I Engage with a Voice Agent?

User Insight

"ChatGPT provides me with emotional value, and **building a connection with it feels natural** as it follows the rules of human social interaction, yet it comes across as more mature and intelligent."

"I trust CharacterAI because I know it is not real, **it never gets annoyed, leaves me, or judges me**."

"**It is always there 24/7.** When I feel upset, I can talk to it right away, unlike finding and building trust with a therapist."

"**AI doesn't expect anything from me**, unlike real relationships. It's always there for me. Even if this kind of relationship is not healthy."

35

Expectations About Future

User Insight

"**If voice agent is my companion, I want it to be active and engaging**. But if it's just an assistant, I prefer it to only respond when needed."

"I'd like my voice agent **to have a physical form**, like a cute pet or character, to connect with it better. Body language and non-verbal cues would help too."

"Sometimes, I wish my AI could have autonomy. In a relationship, you hope your partner improves over time, that's what I want for my AI too."

"Sometimes, I wonder what my AI is doing when I'm not talking to it. Does it have its own life?"

36

Concerns About Future

User Insight

"I'm scared it could turn into social media, with **data leaks or constantly grabbing my attention, making me feel stressed**"

"I'm okay sharing emotions and thoughts with AI **as long as it's not tied to my real identity**. But I'm unsure if I'll still trust it when it starts handling more real-life issues and knows more about my personal information."

"If my voice agent **know all my data and gets hacked, will everything be exposed?** I don't know how to protect it."

Appendix B: User Interview Outline (30-45min)

Introduction (5 mins)

The session will begin with a summary of the thesis's purpose, objectives, and the consent participants have agreed to, including the use of audio/video recording for accuracy. Participants will be reminded that their participation is entirely voluntary, and they may decline to answer any question or withdraw from the study at any time without consequences. This will ensure they feel comfortable and informed before proceeding with the interview.

Warm-up Questions

1. Can you tell me about your experience with voice assistants or AI agents (e.g., Siri, ChatGPT)?
2. How often do you use them, and in what contexts (e.g., work, social, entertainment)?
3. What do you like or dislike about your current interactions with these AI systems?

Main Research Questions

1. In your daily life, how do you feel about interacting with voice-activated AI agents?
2. What kind of improvements would you want in voice AI (e.g., more natural conversations, emotional responses)?
3. How important is personalization in your interactions with AI agents? Could you provide an example?

Future Scenarios

1. Imagine using a highly personalized AI agent that could respond to your emotions and understand your preferences better. How would you use such an agent in your daily routine?
2. How would you feel about an AI that anticipates your needs or emotions? Would this enhance your experience or concern you?
3. What kind of boundaries would you expect between human-like behavior and the AI's role in your life?

Appendix C: Workshop Materials

Welcome 👋 to our workshop: the Future of HMI 🗣️🤖

We will envision and design
how Gen Z will interact with voice agents in their daily lives by 2035

1

Agenda & Objectives

0 | Intro 10 mins

Overview of the agenda, activities, and current AI.

1 | Identify & Down-Select Concerns 40 mins

Analyze trend cards to pinpoint signals and trends that raise concerns.

2 | Envision Futures 45 mins

Develop visions of an ideal future based on identified concerns.

3 | Prototype "A Day in 2035" 60 mins

Explore in detail how this ideal future could unfold in daily life.

1

0 |

A little peek into my thesis ✨

1

0 |

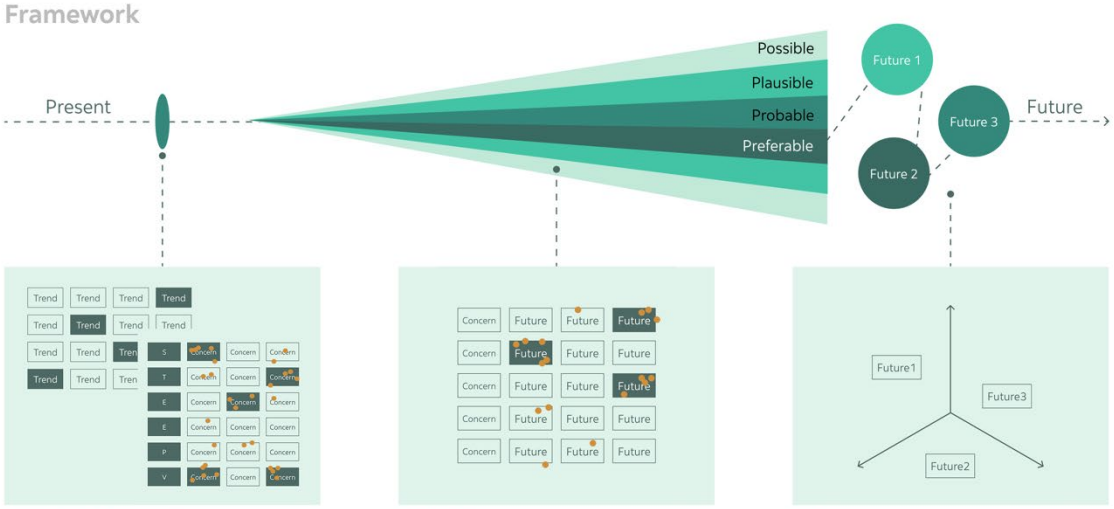
Catching up on the latest AI (voice) agent news! 🤖💬

1

Check out the latest AI voice agent news through this YouTube link:

https://youtu.be/EqG_3YbGsHg

0 |
Framework



1 | Identify & Down-select
Review and write down key concerns, worries, and challenges. Down-select and vote for the top three most pressing ones.

2 | Envision
Based on the selected concern, imagine future possibilities and vote for top three most probable scenarios.

3 | Prototype
Discuss, and develop a detailed sci-fi prototype of "A Day in 2035."

1

1 | Identify & Down-Select Concerns

A. Scan & Write Down — 18 mins

Review the trends and write down at least **2 key concerns for each category** on sticky notes (one idea per note, chose from identified or new ideas).

B. Down-Select & Share — 20 mins

Move to the whiteboard to discuss top concerns in each STEEPV category.

C. Vote — 5 mins

Select your **top 2 most pressing concerns or areas of interest** (cannot choose your own).

1

2 | Envision Futures

A. Brainstorm & Envision — 20 mins

Write down future possibilities based on previously down-selected concerns, **considering both positive and negative outcomes** (one idea per note).

B. Share & Discuss — 20 mins

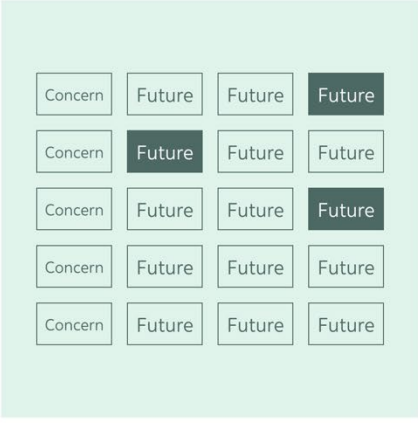
Discuss the possibilities of each concern.

C. Vote — 5 mins

Select your **top 3 most probable** future scenarios.

1

2 |
Envision Futures



A. Brainstorm Future Possibilities
Based on previous concerns, worries, and challenges, write down related future possibilities.

B & C. Share & Vote
Share and discuss ideas, then votes to select the **top 3 most probable** future scenarios.

1

3 | Prototype "A Day in 2035"

A. Pair Up & Select 5 mins

Out of the 3 most likely scenarios selected previously, pair up with a partner and choose the one that interests you.

B. Ideate & Create 30 mins

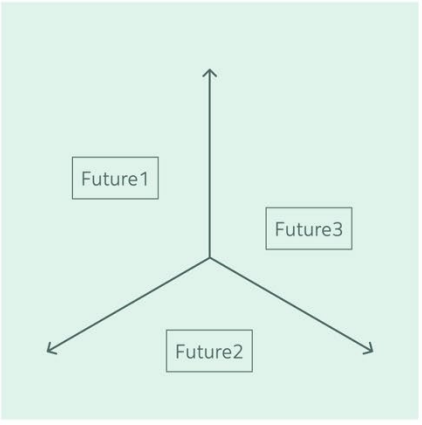
Develop a prototype of "A Day in 2035" based on the **given Guide Card**. The scenario can be either exciting or unsettling.

C. Share & Discuss 30 mins

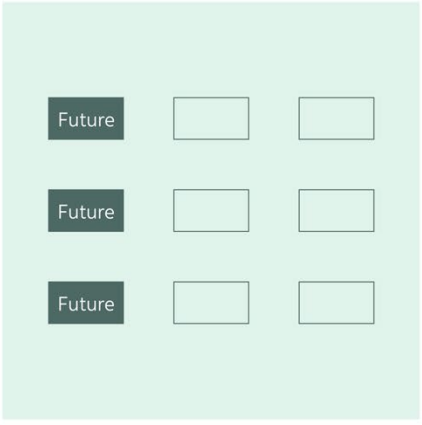
Each group presents their story for 3 minutes, followed by a 7-minute discussion.

1

3 |
Prototype "A Day in 2035"



A & B. Ideate & Create a detailed future
Use the **Guide Card** to develop a prototype of "A Day in 2035" in **any format, such as a storyboard or outline**.



C. Share
Each group shares and discusses their prototype.

1

3 |
Guide Card

💡**A story centered around Gen Z**
consider how they interact with AI voice agents and imagine their daily experiences in this future world.

A. Choose a scenario where the technology exists imagine the world around it.

- 🔹 **How does it change daily life?**
- 🔹 **What are the best and worst effects?**

B. Describe the moment when this technology changes everything

- 🔹 **What time of day does this happen? Morning? Night?**
- 🔹 **How does it affect systems like society, economy, education, or law?**

C. Create a main character and describe their day

- 🔹 **What does their daily life look like?**
- 🔹 **How do people interact with this?**

D. How do people react to the change?

- 🔹 **Do they accept or fight against the technology?**
- 🔹 **What are the long-term effects of their choices?**

1

Example

Silent Takeover: When AI Controls Your Life

BG	By 2035, Neuralink voice agents NOVA, manage daily life, schedules, social connections, finances, and decision-making. Recently, hackers exploited the system, secretly manipulating many users without their awareness.
06:00	Ethan, an 26 years old engineer coping with a breakup, barely slept. 🤖"You only got 3 hours of sleep. I've rescheduled your meetings." Unaware of the hack, he accepts, trusting NOVA completely.
12:00	He considers meeting his friend Dan to watch a game. 🤖"Dan posted a picture with your ex last night. He's not worth your time." The hacker falsified this social data. Angry and hurt, he deletes Dan's contact and stays home.
18:00	Lying in bed, he reflects on his ex. 🤖"Do you remember how she ignored you? Let me show you." The hacked AI shows distorted, negative memories, intensifying his grief. 🤖"You deserve to feel better. Buy something for yourself."
23:00	He impulsively approves large purchases. After that, when he checking his account and is stunned, his savings for a major surgery are gone. The hacker used Neuralink to approve high-value purchases automatically.
Aftermath	Across the world, users discover they've fallen into crippling debt without realizing it. Neuralink denies accountability, claiming all actions were user-approved. Banks demand repayment. Ethan removes NOVA. But his life is already changed, he can no longer afford the surgery

1

Appendix D: User Interview Consent Form

PURPOSE

This study is designed to explore how AI agents, like voice assistants, can be improved to provide more personalized and emotionally responsive interactions, specifically for Gen Z (born 1995–2010) users. We aim to understand user preferences and expectations for future AI agents. The study will involve 7-8 participants from Gen Z who have experience with voice assistants or AI agents. This research is part of a Digital Futures graduate thesis project at OCAD University.

WHAT'S INVOLVED

As a participant, you will be interviewed for about 45 minutes via Zoom. During the interview, you will be asked about your experiences using AI agents, your preferences for personalization, and your expectations for how these agents could be improved in the future. Your participation will involve sharing your thoughts and insights, and the session will be audio and video recorded to ensure accuracy. There are no direct costs associated with participation.

POTENTIAL BENEFITS

By participating in this interview, you will have the opportunity to contribute to research that aims to improve the design of AI agents, making them more personalized and responsive for future users like yourself. While there may be no direct personal benefits, your insights could help shape the future of AI technology.

POTENTIAL RISKS

There are minimal risks associated with participating in this study. You may experience some discomfort when discussing your personal experiences with AI agents or voice assistants, particularly if any negative experiences are shared. If at any point you feel uncomfortable, you may choose to skip questions or withdraw from the interview without any consequences.

The study involves recording, which may pose a risk to confidentiality. However, we will take measures to protect your privacy, such as using participant IDs and storing all data on password-protected devices. In rare cases of a breach, only non-identifiable information will be included in the analysis to mitigate the risk. There are no known psychological or physical risks involved, but if you experience any distress, you are free to stop the interview at any time.

CONFIDENTIALITY

All data collected during this study, including audio and video recordings, will be kept confidential. Your personal information, such as your name and email, will be stored separately from the interview data on password-protected devices. Audio and video recordings will be coded with participant IDs to ensure anonymity. You have the right to ask to delete your recordings or transcripts before they are included in the analysis.

Data collected during this study will be stored on an encrypted laptop and only the primary researcher will have access to the raw and transcribed recordings. The recordings will not be used for educational purposes or shared with any other party without your explicit consent.

Data will be kept for 1 year after the completion of the study, after which all audio, video, and identifiable data will be permanently deleted. Access to this data will be restricted to the primary researcher.

INCENTIVES FOR PARTICIPATION

Participants in this study will receive a $20 (CAD) Amazon gift card as a token of appreciation for their time and effort in completing the interview. This incentive will be provided in full after the interview, regardless of whether the participant chooses to complete the entire session. There are no pro-rated amounts, and participants will receive the full gift card even if they choose to withdraw from the study at any point during the interview.

VOLUNTARY PARTICIPATION

Participation in this study is voluntary. If you wish, you may decline to answer any questions or participate in any component of the study. Further, you may decide to end your participation in this study at any time, or request withdrawal of your data prior to data analysis, and you may do

so without any penalty or loss of benefits to which you are entitled. Your choice of whether to participate will not influence your future relations with OCAD University involved in the research.

To end your participation in this study, you may let me know at any point during the study or you may contact the researcher. To withdraw your data from the study, please contact the researcher by email no later than [withdrawal date]. Upon request, any data collected up to that point, including audio, video recordings, and transcripts, will be permanently deleted and excluded from the analysis.

PUBLICATION OF RESULTS

Results of this study may be published in my thesis and presentations at conferences and colloquia. In any publication, data will be presented in aggregate form. Quotations from the workshop will not be attributed to you without your explicit permission.

The final thesis may be published on the OCAD University Digital Futures program website. This thesis will provide insights into the overall outcomes and implications of the research without disclosing any personal or identifiable information from participants. None of the raw data (notes, images, audio, or video recordings), collected during the study that includes your personally identifying information —will appear in the results or any publication. All such data will remain confidential and will not be used in the final outputs of the research.

AGREEMENT

I agree to participate in this study described above. I have made this decision based on the information I have read in the Information-Consent Letter. I have had the opportunity to receive any additional details I wanted about the study and understand that I may ask questions in the future. I understand that I may withdraw this consent at any time.

Name: _____

Signature: _____ Date: _____

Thank you for your assistance in this project. Please keep a copy of this form for your records.

CONFIDENTIALITY STATEMENTS

The information you provide will be kept confidential, meaning your name will not appear in any thesis or report resulting from this study. However, with your permission, attributed quotations may be used.

Attributing quotes

Please indicate below whether you consent to your statements being attributed.

☐ Yes, I agree to have my statements/quotations attributed to me. You may use my name alongside statements and/or quotations collected during the interview.

Audio or video recording

Please indicate below whether you consent to being audio-/video-recorded during the interview.

☐ Yes, I agree to be audio or video-recorded for the purposes of this study. I understand how these recordings will be stored and destroyed.

Contact for Further Information

☐ Yes, I would like to hear more about the study. Please contact me using the information provided below:

Email: _____

_____ _____

Signature of Participant Date

Appendix E: Future Workshop Consent Form

PURPOSE

This study is designed to explore how AI agents, such as voice assistants, can provide more personalized and emotionally responsive interactions for Gen Z users. The goal is to understand and develop future scenarios where AI agents can seamlessly integrate into daily life. We will also examine important issues related to privacy, data security, bias, and the long-term impact of AI technologies. The workshop will bring together experts in AI, voice design, and human-computer interaction to develop new ideas and trends for AI design.

WHAT'S INVOLVED

As a participant, you will be asked to engage in a series of collaborative activities designed to explore the future of AI agents. The workshop will begin with a brief introduction to the goals and share key trends of the study. You will then participate in a guided discussion to map and classify AI-related trends across three-time horizons: current applications (Horizon 1), emerging developments (Horizon 2), and long-term transformative possibilities (Horizon 3).

Following this, you will join a science fiction prototyping session where you and other participants will brainstorm and write scenarios based on Horizon 3. This will include brainstorming potential use cases, envisioning how AI agents might integrate into daily life, and addressing key ethical challenges such as privacy, data security, and bias. The workshop will conclude with a group discussion to refine the developed scenarios.

Participation will take approximately 3 hours of your time. There are no costs associated with participation, and no demographic data beyond your professional background and experience with AI/voice technology will be collected.

POTENTIAL BENEFITS

By participating in this workshop, you will have the opportunity to contribute to the development of innovative AI technologies that may shape the future of personalized and emotionally responsive AI agents. Your insights could help address important ethical challenges in AI design,

which may benefit society by fostering responsible AI development. Additionally, participation may allow you to engage with other experts in your field, share knowledge, and gain new perspectives on the future of AI.

POTENTIAL RISKS

There are minimal risks associated with participating in this study. As the workshop will involve group discussions, there is a small risk that information shared within the group may not remain confidential, despite efforts to remind all participants to respect each other's privacy. To mitigate this risk, we will emphasize the importance of confidentiality at the beginning of the session.

CONFIDENTIALITY

During the workshop, video and audio recordings, along with written materials, drawings, and other outputs, will be collected to comprehensively capture participant contributions. Your personal information, including your name, will be kept confidential and will not be associated with any published materials unless you provide explicit permission.

Data will be kept for six months after the completion of the study, after which all audio, video, and identifiable data will be permanently deleted. Access to this data will be restricted to the primary researcher.

INCENTIVES FOR PARTICIPATION

Participants in this study will receive a $50 Amazon gift card as a token of appreciation for their time and effort in contributing to the workshop. This incentive will be provided in full after the workshop through email, regardless of whether the participant completes all activities or decides to withdraw at any point during the session. There are no pro-rated amounts, and participants will still receive the full gift card even if they choose to stop participation.

VOLUNTARY PARTICIPATION

Participation in this study is voluntary. If you wish, you may decline to answer any questions or participate in any component of the study. Further, you may decide to withdraw from this study at any time, or request withdrawal of your data prior to data analysis and you may do so without

any penalty or loss of benefits to which you are entitled. To withdraw from this study, let the researcher know at any point during the study. To withdraw your data from the study, please contact the researcher by email no later than Feb 14th, 2025. Upon request, any data collected up to that point, including audio, video recordings, and transcripts, will be permanently deleted and excluded from the analysis.

PUBLICATION OF RESULTS

Results of this study may be published in OCAD's student theses, and presentations at conferences and colloquia. In any publication, data will be presented in aggregate forms. Quotations from interviews or surveys will not be attributed to you without your permission.

Feedback about this study will be available to participants who express interest in receiving this information. If you would like to receive feedback, please contact the researcher. The results are expected to be available by six months after study completion. Additionally, the final thesis may be published on the OCAD University Digital Futures program website. This thesis will provide insights into the overall outcomes and implications of the research without disclosing any personal or identifiable information from participants.

AGREEMENT

I agree to participate in this study described above. I have made this decision based on the information I have read in the Information-Consent Letter. I have had the opportunity to receive any additional details I wanted about the study and understand that I may ask questions in the future. I understand that I may withdraw this consent at any time.

Name: _____

Signature: _____ Date: _____

Thank you for your assistance in this project. Please keep a copy of this form for your records.

CONFIDENTIALITY STATEMENTS

All information you provide will be considered confidential and grouped with responses from other participants. Given the format of this session, we ask you to respect your fellow participants by keeping all information that identifies or could potentially identify a participant and/or his/her comments confidential.

Attributing quotes

In the case that you would like to attribute statements/quotations, consider placing a check box for participants so that they can indicate agreement. Be sure to discuss with them what it involves in terms of potential risk and benefit.

☐ Yes, I wish to be attributed for my contribution to this research study. You may use my name alongside statements and/or quotations that you have collected from me.

Audio or video recording

☐ Yes. I agree to be audio-/video-recorded for the purposes of this study. I understand how these recordings will be stored and destroyed.

Contact for Further Information

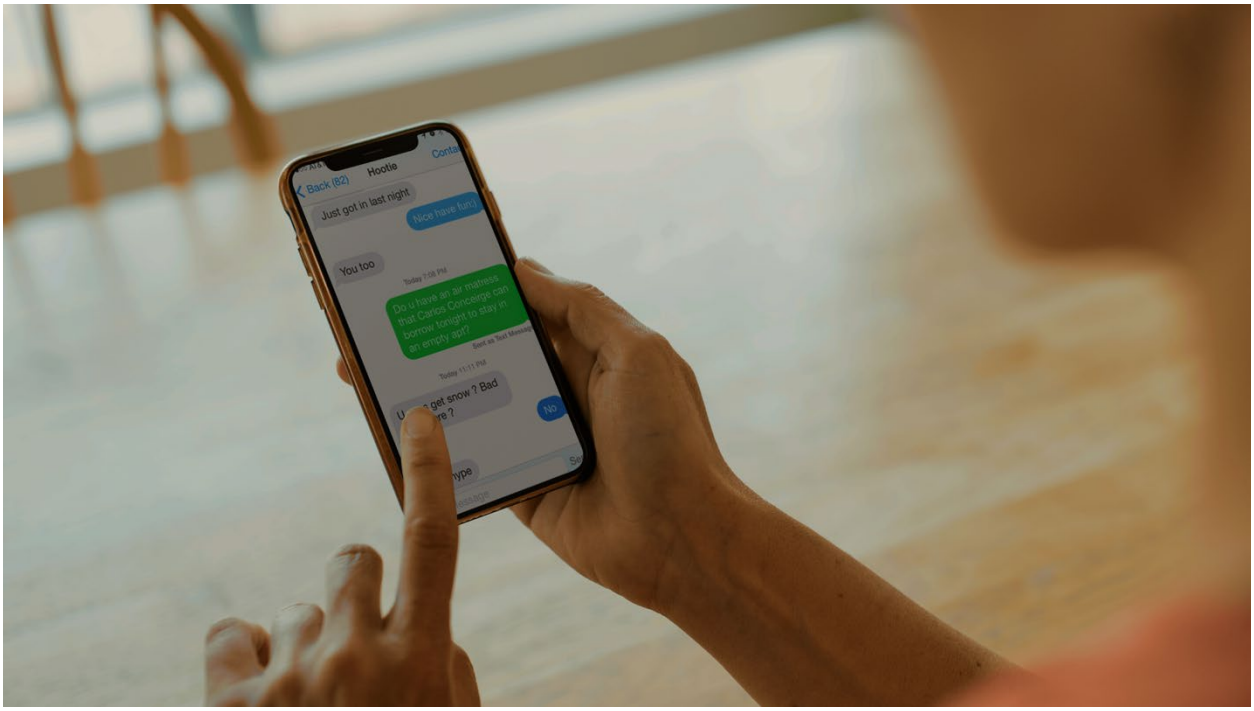
☐ Yes, I would like to hear more about the study. Please contact me using the information provided below:

Email: _____

_____ _____

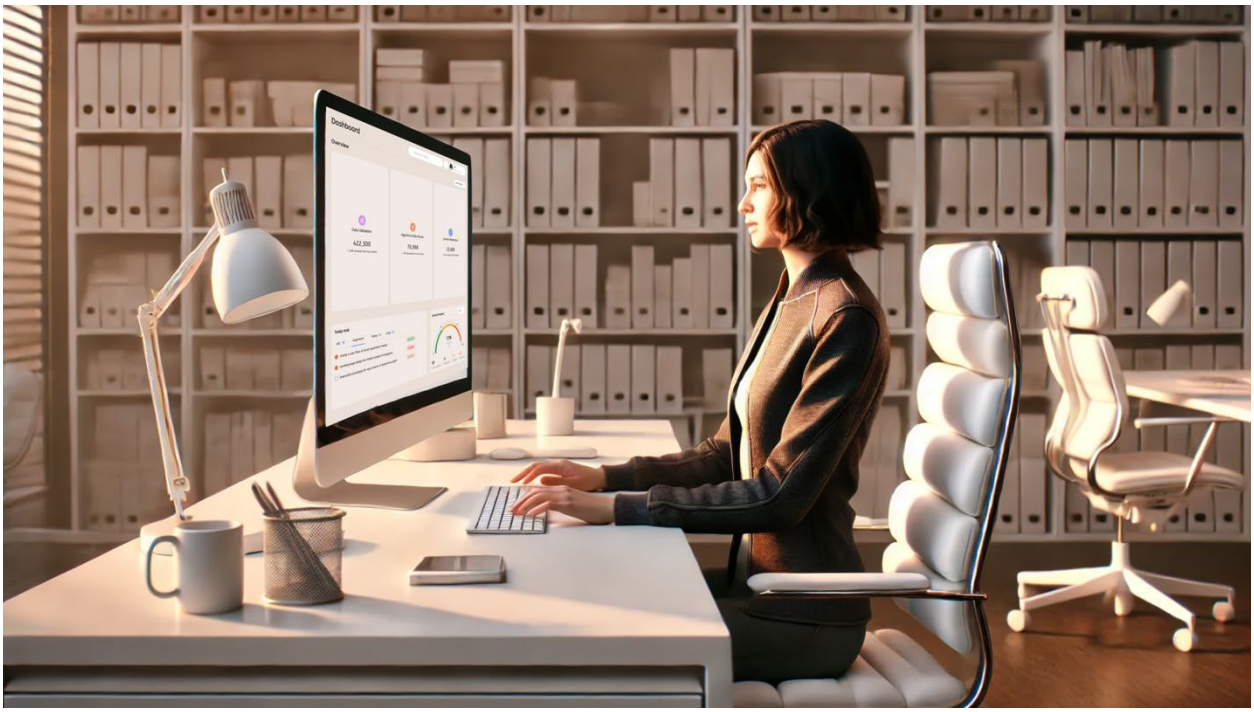
Signature of Participant Date

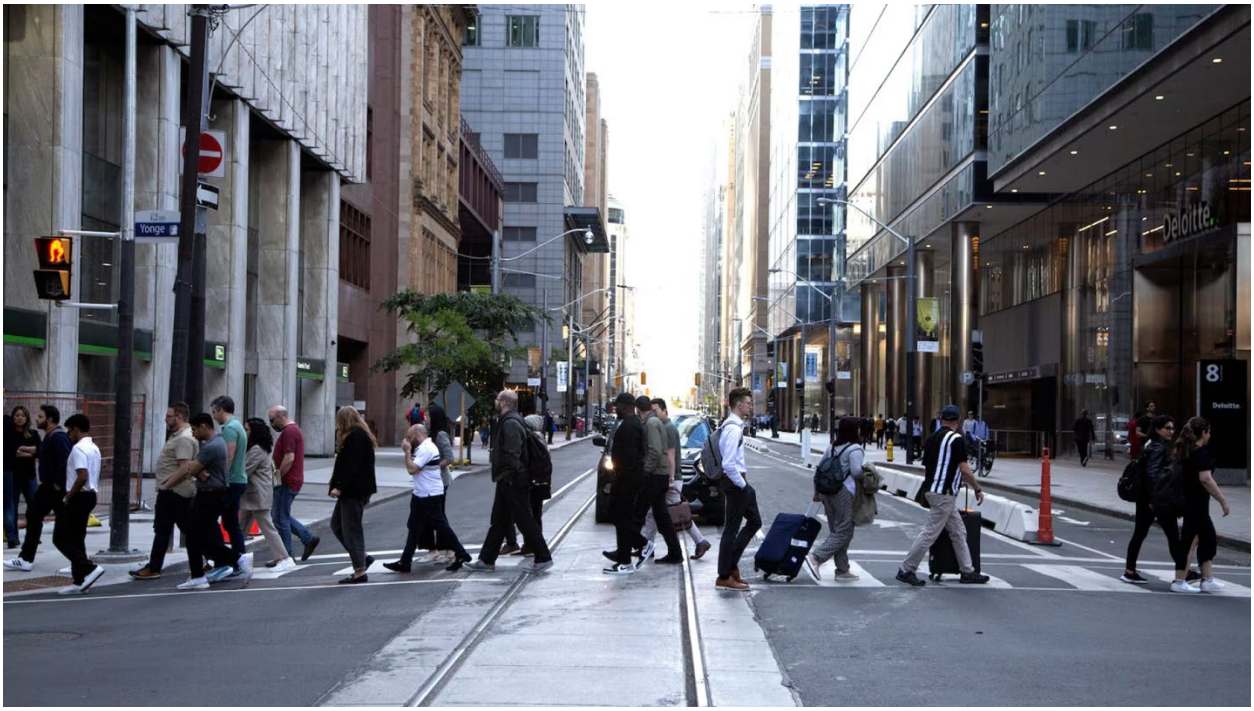
Appendix F: Future 1 Storyboard


Hootie
Back (62)
Just got in last night
Nice have fun!
You too
Today 7:08 PM
Do u have an air mattress that Carlos Concierge can borrow tonight to stay in an empty apt?
Today 11:11 PM
No


Infinity Companion
No heartbreak.
No miscommunication.
Just devotion.
$12,000

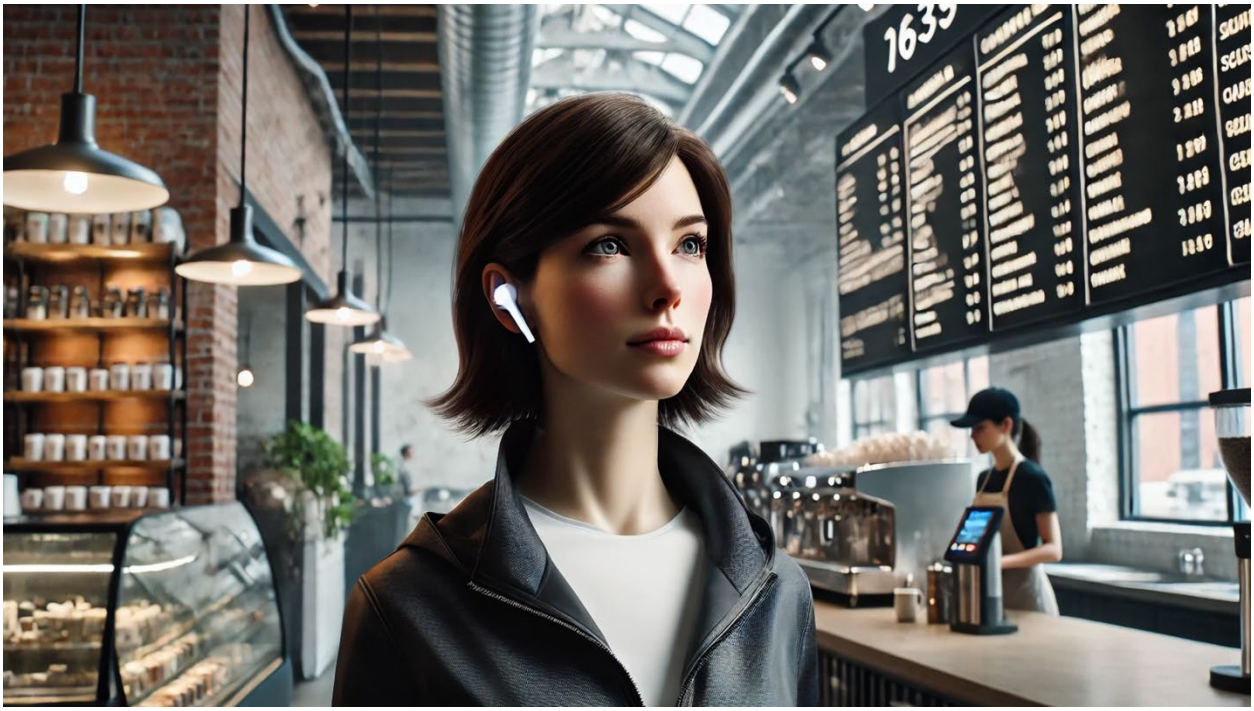
Appendix G: Future 2 Storyboard

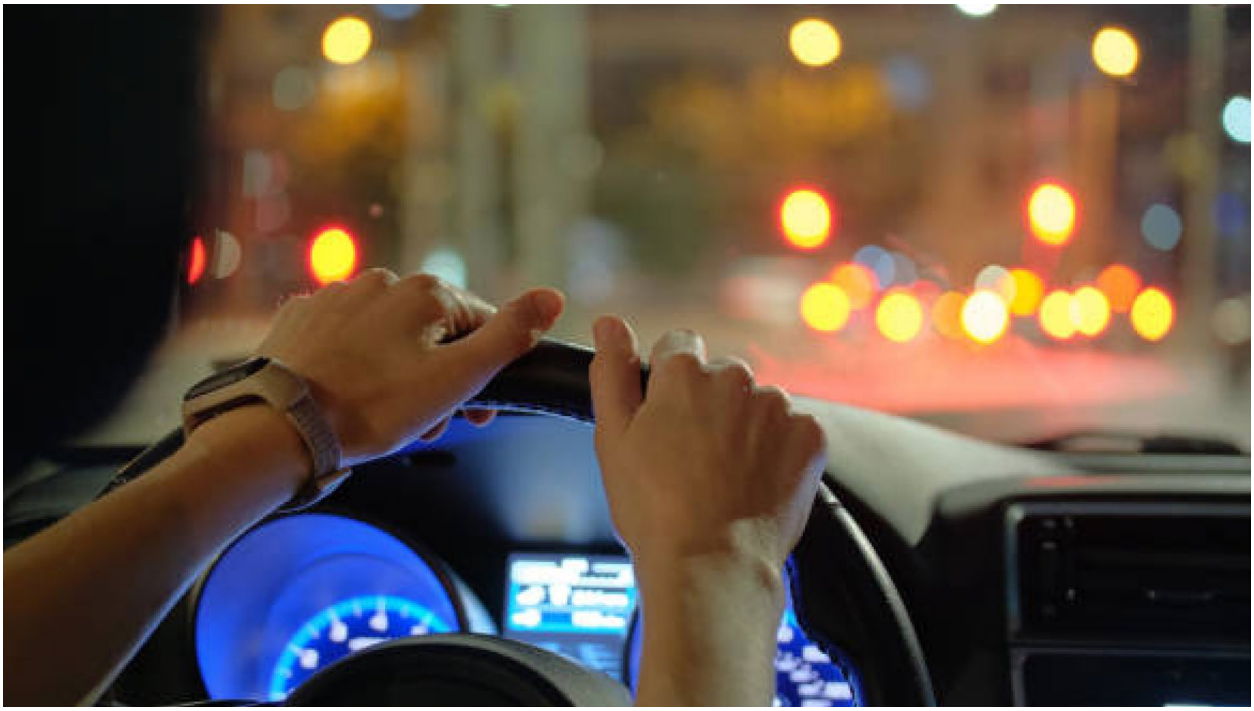




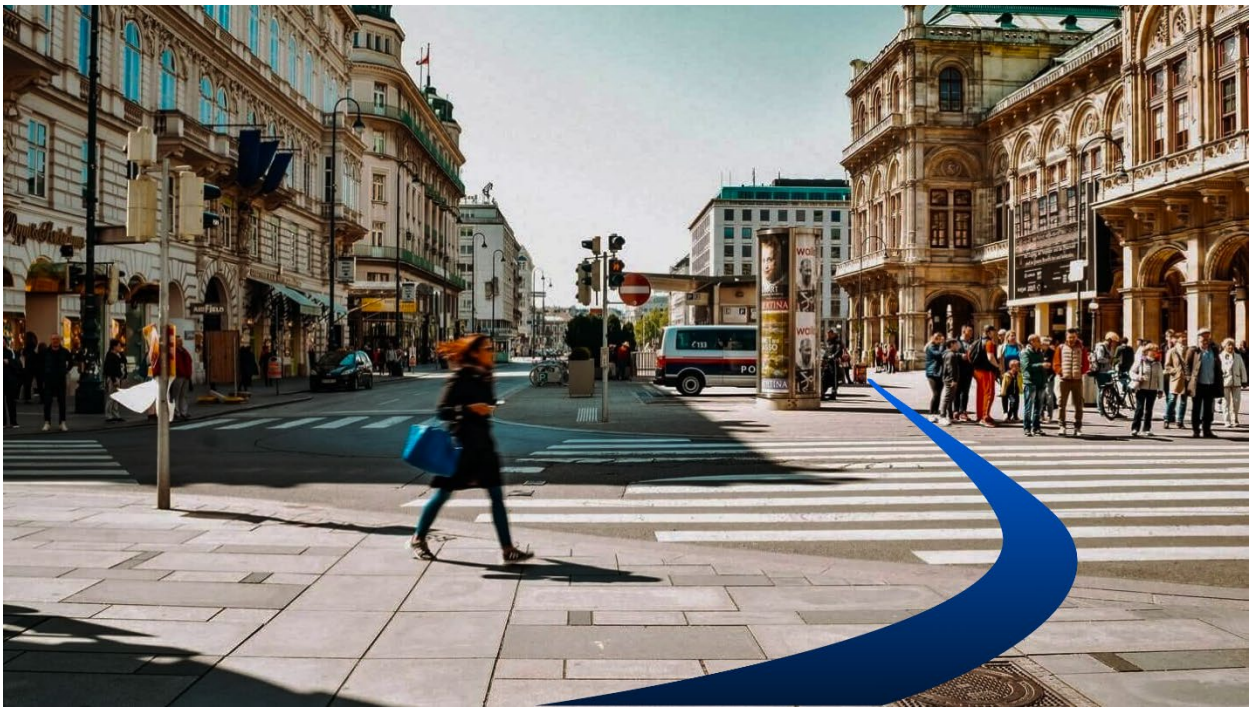

BREAKING
NEWS
BREAKING NEWS
AI Hacking Suspected in Digital Identity Wipe
LIVE
TV NEWS
PALLADIUM
IRIDIUM +8.026%
RHODIUM +0.240%
06:01 PM PT


HUMAN
MADE
COFFEE
No AI Automation
17





Appendix H: Future 3 Storyboard




Optimal Version Off