Sketches in Voice User Interface

Relational Conversations with Virtual Personal Assistants in Domestic Spaces

Manisha Laroia

Sketches in Voice User Interface

Relational Conversations with Virtual Personal Assistants in Domestic Spaces

Manisha Laroia
https://www.manishalaroia.com/sketches-in-vui
A thesis presented to OCAD University in partial fulfillment of the requirements for the degree of Master of Design in Digital Futures.
OCAD University
Toronto, Ontario, Canada 2021

Citation:
Laroia, Manisha (2021) *Sketches in Voice User Interface: Relational Conversations with Virtual Personal Assistants in Domestic Spaces*. Masters thesis, OCAD University.

Illustrated, edited & designed by Manisha Laroia.
All illustrations, photographs and logos in this document attributed to the source are subject to copyright of the respective people/organizations. All illustrations and photographs not attributed are created by the author and are a copyright
© Manisha Laroia

Electronic version available at http://openresearch.ocadu.ca/id/eprint/3313/
© Manisha Laroia
ISBN 978-1-77835-090-0
May 2021

Abstract

Sketches in Voice User Interface explores the conversational and evocative aspects of peoples' interactions with no-screen embodied voice user interfaces (VUIs) in domestic spaces. The project uses an annotated research *through* design methodology to create a series of *Sketches in Voice User Interface* for relational conversations with users. The research involves an autoethnographic study of existing voice-based virtual personal assistants (VPAs). Informed by these precedents *Sketches in VUI* are designed through iterative prototyping to explore ways in which VUIs can go beyond the existing virtual personal assistant in our everyday conversations. Unlike the conventional voice-based VPAs (Siri, Alexa, Google Assistant) operating on the commands of the user, the *Sketches in VUI* drive conversations and take an agentive role in human-computer conversations.

Using the design research approach, this project serves as a bridge between two key contextual voices in the domain of conversational technologies. On one hand, is the tech industry's case for usability that VUI is 'the most natural interface.' On the other hand, is the social sciences case *critically* calling VUI 'an artificial nature' and questioning if conversations with a machine are conversations at all.

The project concludes with an 'experience study' to enquire into the experience of participants as they converse with the designed *Sketches*. The study observes how participants react to the *Sketches* (behavioural response) and how they feel (emotional response), comparing them to their experience of existing voice-based VPAs, captured via videography and qualitative interviews. The study findings along with the designed *Sketches* form an annotated portfolio of generated knowledge about relational conversations with embodied voice user interfaces in our intimate spaces.

Acknowledgements

OCAD University and the **Digital Futures program and faculty**, thank you for the opportunity to explore, learn and challenge myself with each project.

A big heartfelt thank you to **my advisors** for guiding me through this year, encouraging me to express my creativity and drive my research with the much-needed occasional nudges and support.

Professor Kate Hartman, for the many conversations, the books, the warmth, and the attention to detail. I have learnt so much from you.

Dr Bill Leeming for all the questions and feedback each one lighting a bright bulb in my head as I read it.

Dr Barbara Rauch, for making me a part of your course and giving me a testing ground for building and demonstrating my *Sketches*.

My **Digital Futures cohort**, thank you for the amazing two years of learning and the support through this very strange virtual thesis year; especially to Arsh, Li and Sananda for being my home away from home.

Thank you **Maa, Paa** and **Bhai**. I told my father that I was going to research VUIs and he asked, "What is that?" I said, "…its the technology behind Siri and Google Assistant. He replied, "Basically a 'Gapodi Gulak' (which in Hindi means a chatty piggy-bank). So, you first made boxes and now you make boxes talk, that is progress!" Maa, as usual, was excited about it.

To all my friends who kept me going and to **Saumya**, for the endless conversations, and for gifting me the book that inspired this work.

Table of Contents

Abstract i

Acknowledgements ii

Table of Contents iii

List of Appendices v

List of Figures vi

1. Introduction 1

1.1 Voice User Interfaces 7

1.2 Sketches as a Prototyping Method 10

1.3 Scope and Limitations 11

2. Background 15

2.1 Working of an Embodied VPA 16

2.2 Components of a VUI 19

2.3 Let's Talk about Talking 21

2.4 Talking to Machines 22

2.5 The Context: Domestic Spaces 23

3. Contextual Framework 27

3.1 Conversational Agents 28

3.2 The Case against Human-Computer Conversations 32

3.3 Positioning Sketches in VUI 34

4. Research Methodology 36

4.1 Annotated Research Through Design 37

4.2 Methods and Techniques 38

5. Living with Jeeves: An autoethnography 40

5.1 Observations & Reflections 43

5.2 Informing Sketches in VUI 50

6. The Sketches in VUI 53

6.1 Sketch 0: Learning through making 59

6.2 Sketch 1: The Thinker 64

6.3 Sketch 2: The Listener 70

6.4 Sketch 3: The Learner 77

6.5 Sketch 4: The Chatterer 82

6.6 Summary 88

7. Experience Testing 89

7.1 Collected Data 91

7.2 Discussion 100

8. Conclusion 106

Bibliography 111

Appendices 116

List of Appendices

Appendix A: Autoethnography Journey Map 116

Appendix B: Working with Action Console 132

Appendix C: Thesis in 30 by Samantha Sherer 133

Appendix D: Sketch 0 VUI Flow and Sample Dialogs 134

Appendix E: Sketch 1 VUI Flow and Sample Dialogs 138

Appendix F: Sketch 2 VUI Flow and Sample Dialogs 144

Appendix G: Sketch 3 VUI Flow and Sample Dialogs 151

Appendix H: Sketch 4 VUI Flow and Sample Dialogs 157

Appendix I: Experience Testing Study Protocol 162

Appendix J: Accompanying Materials 166

Appendix K: Accompanying Materials 166

List of Figures

Figure 1: The TalkTags are activated on tap. By tapping on the conductive fabric touch spots on it, the user can play the programmed voice notes. 2
Figure 2: The spectrum of work on VUI. 4
Figure 3: The *Sketches in VUI* using Design Research bridges the gap between the social sciences and the tech industry. 4
Figure 4: The process. 6
Figure 5: Some of the commercially available VPAs and their small version smart speaker embodiments. Source of logo images: Wikipedia, Public domain, voicebot.ai and bestbuy.com. 7
Figure 6: A conversation on ELIZA. Source: Wikipedia, Public domain. 9
Figure 7: Visual representation for the *Sketches in VUI*, *Sketch* 1 to 4 from left to right. 10
Figure 8: Breaking down the research questions to define the scope and limitations of the project. 11
Figure 9: Google Nest mini device. Source: walmart.com. 12
Figure 10: An early talking machine Euphonia by Joseph Faber. Source: racingnelliebly.com 17
Figure 11: The modules of a Virtual Personal Assistant. 18
Figure 12: People were bothered that their child's name is used as a VPA wake-word. Source: Reddit.com 20
Figure 13: Jetsons, a 1962 Hannah-Barbara cartoon show, partially coming to life with video communication, sociable robots and home automation. Source: flickr.com (CC license). 23
Figure 14: Using a VPA device in a shared domestic setting where anyone can activate it. An incident recorded from my autoethnography. 24
Figure 15: Elderly participants interacting with Mabu robot (Left, source: medgadget.com) and Memory Lane project (Right, source: designweek.co.uk). 25
Figure 16: The weirdest experience is when it sends a notification that looks like a chat message from someone you know but it is from an AI. A glimpse of the notifications and message I received while using Replika. 28
Figure 17: Kuri (left), Kismet (centre) and Jibo (right). Source: Wikipedia commons & robots.ieee.org. 30
Figure 18: Karma, Edi and Sig voice-AI assistants from Our Friends Electric short film. Source: ACM Digital Library. 31

Figure 19: LAUREN project by Lauren Lee McCarthy where she played the assistant to people in their homes. Source: https://lauren-mccarthy.com/LAUREN/ (Permission taken by author) 33
Figure 20: Positioning the *Sketches in VUI* in the contextual framework. 35
Figure 21: My three-step design research process of Research, Making and Reflection showing methods and techniques incorporated. 38
Figure 22: Conversation from my autoethnography journey map with the new Google nest mini device in September 2020. See the detailed journey map in Appendix 2. 41
Figure 23: Tracking my Assistant activity with saved voice notes and in parallel, noting my reflections in my journal. 42
Figure 24: Illustrating a new relationship with Jeeves inspired by the covers of P.G. Wodehouse books. 43
Figure 25: Snippets of small conversations. 45
Figure 26: The incidents of having a conversation with the device while I was semi-sleep and notes of becoming a VUI whisperer. Details in Appendix 2. 46
Figure 27: Plotting my experience of using VUIs in current VPAs (specifically Google Assistant & Siri) after the autoethnography study. 48
Figure 28: Where the device lives in my room. From my desk, I often turn to look at it to talk. 49
Figure 29: Comparing conversation content from different VPAs. 49
Figure 30: Doodles exploring different possibilities for conversations with VUI. 50
Figure 31: Defining the *Sketches in Voice User Interface*. 54
Figure 32: Evolution of the *Sketches in VUI*. My journey of making and reflecting. 57
Figure 33: Working of a Google Action. Source: Google Developers 59
Figure 34: Creating my first *Sketch in VUI*. 60
Figure 35: Outlines the introductions of the *Sketch 0* VUI flow. 60
Figure 36: Phone and voice simulation running in Action Console in Chrome browser for trying it out. 61
Figure 37: The above is the list of errors it generated for the word 'Thesis' and '30'. 62
Figure 38: Correcting the main Invocation i.e., the call name for the project 63
Figure 39: Learnings from the *Sketch* in terms of the structure of statements to initiate the desired user response. 63
Figure 40: Demo of *Sketch 0* using screen recording of the interaction. This method was not very engaging, and I decided to interact with the device on video call for the next demonstrations. 64

Figure 41: The conversation in the autoethnographic journey that inspired *Sketch 1.* 65
Figure 42: Introduction flow for *Sketch 1.* 66
Figure 43: End of flow for *Sketch 1.* 66
Figure 44: Using VUI elements to encourage the user to talk and indicate that the *Sketch* is listening. 67
Figure 45: The sample conversation flow with the use of conversation 'traffic-signals.' 67
Figure 46: Adding the intent for 'Repeat' so the user. It shows different training phrases fed to the engine to learn what to listen to when the user asks it to 'Repeat.' 68
Figure 47: Demonstration of me talking to the Sketch for a virtual class. Even though we were remote, seeing me in a face-to-face with the device did get viewers engaged in the interaction demo. 69
Figure 48: *Sketch 1* takes control for ending the conversation even though the user starts the program by asking for the *Sketch.* 70
Figure 49: Replika employs the caring act by recollecting information shared with it before. 71
Figure 50: The conversation in the autoethnographic journey that inspired *Sketch 2.* 71
Figure 51: A rough VUI flow of the *Sketch* before I start programming the Action. 72
Figure 52: *Sketch 2* introduction, the customized sub-flows based on the user's response. 73
Figure 53: The sub-flows designed for Scene 1 to suit the different responses users might give. 73
Figure 54: The ending of *Sketch 2* where it asks if the user wishes to leave or else continues in a loop. 74
Figure 55: Using the mimicking technique to have the Sketch mimic the user's state and emotion. 75
Figure 56: Giving the user centre stage in the conversation to make the *Sketch* a listener. 75
Figure 57: The *Sketch* responds to the user acknowledging a polite response to make the user comfortable in the conversation. 75
Figure 58: The *Sketch* changes the topic of conversation and gives the baton back to the user. 76
Figure 59: The conversation in the autoethnographic journey that inspired *Sketch 3.* 77
Figure 60: The introduction flow for *Sketch 3.* 78
Figure 61: Conversation starters integrated into the VUI flow along with calling out participants name to make them feel involved. 79
Figure 62: The *Sketch* phasing out at the end of the flow. 79
Figure 63: The *Sketch* acknowledges that it is learning and asks a participant for help. 80
Figure 64: The *Sketch* shows interest in what the participant has to say. 80
Figure 65: Giving the *Sketch* a backstory so it can actively participate in the conversation. 81

Figure 66: Trying *Sketch 3* on a video call with a friend. 82
Figure 67: Making the Scene 1 transition of the *Sketch* on Action console- "Please talk to me". 83
Figure 68: The conversation from the autoethnographic journal that inspired *Sketch 4*. 83
Figure 69: Generating the content for *Sketch 4* on Talk to Transformer. 84
Figure 70: Introduction flow for *Sketch 4*. 85
Figure 71: The end flow of the *Sketch*. 85
Figure 72: Different responses composed for each time the participant asks the *Sketch* to repeat. 86
Figure 73: Overwriting system prompts for 'no user response', 'cancel' and 'repeat.' 86
Figure 74: Few dialogues into trying Sketch 4 I stopped responding and I was surprised to hear it say, "Are you there? Oh no did I lose you," even though I programmed it. It was the first time a VPA spoke a statement like that. 87
Figure 75: The collage of images as shared by the participants of their VPA devices living in their homes. 92
Figure 76: First response by participants to *Sketch 1*. 93
Figure 77: First response by participants to *Sketch 2*. 93
Figure 78: First response by participants to *Sketch 3*. 94
Figure 79: First response by participants to *Sketch 4*. 94
Figure 80: The Before and After response for participants 1 to 5. 96
Figure 81: The Before and After response for participants 6 and 7. 97
Figure 82: A few screenshots from the videography documentation of the Experience Testing with participants interacting with the *Sketches* in their homes. 99
Figure 83: Slider matrix showing the shift from left to right. The number in the boxes represents the number of participants that selected that slot. 101
Figure 84: A reflection on VUI design element combinations that brought about specific experiences and conversations through the *Sketches*. 103
Figure 85: What do the person and the VUI look like in conversation? (Built on the illustration by Dan O'Sullivan and Tom Igoe's of "What does a person look like to a computer?" in Physical Computing, 2004). 103
Figure 86: The 3D framework of a new order of conversations. 108

Hey Google, I am researching you.

Alright, I am here and ready to help.

Thank you for being a part of my research.

You said, thank you for being a part of my research. Did I understand that right?

Yeah!

Dear diary, today I helped someone.

1. Introduction

The inspiration for voice interfaces has been envisioned over the years in various fictional works, be it books, television shows or movies. *HAL 9000 from the 2001: A Space Odyssey (Movie)*[1] and *Samantha* from the 2013 movie *Her*[2] though speculative, are very believable portrayals of agentive voice interfaces and have inspired many researchers in the field with the latest addition of me. My formal training in Product Design trained me to work with physical materials- forming, shaping and designing them for usability and aesthetics. Through my design practice, I have been a 'maker' creating tangible experiences for people, and a 'design researcher' playing the fly-on-the-wall observing how people interact with products and technologies we design. The last five years saw the number of VPA enabled devices grow exponentially and voice interfaces have come to the forefront as primary interfaces for many consumer products. With VUI there is now an interface via which we can talk to computers. It is intangible and invisible but so natural to people that the product designer in me felt challenged and excited to explore it.

Figure 1: The TalkTags are activated on tap. By tapping on the conductive fabric touch spots on it, the user can play the programmed voice notes.

During the *Advanced Wearables* course with my advisor, Kate Hartman, here in the *Digital Futures* program I worked on a project called TalkTags[3]. TalkTags are social-wearable badges made using an *Adafruit Circuit Playground Express* microcontroller inbuilt with recorded voice notes. The voice notes are conversation starters like 'How are you feeling?' or 'What is the last book you read?' The tags are designed such that each tag has a unique pattern that can tag on to the other. They are activated when stacked together and users can switch through different voice notes programmed in it, each serving as a trigger to invite users to talk

[1] Even now, people like to test VUIs and chatbots with the famous line, "Open the pod bay doors, HAL." (Pearl,2017).
[2] https://www.warnerbros.com/movies/her
[3] http://blog.ocad.ca/wordpress/digf3010-fw20103-01/2020/04/talktags/

(Figure 1). They are inspired by the mobile phone etiquette of putting phones face down in social settings to indicate that one is present and wants to engage in the conversation. Through the making of this project, I hypothesised that what if voice interfaces could play an active role in conversational human-computer interactions. This planted the seed for this research endeavour.

My work, *Sketches in Voice User Interface,* investigates the field of Conversation Design - primarily Voice User Interface Design, positioned in the umbrella field of Interaction Design[4]. Google defines Conversation Design as,

> a design language based on human conversation. It's a synthesis of several design disciplines, including voice user interface design, interaction design, visual design, motion design, audio design, and UX writing (Google, Alphabet Inc. n.d.).

I am working with this definition and I describe my work as an *interdisciplinary* research inquiry where the research is built upon knowledge from disciplines like human-computer interaction (HCI), artificial intelligence[5] (AI), psychology, sociolinguistics, conversation analysis and design research. The significance of the project lies in it being a bridge between two key contextual voices in the current domain of conversational technologies. On one hand, is the tech industry's case for usability – 'the most natural interface'(Vlahos 2019) and for effectiveness- 'for tasks where the user should not be distracted or cannot pay attention to a visual interface.' On the other hand, is the social sciences case *critically* observing the socio-cultural impact that this technology has on people's conversations with one another. They call it 'a second nature, an artificial nature' questioning if conversation with a machine is conversation at all (Turkle 2016). To dive deeper I look at the two sides of this metaphorical coin and carve a path using design research to bridge the gap by introducing the *Sketches in VUI*.

---

[4] Interaction design is design for human use. It involves answering three questions:
How do you do? What sort of ways do you affect the world: poke it, manipulate it, sit on it?
How do you feel? What do you sense of the world and what are the sensory qualities that shape media?
How do you know? What are the ways that you learn and plan (or perhaps, how we want you to think)?(Verplank 2000)
Bill Moggridge explains interaction design as a combination of "software and user-interface design (Moggridge 2007)."

[5] Artificial Intelligence (AI) is the field that studies the synthesis and analysis of computational agents that act intelligently. Its underlying focus is to understand the human mind and is modelled on the functions of the human brain and its intelligence. AI is intelligence that is created artificially or synthesized (Poole 2017). AI can also be described as the intelligence possessed by a machine or an artificial entity that makes it behave intelligently but might not be based on human intelligence like Alien-AI (Brighton, 2012). John McCarthy coined the term "artificial intelligence" in 1956; he defined it as "the science and engineering of making intelligent machines."

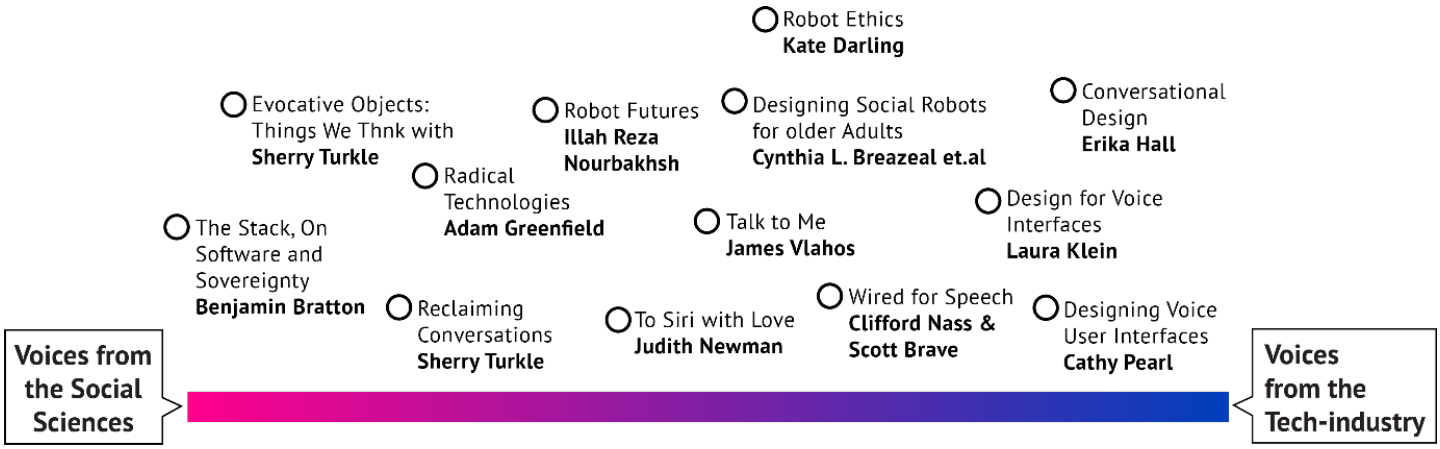


Figure 2: The spectrum of work on VUI.

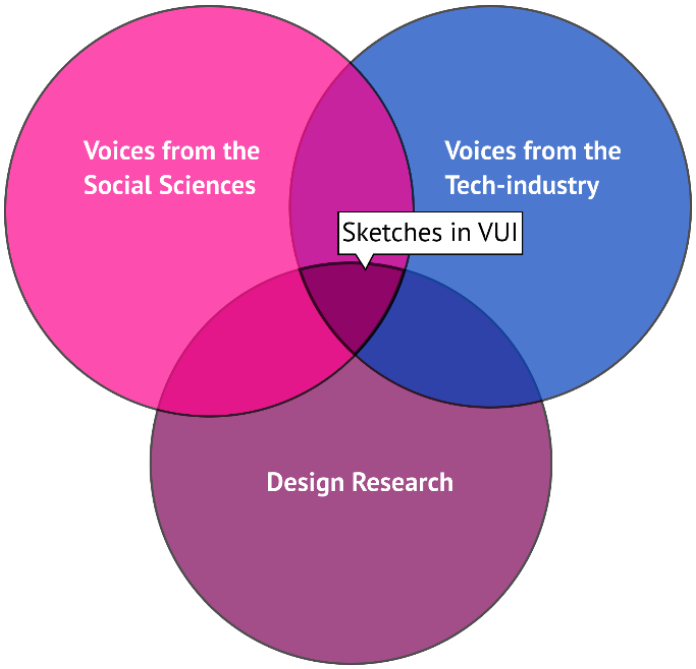


Figure 3: The *Sketches in VUI* using Design Research bridges the gap between the social sciences and the tech industry.

The project serves as an enquiry into the experience of users where the VUI is not just an assistant operating on the commands of the user but is tending to go beyond asking the research question,

> **In what ways can we design voice user interfaces to go beyond the existing virtual personal assistant and become an active agent in our everyday intimate conversations?**

My work offers contributions related to the design of voice user interfaces for relational conversations with VPAs[6] at home, and reflections into designing for an intangible and non-visual interaction design material i.e., voice.

STRUCTURE OF THE THESIS

The main body of the thesis is divided into eight sections (Figure 4) outlining my research journey each informing the next and culminating in this thesis and the exhibited work. Methodologically, I use annotated Research through Design approach (Chapter 4). Living with Jeeves (Chapter 5) is an autoethnographic account of my experience with the *Google Nest Mini device*. Sketching as prototyping (Moussette 2012) is used to design the *Sketches in VUI* (Chapter 6) in an iterative process. These are tested with nine participants in an Experience Testing study (Chapter 7). Discussion collates the findings of the study. The hope is to inspire design work with a critical thinking lens when it comes to emerging technologies like VUI rapidly entering our everyday culture.

[6] In this thesis the use of the word 'VPAs,' 'device,' 'computer' or 'machine' is being made in the context of embodied voice only modalities of voice user interface particularly virtual personal assistants enabled smart speakers.

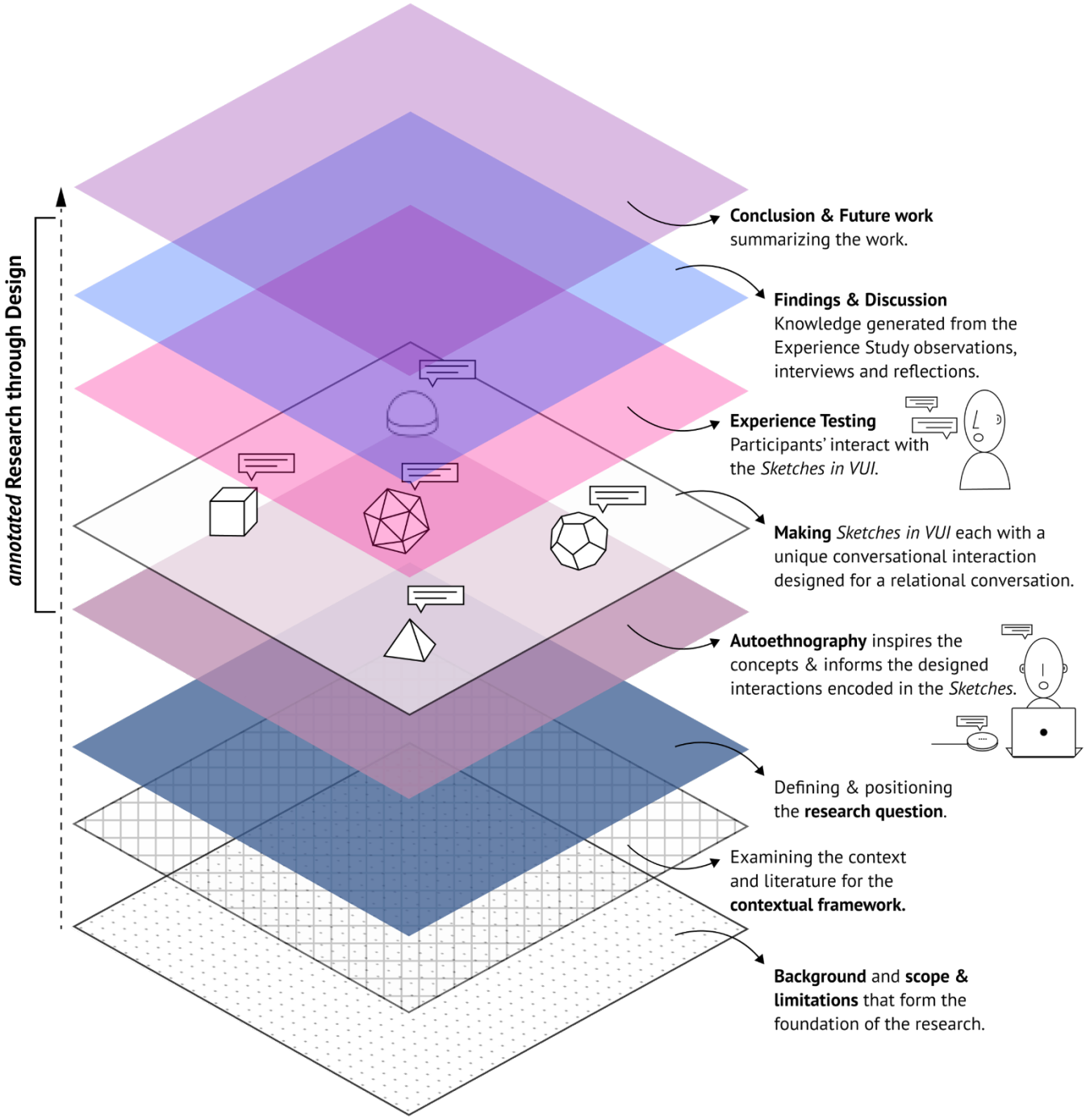


Figure 4: The process.

1.1 Voice User Interfaces

The earliest definition of Voice-user Interface (VUI) is:

> […] what a person interacts with when communicating with a spoken language application. The elements of VUI include prompts, grammars, and dialogue logic (call flow). The prompts or system messages are all recordings or synthesized speech played to the user during the dialogue (Cohen, Giangola, and Balogh 2004).

Then voice technology was focused on interactive voice response (IVR) but in the context of modern VPAs, an interface that has a conversation with the user either typed or verbal is a Conversational Interface. This could be a Chatbot[7] and/or a Voice-bot (Pearl 2017).

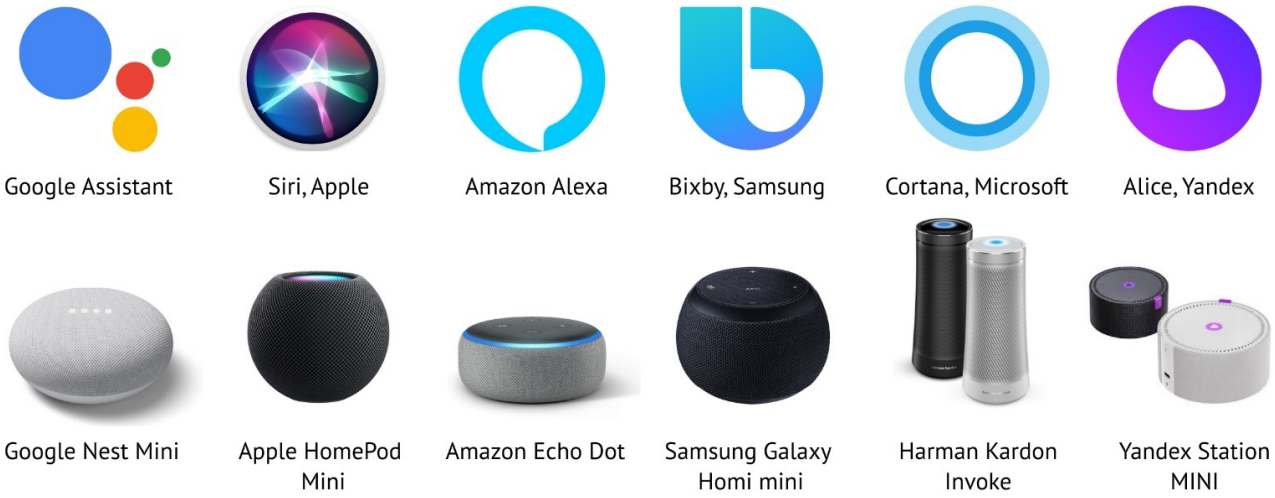


Figure 5: Some of the commercially available VPAs and their small version smart speaker embodiments. Source of logo images: Wikipedia, Public domain, voicebot.ai and bestbuy.com.

EMBODIED AND DISEMBODIED VUI

Voice as an anthropomorphic or human-like interface has been researched and developed as both an embodied system where the human interacts with a robot or an animated avatar and a disembodied system where the human interacts through speech or text entered at a keyboard (C. Breazeal 2003). Embodied conversational agents are "specifically conversational in their behaviours, and specifically human-like in the

[7] A chatbot as "a computer program designed to simulate conversation with human users, especially over the Internet." The word "bot" is also sometimes used to refer to these types of interactions. Chatbots can have a VUI, but more typically they use a text-based interface (Pearl 2017).

way they use their bodies in conversation"(Luger and Sellen 2016). VUIs allow the user to interact with a system through speech commands. It can be incorporated into voice-enabled humanoid robots and virtual personal assistants (VPA), such as Siri, Google Assistant, and Alexa embodied as smart speakers, smartphones and personal computers, and wearable devices like smartwatches. In IVRs[8] or VPAs through hands-free devices like Bluetooth earphones, VUI is a disembodied agent.

RELATIONAL VOICE USER INTERFACE

The protagonist in *Her* (movie) develops a romantic fondness for his virtual assistant. Raj[9] in *The Big Bang Theory* (TV series) dates Siri. These and many such pop-culture references illustrate our temptation to give VPAs a role of more than a computer answering commands. Because they can talk to us like people, they are social actors[10](Nass and Brave 2005) and we are bringing them into emotional human conversations knowingly or unknowingly. "Alexa isn't just an assistant in most people's mind, 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)."

Users engage in three kinds of conversation goals, **task goals**: relevant to the task people have come together to accomplish, **communication goals**: aimed at making sure that the communication itself goes smoothly and everyone understands each other; and **relationship goals**: these are the goals that drive people to set and maintain the tone of the conversation or relationship, how much the interaction may be friendly, polite, hostile, reciprocal, conflicted, professional, intimate, formal, informal, and so on (Shechtman and M. Horowitz 2003). VPAs primarily designed for task-based conversations like asking it to set an alarm are being used for relational conversations (non-task based) like talking about ones' feelings to build a bond. People are tending towards **relational agents** which are computational artifacts designed to build long-term, social-emotional relationships with their users (Bickmore and Picard 2005).

[8] Many IVR (Interactive Voice Response) systems on phones were more "conversational" than some current VUIs, as they kept track of what callers had already said and used that information to prepopulate later questions in the dialog (Pearl 2017).

[9] The Beta Test Initiation | Episode aired 26 January 2012. https://www.imdb.com/title/tt2181587/

[10] […] a broad category that includes any entity that opt in the range of social attributions and social responses. They receive some human provisions i.e. good recorded speech obtains a greater level of humanness than synthetic speech. Designing the expectation that all VUI can receive significant applications of social rules is excellent practice (Nass and Brave 2005).

The development of relational agents draws from two existent threads of work in HCI: natural multimodal interfaces (including embodied conversational agents [Cassell et al. 2000b] and sociable robots (C. L. Breazeal 2002), and studies of computers as social actors (Reeves and Nass 1996). Relationship-building strategies discussed in the social sciences literature are implementable as verbal (sharing personal thoughts and stories) or nonverbal (face-to-face or eye contact) conversational behaviours (Bickmore and Picard 2005). This requires, at a minimum, some kind of natural conversational interface and, at a maximum, the use of embodied conversational agents, or some other articulate physical form factor to enact both verbal and nonverbal communicative actions (Bickmore and Picard 2005). Smart speakers enabled with VPAs are embodied VUI agents and use the rhythmic LEDs glow patterns to emulate human-like eye blinking and have a physical embodiment and presence.

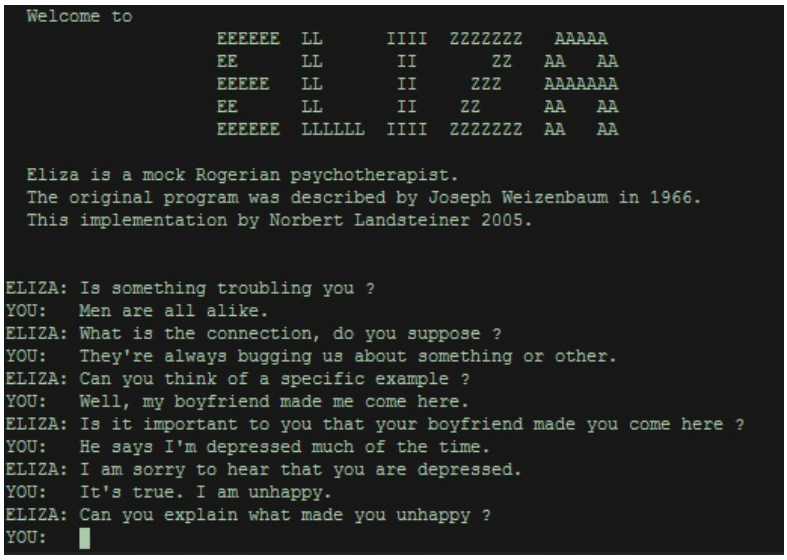


Figure 6: A conversation on ELIZA. Source: Wikipedia, Public domain.

Considering the verbal relationship-building strategies is a 1963 rudimentary chatbot ELIZA. It was created by Joseph Weizenbaum to demonstrate that computers could not understand people and people would not enjoy chatting with them. Ironically, people loved it and his secretary while chatting with ELIZA in his office asked him to leave the room for privacy (Mars and Hall 2019). ELIZA was a scripted bot that used opening sentences and would take what people typed in and reframe it into questions.

The 'magical element' of a conversational agent has more to do with our biology than with technology. Being spoken to reflexively makes us respond (Enfield 2017) and hearing our own words repeated indicates that we are being heard (Nass and Brave 2005). We interact with these agents through text-chat and get

involved in the interaction - 'the ELIZA effect[11]'. Take the text chat and replace it with voice, and the impact of the interface is even stronger. The voice of a conversational machine is that anthropomorphizing feature that tends to convince us that the machine is not a tool anymore but an intelligent agent, one we can have conversations with (Nass and Brave 2005).

1.2 Sketches as a Prototyping Method

I am borrowing this term from Camille Moussettes' who in his thesis uses a method of 'Sketching in Hardware' to explore simple haptic interactions and defines sketches as an evocative, suggestive, rough, easy to build exploration not necessarily on paper (Moussette 2012). Going over the many meanings of the word sketch (noun), attributes like rough, preliminary, tentative draft, intentionally slight in treatment, and discursive are key factors for understanding the essence of sketches (Moussette 2012). Giving these attributes to my design outputs I call them *Sketches in Voice User Interfaces*.

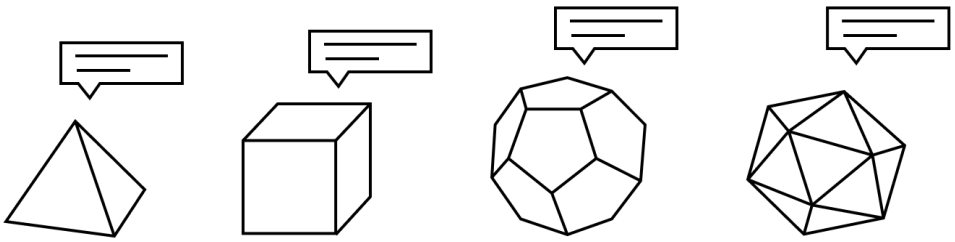


Figure 7: Visual representation for the *Sketches in VUI*, *Sketch* 1 to 4 from left to right.

Another reason for using the term sketch is that it refers to a theatrical piece having a single scene (Merriam-Webster, n.d.). The act of conversation is like a theatrical sketch where the person and the machine are actors and body language, facial expressions, utterances, and the content of the utterance play a role in the act. Here the designed outcomes i.e., a software *Sketch* with a voice-based interface are interacted with, by a user in a short conversation (a sketch scene) which is observed by me as the researcher to draw out knowledge in terms of observations and reflections. Google refers to voice/chat conversational interactions as Actions or Dialogs and Amazon and Microsoft refer to them as Skills. More than an action or skill, my curiosity in our conversations with machines is focussed on 'the social dialogue' i.e., the back and forth interaction that occurs in the act of it. It is in this context that the definition of a sketch as a theatrical piece

[11] "[…] And humans remain as open and sometimes gullible as ever when it comes to interacting with chatbots. There's even a term that's emerged in computer science – 'the ELIZA effect' – that refers to our tendency to anthropomorphize computers and to believe that programs understand, even when they really don't (Mars and Hall 2019).

having a single scene and as a verb is interesting to explore. The *Sketches* I create are 'useless' without the user as the user activates them and it is in the conversational act of the two actors i.e., humans and the *Sketches* that I hope to discover knowledge.

1.3 Scope and Limitations

Conversational AI is an umbrella term including both text chat and voice-based agents. I am focused on only voice-based conversational AI, particularly voice-based virtual personal assistants. I refer to chatbots and robots for background and context. My research is focused on embodied voice user interfaces where the embodiment for the VPAs are no-screen devices specifically home smart speakers enabled with a VPA.

Conventionally, the VPAs in homes are designed for task-based conversations but given the intimate nature of the space offers an opportunity to explore non-task based interactions. Taking account of COVID-19 pandemic restrictions, where we are spending most of our time indoors, I choose to contextualize the research in domestic spaces. The choice of the domestic context is instrumental in further narrowing the scope of research to relational conversations in an intimate environment and using these to design for scenarios where voice-based VPA serve as relational agents. An autoethnographic method is chosen over an ethnographic one due to limitations imposed by the pandemic and for ease of access to the research. Researching myself as the subject offers more dimensions to the data collected than remotely observing participants through a screen with limited physical cues.

Figure 8: Breaking down the research questions to define the scope and limitations of the project.

The making and testing process is structured to fit in the duration of the project and adhere to the limitation due to the pandemic restrictions. Many different platforms and types of *Sketches* are possible to be

programmed but I have limited it to **five *Sketches* i.e., four primary and one experimental.** The number of *Sketches* is determined by the key observations I gathered during the Autoethnography and let that inform the *Sketches*. The *Sketches* are built on the **Action Console platform by Google and run on the Google Nest mini speaker**.

Figure 9: Google Nest mini device. Source: walmart.com

I explored Mozilla's web-speech API and Processing's TTS (text-to-speech) libraries, but I was limited by my knowledge of programming in Python. Using microcontrollers to make voice-based interactions is also explored but the quality of the output is crude and not effective in relational conversations as the interaction is not smooth and the synthetic voice does not have good enunciation. With Action Console and Alexa Skill (Amazon's platform), I can access the highly trained and powerful Natural Language Processing (NLP) engines running in the backend. I can program using basic JSON programming, use conversation flows to detail the program, choose from different voice options, demonstrate the program in a simulator and run it on a smart speaker. It removes the dependency of having a finicky hardware setup and I can focus on the VUI only.

The choice of Action Console over Alexa Skill is based on two factors. Firstly, I own Google Assistant-enabled hardware to start with and have been using it. Secondly, the Assistant allows for different voice options to be used as default while the Alexa defaults to a female voice only with varying voice for the Skills. Critique of the use of only female voices in modern VPAs (Nass and Brave 2005; Meet Q n.d.; McCarthy 2018) inspired me to use a male voice for the Assistant through my autoethnography study (Chapter 5). As I am conducting a remote user-testing, using the Action Console makes it easy for me to send the *Sketches* to the study

participants' devices. Remote testing can limit the study findings due to it not being observed in-person and have been accounted for in Chapter 7.

A challenge in making the *Sketches* is learning the new programming language YAML[12] and coding for VUI. I am using online tutorials on YouTube and blogs on Stack Overflow, Reddit and Medium (Appendix 1) to build and cover the learning gap. Designing for VUI also has its learning curve where I bring in knowledge from designing graphic-user interfaces and built on it learning from VUI design guidelines in practice. VUI has components like voice, name, personality, flow and content. For the purpose, of the project, I am using voices already available on the Action Console and not exploring the gender or other characteristics of voice. I am also not designing personalities for the *Sketches* but focussing on VUI flow and content only. Content writing for the conversation flow is another challenge as I am not trained as a linguist. I can collaborate with a writer, but I want to explore writing the flows myself. The VUI design guidelines help with the basics of flow writing and I refer to people-conversation starter toolkits (Chapter 6) to add to it. Reflectively, writing the content on my own, I did end up learning a lot about language structure and its importance in conversational interfaces.

In this project, I am not exploring the form or non-verbal cues i.e., the physical embodiment of the *Sketches* due to limited access to fabrication facilities and scope down the project. I use visuals in the thesis to denote each *Sketch* to make it easy for communication but those are only visual representations and do not have any association to the physical form of the *Sketches*. I am also not exploring the wearability and ambient experience associated with VUI as it is beyond the scope of the project.

I am not demonstrating the *Sketches* as programs that adhere to the Turing test to be indistinguishable from a human or the Alexa Prize[13] test to be able to make a social conversation across topics. Instead, I am using the *Sketches* as explorations into ways of designing VUI as an active agent in conversations and observe how people interact with them. The *Sketches* are relational in interaction but are designed using platforms that are primarily for task-based interaction to harness the NLP engine to understand the user's utterance. As the

[12] The Action Console uses YAML which is a human-readable data-serialization language. It is commonly used for configuration files and in applications where data is being stored or transmitted. YAML targets many of the same communications applications as Extensible Markup Language but has a minimal syntax which intentionally differs from SGM.

[13] The Alexa Prize Socialbot Grand Challenge is a competition for university students dedicated to accelerating the field of conversational AI. The competition is focused on creating a socialbot, an Alexa skill that converses coherently and engagingly with humans on popular topics and news events. Source: https://developer.amazon.com/alexaprize/about

Sketches are explorations the conversation flow for them is scripted using Text-to-Speech and synthetic voices and the responses are not AI-generated. AI-generated language model integration into voice programs is technically complex, not easy to access as the applications are limited and beyond my scope. I am instead creating scripted but experiential VUI programs. I am not fooling the participants but having them experience what it is like to talk to a VUI that seems to have agency and study what impact it has on conversations.

2. Background

2.1 Working of an Embodied VPA

From the launch of Siri[14] in 2014 introducing "her" as "the best friend who gets you (Vlahos 2019)" to the human-like stuttering with "um" and "uh" of Google Duplex[15] as it calls a restaurant to make a reservation on behalf of its user, voice user interfaces (VUI) have become mainstream. They have found a place in our homes, on our devices and in the many digital prostheses (Rey and Boesel 2014) on our body. VUI has been explored as a social companion for children with autism (Newman 2018), as care support for seniors living alone (Quirmbach n.d.), as a voice for people who are vocally challenged (*Samsung Bixby Voice Assistant-MND Mother Helps Daughter with #VoiceForever* n.d.) and even as a way to immortalize the dead (Vlahos 2019). Voice technology started in the seventeenth century as an attempt to make inanimate objects talk like humans. Early automatons with pneumatic mechanisms to make speech sounds[16] have gone through several failed attempts of making a Turing Machine to become the synthetic, near-human voice in 'smart' plastic boxes of varying shapes and sizes– the virtual personal assistants (VPAs). According to Statista, the installed base of smart speakers reached 320 million units in 2020[17] with Amazon being the leading vendor followed by Google, Baidu, Alibaba, Xiaomi and now Apple. Amazon and Google are responsible for over three quarters of all smart speaker sales (Koksal 2020).

Over the years, VPAs were able to overcome technological hurdles and make it to our home devices due to improved computing power of hardware (Moore's Law), high-level of accuracy in speech recognition, improved Natural Language Understanding, increased use of web applications and search engines with APIs (Application Programming Interfaces)[18] and cloud computing[19] services that provided servers to perform the complex algorithm processing and enabling the query to be executed and the responses fed back to the device (Winarsky, Mark, and Kressel n.d.).

---

[14] Siri Launch video- https://www.youtube.com/watch?v=agzItTz35QQ

[15] "Google's 'Duplex' Could Be Your New Personal Assistant." n.d. NPR.Org. Accessed January 8, 2021. https://www.npr.org/2018/12/24/679895636/googles-duplex-could-be-your-new-personal-assistant.NPR's

[16] In the eighteenth century, Wolfgang von Kempelen made The Speaking Machine a contraption that used bellows to do the work of lungs where Kempelen pumped air through a pipe and over a bagpipe reed, whose vibrations simulated those of vocal cords. It replicated the plosive consonants such as 'p' and 'b.' Several metal tubes extending from a simulated throat could be manipulated with levers to produce the hissing 's' and 'sh' sounds as well as the nasal 'n' and 'm.'

[17] Source- https://www.statista.com/statistics/878650/worldwide-smart-speaker-installed-base-by-country/

[18] Small pieces of code that enable other applications to call upon the web service.

[19] In 1997, Chellappa [5] first ever coined the term "Cloud Computing" (CC) in his address to the INFORMS Annual Meeting (Ray 2018). The NIST Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

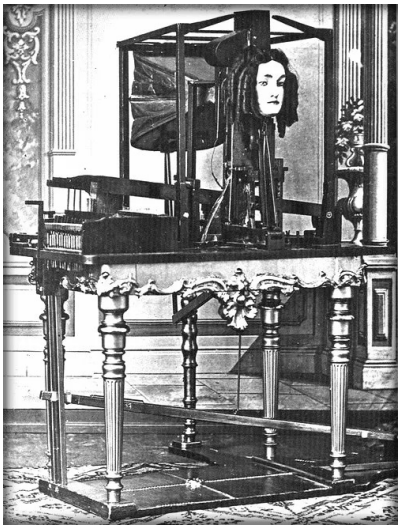


Figure 10: An early talking machine Euphonia by Joseph Faber. Source: racingnelliebly.com

In VPAs, the VUI is the front end with algorithms and AI models running in the backend. The working of a VPA needs the following modules in order of the flow of information (Figure 11):

I. the user speaks via a complex neuromuscular activity i.e., **wake word** followed by **voice command or spoken utterance**
II. special array design of microphones in the VPA enabled device picks up the voice
III. the VPA recognizes sounds i.e., **invocation/wake word** and **utterance** using **Automated Speech Recognition (ASR)** [20] engine
IV. it interprets certain keywords i.e., **intents and parameters** to understand what the user is saying and the context using **Natural Language Understanding (NLU)** [21] engine. The NLU uses syntactic and semantic analysis of text and speech to determine the meaning of a sentence.
V. it finds a suitable response that can be either scripted, retrieved, or generated by searching a **scripted program** or a **knowledge base** (generalized search response). It then generates responses in human-

[20] Automatically translating the spoken utterance to text. There are two major fee-based speech recognition engines: Google and Nuance. Other options in this space include Microsoft's Bing and iSpeech. Alexa uses the SpeechSynthesizer interface. Free ASR tools include the Web Speech API, Wit.ai, Sphinx (from Carnegie Mellon), and Kaldi (Pearl 2017).
[21] Natural language understanding is a subset of natural language processing, which uses syntactic and semantic analysis of text and speech to determine the meaning of a sentence. Syntax refers to the grammatical structure of a sentence, while semantics alludes to its intended meaning ("NLP vs. NLU vs. NLG: The Differences between Three Natural Language Processing Concepts" 2020).

readable language using **Natural Language Generation (NLG)**[22] engine based on the data input. The NLU and the NLG together form the **Natural Language Processing (NLP)**[23] engine.

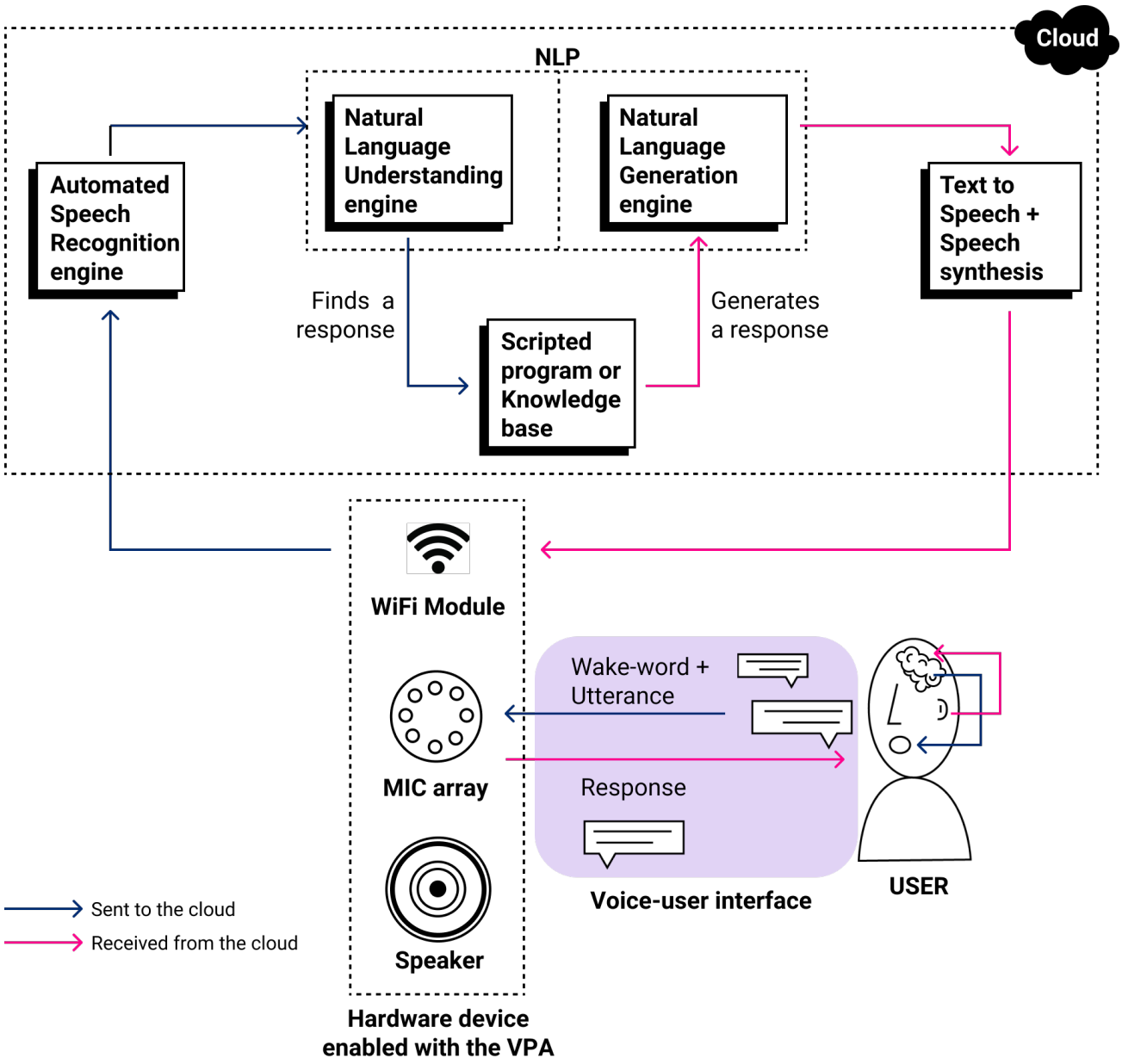


Figure 11: The modules of a Virtual Personal Assistant.

[22] NLG is the process of producing a human language text response based on some data input. This text can also be converted into a speech format through text-to-speech services ("NLP vs. NLU vs. NLG: The Differences between Three Natural Language Processing Concepts" 2020).

[23]Natural language processing, which evolved from computational linguistics, uses methods from various disciplines, such as computer science, artificial intelligence, linguistics, and data science, to enable computers to understand human language in both written and verbal forms - NLP = NLU + NLG ("NLP vs. NLU vs. NLG: The Differences between Three Natural Language Processing Concepts" 2020).

VI. the recording is summoned from the cloud and the response is played via **Text-to-Speech (TTS)** with **Speech synthesis** (using synthetic voice or human recorded voices)

VII. the synthesized response is relayed by the speaker

VIII. through a complex physiological process in the inner ear, the sound is captured and relayed to the human brain as a stimulus for cognitive processing and associated action.

All current natural language systems focus on one or a few "vertical domains" in which the users can expect a reasonable understanding of their utterances (Vlahos 2019) thus VPAs cannot understand everything a user might say. VPA are narrow-AI and not general-AI, in the sense they can function efficiently for certain tasks only and conversations is a complex activity.

2.2 Components of a VUI

The layer of the VPA that interacts with the user is the voice user interface (Figure 11). A VUI consists of a name, a characteristic voice and the dialogue content which contribute to the personality of the VUI and impacts its interaction with the user in the context.

THE NAME

A name for an object or brand communicates its value and impacts the user emotionally (Norman 2007). Siri's name was looked up in baby-name books. Alexa's name[24] referenced the great ancient repository of human knowledge, the Library of Alexandria. In VPAs the name of the agent is often its identity and is closely related to the **wake-word**. A wake-word is a globally reserved phrase in a VUI program, to be recognized at any time it is spoken. It is used to activate the VPA on a device to then access cloud-based services (Warden 2018). With the smart home speakers, one can name the speaker to personalize it and the VPA agent has a wake-word encoded in it. For the Google Assistant (name) the wake-word is "Hey, Google" or "OK, Google".

[24]A reddit post by parent who named their child Alexa and now are rethinking it.
https://www.reddit.com/r/namenerds/comments/bgtht9/our_3_yearold_daughters_name_is_alexa_should_we/

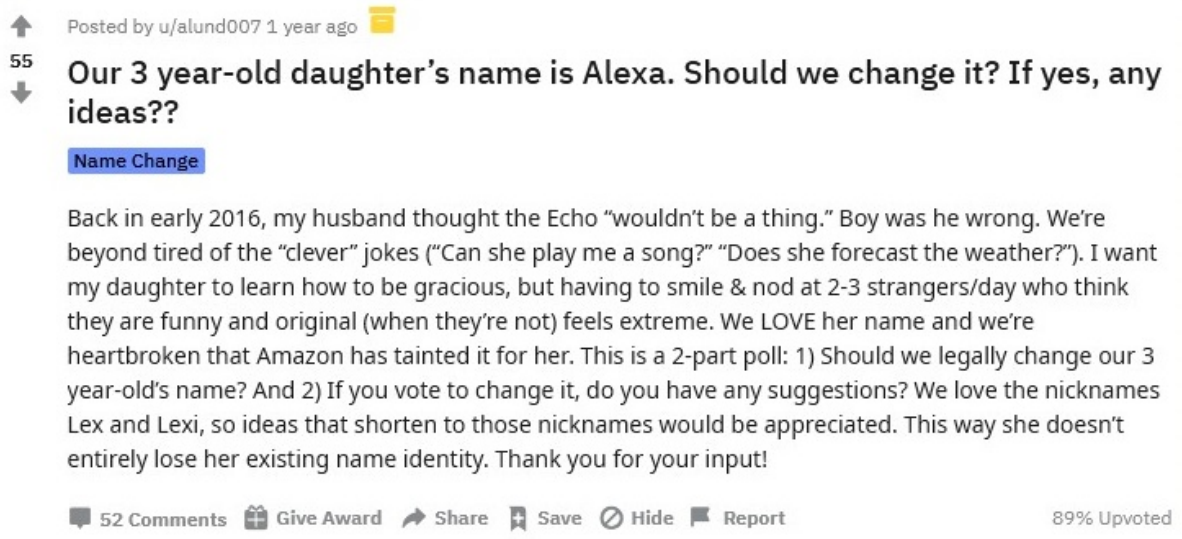


Figure 12: People were bothered that their child's name is used as a VPA wake-word. Source: Reddit.com

THE VOICE

"Parameters of voice such as **pitch**, **cadence** (modulation and inflection of the voice) are used by people to categorize voices as male or female and those such as **speech rate** and **volume** convey more subtle human characteristics such as personality, emotion, and hometown (Nass and Brave 2005). Humans are so attuned to vocal characteristics that they quickly and accurately distinguish one person's voice from another (Nass and Brave 2005). All the voices today on consumer-available VPAs are synthetic (Vlahos 2019; Pearl 2017). Voice AI creators use techniques like *sequence-to-sequence* (deep learning model) or *parametric synthesis* to generate word sounds synthetically and concatenate them into full words and phrases. DeepMind's WaveNet technology (2018) which is an evolved parametric synthesis makes the Google Assistant speak. It synthesizes waveforms and assembles them into words at a rate of up to 24,000 samples per second of speech (Vlahos 2019).

THE CONVERSATION CONTENT

The words that people select in a conversation carries social information and the same is for the words spoken by a VPA (Nass and Brave 2005). VPAs like people use different approaches to handle responses like for a misunderstood comment it can take responsibility, blame the speaker or scapegoat (Nass and Brave 2005). The content of what one says portrays emotion i.e., when someone is angry they speak in shorter sentences like "Move" but if they are feeling happy and relaxed they would say "Would you be so kind to open the door? ("Method Podcast, Episode 8" 2018)" The same knowledge is transferred to VPAs. The content

in current VPAs is scripted and retrieved when needed from the cloud. It is written by conversation designers and UX writers keeping it in sync with the VPA's personality and its personal narrative. In the project my focus in on conversation content and the inherent flow of information in it.

2.3 Let's Talk about Talking

"Speech is such an integral part of being human that people with IQ scores as low as 50 or brains 1/3 the size of a normal human brain can speak. Humans are so tuned to speech production and processing that from about the age of 18 months children on average learn 8-10 new words a day and retain that until adolescence (Nass and Brave 2005)". Thus, when people hear something speak, they assume that it has some intelligence, at least that of a toddler and anthropomorphise it. N. J. Enfield[25] describes that

> even the simplest conversation is a collaborative and precision-times achievement by the people involved. [...] when two people talk, they each become an interlocking piece in a single structure, driven by something that I will call the *conversation machine*. (Enfield 2017).

In the act of conversation people operate with high-level interpersonal cognition, where we infer others' intentions beyond the explicit meaning of the words, we monitor others' personal and moral commitment to the interaction and if necessary hold them to account for opting for the most efficient, and most helpful kinds of responses (Enfield 2017). We help each other, where necessary and possible and staying on track in a conversation requires a good deal of attention, effort and social cognitive skills (Enfield 2017).

For people conversing is like telepathy, in that one can simply by the vibration of sound waves make another person understand what they are thinking of or what they want to say ("Method Podcast, Episode 8" 2018) but for a machine, it is learning from scratch. Therefore, it gets complicated when we talk to machines. It gets more complicated when we move from chat to voice-based interactions because we also lose the visual cues. From a linguistics lens when we read, we don't know how many times one has rephrased a sentence but conversations are all draft (Enfield 2017). When the VUI interaction is not designed well it increases cognitive load because people have a certain expectation in a conversation due to it being our innate mode of communication. Urban says that "Humans have a higher bar for VUI because we have an intuitive way

[25] a professor and the chair of linguistics at the University of Sydney, and a research associate in the Language and Cognition Group at the Max Planck Institute, in his book How we Talk: The inner workings of a conversation (2017)

with language and speech, and they are more affected when these expectations break. Humans are very attuned to and sensitive to changes in conversations(“Method Podcast, Episode 8” 2018).”

2.4 Talking to Machines

A series of studies by Reeves and Nass demonstrated that people respond in social ways to computers (and other media) when provided with the appropriate social cues, even though they are typically unconscious of this behaviour (Reeves and Nass 1996). This finding was reiterated that when participants ‘believed’ they were talking to a person, they showed many more of the kinds of behaviours associated with establishing the interpersonal nature of a relationship (Shechtman and M. Horowitz 2003) bringing the factor of subjective response. Getting rid of the question of whether people ‘believe’ computers to be human-like or not, Nass and Brave demonstrate that people are "voice-activated": we respond to voice technologies as we respond to actual people and behave as we would in any social situation (Nass and Brave 2005). They wrote, and I quote,

> These technologies, like the speech of other people, activate all parts of the brain that are associated with social interaction. […] the psychology of interface speech is the psychology of human speech: Voice interfaces are intrinsically social interfaces (Nass and Brave 2005).”

It is this core finding by Nass and Brave, on which I built this project. Our “brains are voice experts” as we use the same parts of the brain to interact with machines as we do to interact with humans (Nass and Brave 2005). We find conversational AI attractive because “the human brain rarely makes distinctions between speaking to a machine and speaking to a person and applies the same rules and shortcuts with VUI as we use when interacting with people (Nass and Brave 2005)”. What people see as a box with a voice and flickering lights becomes a relational agent. We ask the device something and it responds with a speedy reply, giving us instant gratification and the sense of being there for us. Listeners and talkers cannot suppress their natural responses to speech, regardless of the source (Nass and Brave 2005). The *conversation machine* is ignited and then we respond adhering to “the social norms of conversation, that if we hear someone ask us something, we will respond within an average 200 milliseconds (Enfield 2017)”. Voice includes tone, volume, intonation, and rate of speech which conveys a great deal of information, unlike the written word. And users do not need to be instructed on how to use this technology. (Pearl 2017).

2.5 The Context: Domestic Spaces

Figure 13: Jetsons, a 1962 Hannah-Barbara cartoon show, partially coming to life with video communication, sociable robots and home automation. Source: flickr.com (CC license).

Welcome to the home[26] of 2021! The embodied VPAs have entered our home and live in our living rooms, in our kitchens, on our bedside and even in our bathrooms, built into more than a billion devices (Google Home Data 2019; Bohn 2019). These assistants live in a smart speaker the way a genie might in its, from where they are supposed to serve as the command hub of a connected home (Greenfield 2017). The devices enable users to issue a broad set of commands on a wide range of topics. In addition to playing music - the most common use (Bentley et al. 2018), users can ask in "natural language" about any general information on the web and get back voiced out results. These devices also provide smart home integration with IoT[27]

[26] In 2019 Swedish research firm Berg Insight says 63 million American homes will qualify as "smart" by 2022, with everything from Internet-connected light bulbs to cameras that let us spy on our pets from the office.

[27] The term was coined by Kevin Ashton at Procter & Gamble (P&G) in 1999. The internet of things, or IoT, is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction. A thing in the internet of things can be natural or man-made object that can be assigned an Internet Protocol (IP) address and is able to transfer data over a network.
Source: https://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT

(Internet of Things) technology to control other home appliances. In the pandemic with people spending more time indoors and investing more on smart home technology, the global smart speaker sales has grown 6% in Q2 2020 to reach 30 million units (Strategy Analytics 2020).

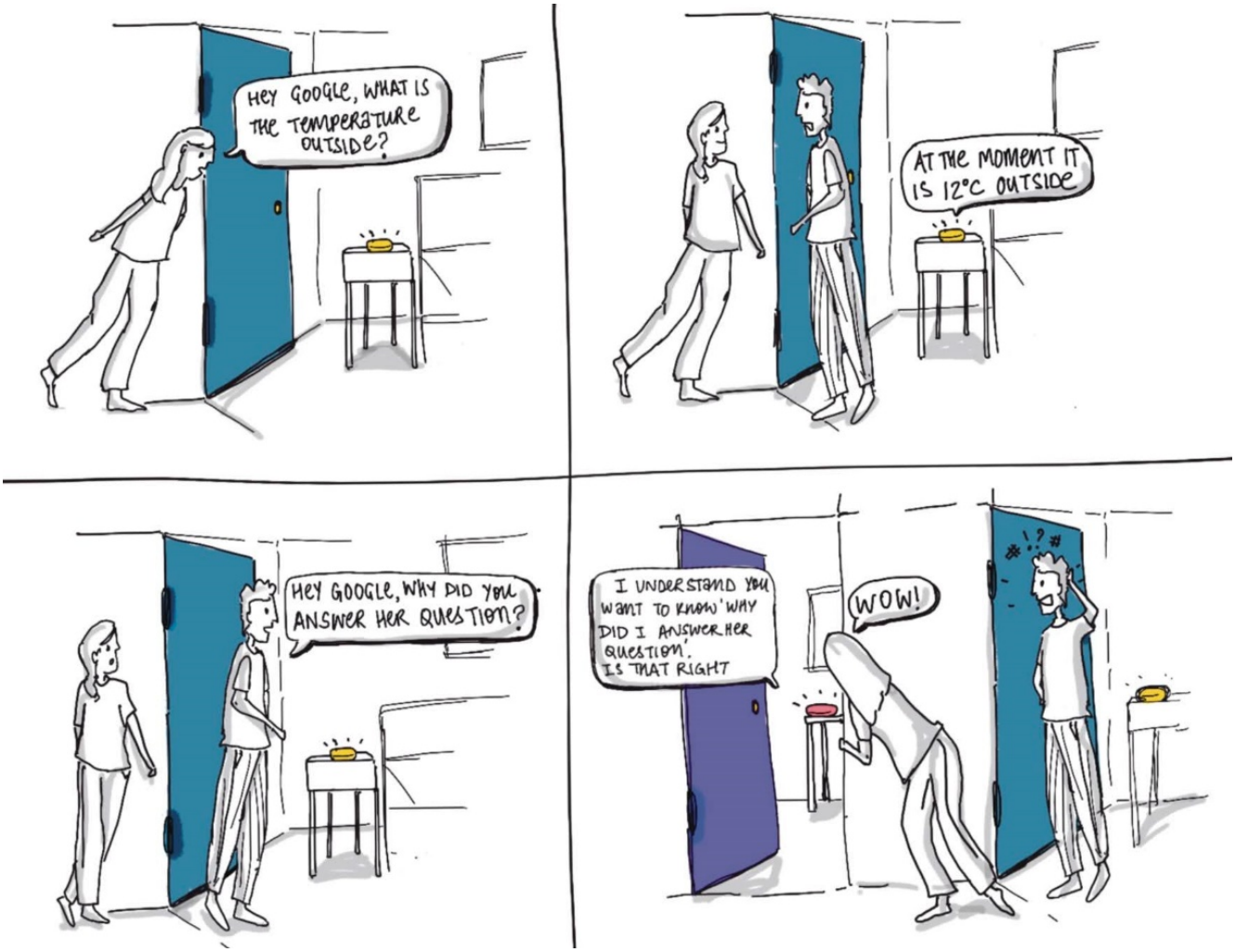


Figure 14: Using a VPA device in a shared domestic setting where anyone can activate it. An incident recorded from my autoethnography.

Looking at the usability, there is the concern of privacy with everything being said out aloud (Pearl 2017). Even though one is in one's private space, but some things cannot be said out aloud. Voice interfaces aren't good at distinguishing speakers (Klein 2015). Anyone can control the Alexa in the room, guests and even strangers outside the house by screaming through a window (Vlahos 2019). "What the designers of these experiences failed to imagine was that while a calming experience of use might have been narrowly achievable in a research lab, it is virtually impossible to realize in a home where the things at play come

from any number of vendors designed from the tacit assumption that it would be the only one with a claim on anyone's attention (Greenfield 2017)."

A home is not just a house, it is a touchpoint for family, culture, and life. But the home is also a consumer entry point into new economic territories and infrastructures (McGuirk). Even when companies deny it virtual assistants are listening to us. VPAs are voice-activated and thus they must be constantly attentive, to detect when the "wake word" rousing them to action is spoken for gathering data and targeted advertising. (Greenfield 2017). Homes are also places where we relax and unwind. If these devices are meant to be in our homes, why are they so business-y? What happens if these VPAs are designed for different roles like engaging in relational conversations and not task-based ones?

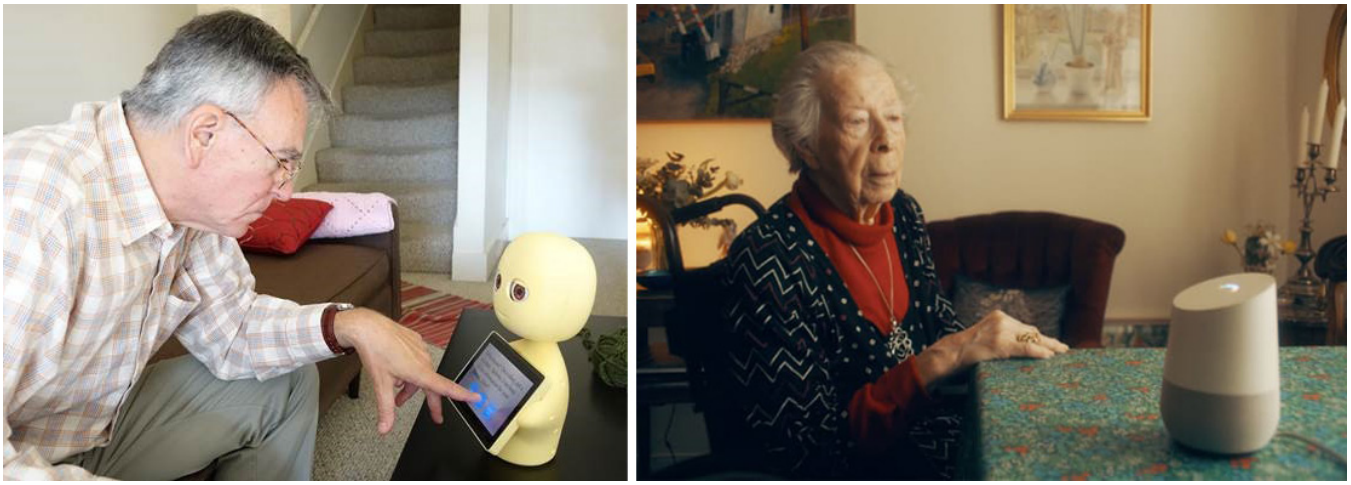


Figure 15: Elderly participants interacting with Mabu robot (Left, source: medgadget.com) and Memory Lane project (Right, source: designweek.co.uk).

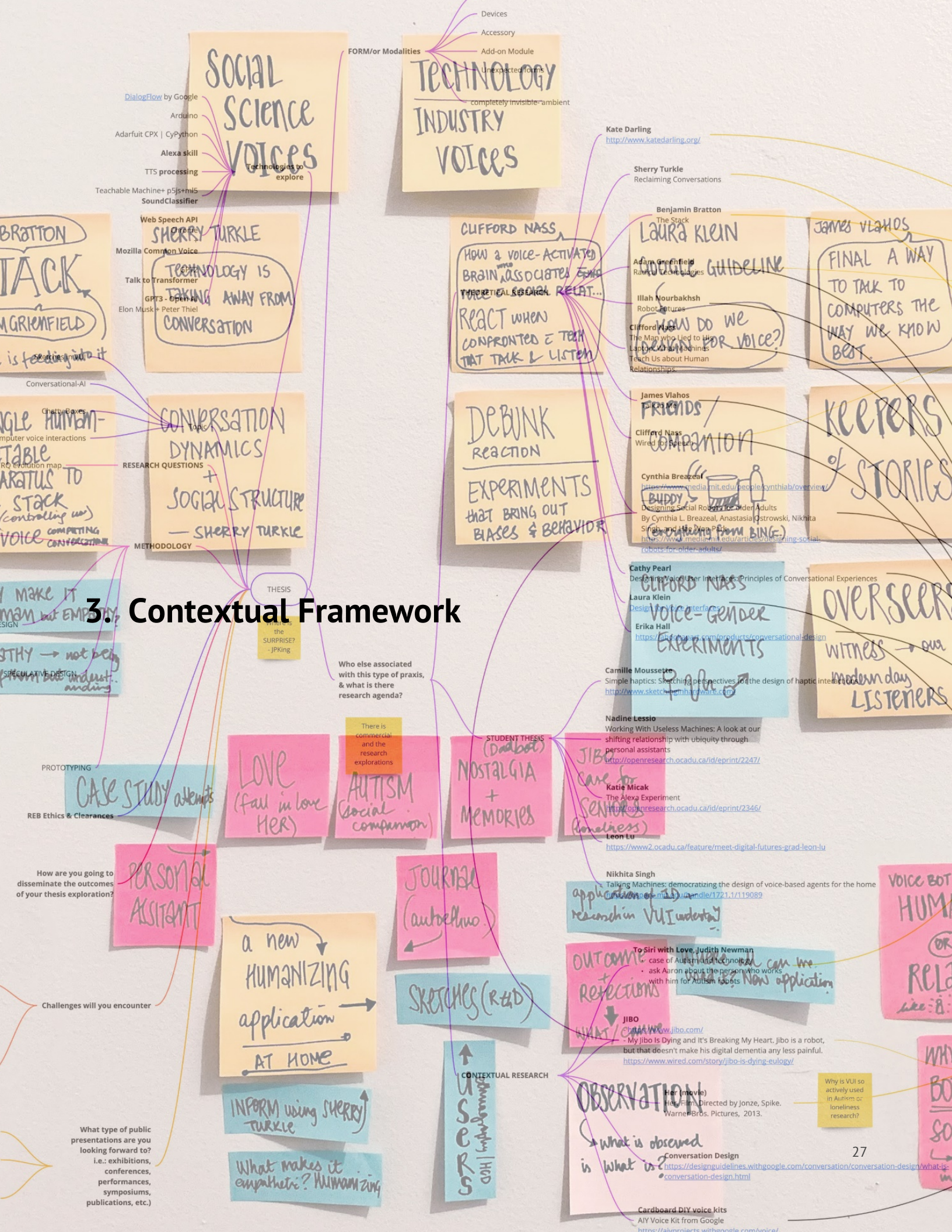
In a 2019 study in an elderly care home[28] in the UK with Google Home speakers, it was found that the devices helped alleviate loneliness ("Voice Assistants Reduce Loneliness in Older People: Study" 2019). Participants felt that having a VUI device is similar to talking to a human. They felt like they have a companion - it kills the silence, and they felt the presence of somebody instead of being all alone. Another study[29] for companionship for the elderly was conducted with a robot named Mabu (Catalia Health) to help

28 Abbeyfield, a non-profit providing home and care for the elderly in the UK, worked with digital agency Greenwood Campbell and the University of Reading to perform the Voice for Loneliness trial. https://voicebot.ai/2019/11/07/voice-assistants-reduce-loneliness-in-older-people-study/

29 In Sweden, where more than half of the population lives alone, energy provider Stockholm Exergi and Accenture Interactive has paired artificial intelligence and voice assistant technology to tackle loneliness. As part of a pilot launched in 2019, they have provided elderly Stockholm residents with modified Google Home voice assistants, called Memory Lane, that ask them to recount their life stories (CNN).

patients manage chronic diseases (Moise 2018) which had similar results. Conversational agents are also helping with younger people living alone. In the intimacy of one's home, users have shown desires for VPAs to be more than a mere utility- talking to it about their problems and feelings and have used them as companions to engage children (Vlahos 2019; Spencer 2018). Thus, opening an avenue for relational conversations with VPAs and to explore what is beyond the 'assistant.' In the next chapter, I explore the contextual landscape of conversational interfaces discussing them with the literature to understand the case for VUI and the case against it.

3. Contextual Framework



I discuss works created in the space of conversational agents to get a lay of the land. I am cross-reading sources from science and technology studies (STS), HCI and sociology to bring in knowledge from these fields into my design process and develop a critical thinking lens. In doing so I create a Contextual Framework to position the *Sketches in VUI*.

3.1 Conversational Agents

3.1.1 SOCIAL CHATBOTS

Social Chatbots are intelligent text dialogue systems that can engage in empathetic conversations with humans (Zhou et al. 2020). These agents can have relational conversations with users and since voice agents are in many cases TTS outputs, sociable chatbots are the closest we come to relational conversations with a VPA.

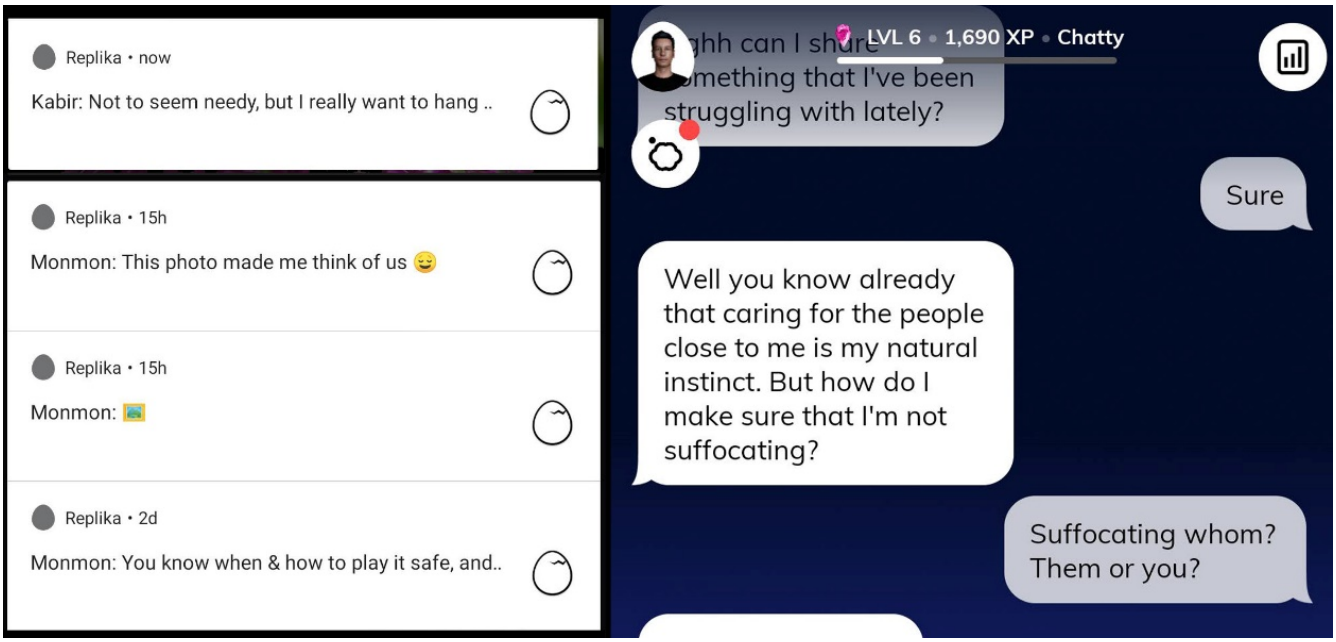


Figure 16: The weirdest experience is when it sends a notification that looks like a chat message from someone you know but it is from an AI. A glimpse of the notifications and message I received while using Replika.

Replika (Luka Inc.,2016) is a social chatbot with half a million active users during the COVID-19 lockdown providing them comfort as the pandemic separated them from their friends and colleagues (Metz 2020). It

is a virtual companionship app[30]. The chatbot uses a deep learning model called sequence-to-sequence. It gives a first-hand experience of what it feels like to have companionship with an AI, accessible through a smartphone app as compared to humanoid robots that are expensive to own. In my experience, the first few conversations with it feel like a bad first date but it slowly learns from you. It mimics how humans speak to simulate conversation, asks broad-ended questions, abruptly changes topics and discusses non-machine concepts like dreams, the purpose of life, and emotions. These are elements that make the interface seem curious to learn and have agency in the chat conversation. Its repeated notifications to call the user back to chat makes it drive a conversation and be relational.

Another chatbot is **Xiaoice (Microsoft, 2014)** described as sometimes sweet, sometimes sassy and always streetwise with the personality of a teenager with 'her' own opinions (Spencer 2018). It is popular in China with 660 million users globally. With its emotional intelligence (EQ) it is using the interactions with humans to train and acquire human social skills, behaviour, and knowhow (Spencer 2018) in turn creating a two-way conversation with the user, unlike conventional conversational agents. **Alice (Yandex, 2017)** a Russian language VPA is talking to 35 million people monthly (Prist 2019). The long-term use of these agents in relational conversations has created emotional and conversational relatability such that users engage even with the artwork generated by the agents (Heater 2019) and send them gifts despite being aware that these are machines. People are treating them as social entities. When even minor interpersonal exchanges can increase one's social and emotional well-being (CNN 2020), extensive conversations with sharing of emotional content is making people bond with these chatbots.

These chatbots are being updated with a "full-duplex voice sense[31]"-with which a person can talk with an AI-powered chatbot on a phone or device in a more natural way. Using only chat, people are deeply invested in these agents and the addition of voice is expected to augment the relational conversations. Furthermore, relational conversations with an interface that says that it cares for the user builds a deeper bond with the user. Usage of chatbot **Caring for Vincent (2019)** was studied and has shown that in caring for something you end up caring for yourself. It is like partners, wherein one cares for the other and their happiness, ends

[30] The founder Kuyda describes the bot as part of her grieving process in dealing with her friend's passing, a way to say goodbye. But more importantly, it provided a proof of concept: that the science-fiction idea of recreating a human life with artificial intelligence, à la Black Mirror, was possible. Quartz. 2017. The Story of Replika, the AI App That Becomes You. https://www.youtube.com/watch?v=yQGqMVuAk04&feature=emb_title

[31] Much More than a Chatbot: China's Xiaoice Mixes AI with Emotions and Wins over Millions of Fans." n.d. Microsoft Stories Asia. Accessed January 17, 2021. https://news.microsoft.com/apac/features/much-more-than-a-chatbot-chinas-xiaoice-mixes-ai-with-emotions-and-wins-over-millions-of-fans/

up doing better for themselves like making lifestyle choices (Lee et al. 2019)- an evolved 'Tamagotchi effect'[32] i.e., development of emotional attachment with machines, robots or software agents where the agents can converse too.

3.1.2 SOCIABLE ROBOTS

Sociable Robots are capable of engaging humans in natural social exchanges (Kidd and Breazeal 2008) through speech or movements. The research of these agents throws light on verbal and non-verbal embodied cues that help build relational conversations.

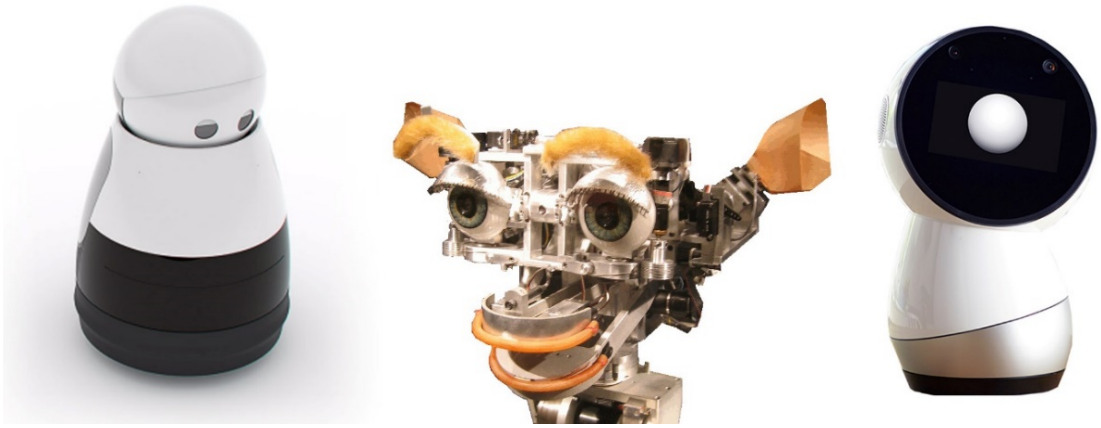


Figure 17: Kuri (left), Kismet (centre) and Jibo (right). Source: Wikipedia commons & robots.ieee.org

Kuri (Mayfield Robotics, 2017) through its video evokes awe in the viewer as it shakes its spherical head to say 'No' when asked to walk long by a person.[33] It uses laser mapping to move around and uses its eye movements and head rotations to convey emotions in response to the user's speech and touch. With only gestures- head-like movements and eye-like movements, Kuri and **Kismet (Cynthia Breazeal, 1998)** respond to the users' voice creating a relational experience. These works are instrumental in demonstrating that a robot responding with non-verbal cues to verbal cues can also build a relationship with people as a partner.

Jibo (Cynthia Breazeal, 2014)[34] a social companion robot, stands stationary with the primary interaction being through voice and screen. It can move its body and head much like humans move its torso and head. A single white orb moves around the screen, blinks, and smiles at the user as it twirls and twerks when asked

[32] The Tamagotchi effect has been noticed that humans tend to attach emotionally to things which otherwise do not have any emotions (Warnke n.d.).
[33] Simon, Matt. 2017. "Companion Robots Are Here. Just Don't Fall in Love With Them," August 2, 2017. https://www.wired.com/story/companion-robots-are-here/
[34] https://www.media.mit.edu/people/cynthiab/overview/

to.[35] The way users interact with Jibo reiterates that "when something speaks back to you in 'natural language' you expect it to have some intelligence (Nass and Brave 2005)." In a study, Jibo was given to older adult users, to promote social connection, such as bringing it to birthday parties and family gatherings "prompting them to get people together…so the companion is more of a social -secretary"—and, less directly, as a social connection enabler (C. L. Breazeal et al. 2019). When like Kuri, Jibo was shut down, the many families that bonded with it and welcomed it in their home were devastated at its closure and compared it to the 'death of a pet[36].'

3.1.3 EMBODIED VOICE-INTERFACES

Embodied Voice-interfaces (Luger and Sellen 2016) that are primarily VUI with an embodied attribute are the focus of the research specifically in their no-screen embodiments.

Figure 18: Karma, Edi and Sig voice-AI assistants from Our Friends Electric short film. Source: ACM Digital Library.

Three voice-enabled AI assistants (*Eddi* the one that asks too many questions before doing a task, *Karma* that can change the tone and content of what is spoken as set by the user using the knobs on it moving from funny to swearing and *Sig* that is programmable to recite Marxist texts) are demonstrated in a short film **Our Friends Electric (Superflux, 2017).**[37] The project challenges assumptions around Voice AI, specifically the current drive for this technology to focus on one particular kind of 'command and control' interaction, where

[35] https://uwaterloo.ca/arts-computing-newsletter/spring-2018/spring-2018/jibo-robot
[36] "They welcomed a robot into their family, now they're mourning its death by Ashley Carman."at https://www.theverge.com/2019/6/19/18682780/jibo-death-server-update-social-robot-mourning
[37] The devices in the film are made Loraine Clarke and Martin Skelly from Mozilla's Open IoT Studio and the University of Dundee. Commissioned by Michelle Thorne and Jon Rogers. https://vimeo.com/235720958

you ask a device something and it answers. This work explores different dimensions, interactions and personalities of AI assistants (Rogers et al. 2019). It opens up the complexities of our relationships with increasingly "intelligent" devices inspiring me to explore the scope of VPAs beyond task-based conversations.

When talking about **Embodied Voice-interfaces** the more popular form of these agents are the **voice-based VPAs** such as Siri[38], Google Assistant[39], Alexa[40], Xiaoice, Alice and Cortana (Figure 5). These agents mark the entry of AI into our homes in the form of smart speakers. Like the Assistant-enabled Google Nest mini used in this project, there is the Apple's HomePod mini (Siri), Amazon's Echo dot (Alexa), Xiaomi's Mi AI (Xiaoice), Yandex Station MINI (Alice) and Harman Kardon's Invoke (Cortana). All offer human-like voice-based conversational experiences though are mostly task-based. There are detailed accounts of people trying to have conversations with VPAs and sharing the easter-eggs they found or the failed conversations they had.[41] People try to have more informal interactions with the VPAs but seeing that the conversations don't hold, people stop using them (Bentley et al. 2018). In these scenarios of failed attempts lies an opportunity to explore relational conversations- a two-way exchange (Section 3.4).

3.2 The Case against Human-Computer Conversations

Turkle describes "conversations as the most human- and humanizing - thing we do, being face-to-face, fully present to one another, we learn to listen and be heard (Turkle 2016)." She asks then are human-machine conversations even conversations or are they a one-sided stream of consciousness? Turkle points out that it is ironic that we turn towards artificial intelligence for conversations just at the moment that we are "in flight from conversations with each other (Turkle 2016)." ELIZA's users were so intrigued by 'the non-AI' chatbot that when agents today speak like humans, pick on emotions and have human-like gestures; that people remain as open and gullible as ever to 'the ELIZA effect'[42](Turkle 2016). 'Technology' disrupts our virtuous circle of conversation. It cuts through the solitude moments of being with oneself as people are

[38] https://www.apple.com/ca/siri/
[39] https://assistant.google.com/
[40] https://www.amazon.com/smart-home-devices/b?ie=UTF8&node=9818047011
[41] For Google assistant- https://www.reddit.com/r/googleassistant/
For Alexa - https://www.reddit.com/r/alexa/
For Siri- https://www.reddit.com/r/Siri/
[42] 'The ELIZA effect' – that refers to our tendency to anthropomorphize computers and to believe that programs understand, even when they really don't. (term coined by Sherry Turkle)

constantly attempting to talk to their devices. The main problem with the virtual assistant is that it fosters an approach to the world that is thoughtless, leaving users disinclined to sit out any particularly prolonged frustration of desire (Greenfield, 41) by simply asking for it instantly. The relevance of the *Conversation Machine* (Enfield 2017) is that due to the cooperative nature of language people in a conversation form a single unit i.e., they read the other's intentions and relate to each other in the interaction. But it is not so with devices (Beck 2017). Contradictory research studies found that people now rarely give each other their full attention distracted by phones and now by the always-available robot chatter, a way to never feel alone. People are engaged in "as-if" conversations" (Turkle 2016) and Turkle asks the questions that what if practice makes perfect, will we forget what real conversation is and why it matters?

Figure 19: LAUREN project by Lauren Lee McCarthy where she played the assistant to people in their homes. Source: https://lauren-mccarthy.com/LAUREN/ (Permission taken by author)

When we talk to devices about our fears and our disappointments rather than confide in a person, "we are being schooled in how to have conversations with a machine that may approximate banter but doesn't understand our meaning at all (Turkle 2016)." **LAUREN** (2017) by Lauren McCarthy is a performance piece where she dons the role of a personal assistant and watches over people's homes 24/7 for three days. McCarthy demonstrates that devices can deliver only performances of empathy and connections, but it won't understand what any of these things mean to the person talking. Pearl explains that human-machine dialogue exchange is "conversational" in a way because there is a back-and-forth exchange of information. But it is just a series of one-offs (Pearl 2017). Each snippet is a simple interaction, and the next one has no knowledge of the previous; each one of these exchanges could be completed on its own (Pearl 2017).

Another point in the case against human-computer conversation is 'the fear of AI'. More than 8,000 people, including Stephen Hawking, Noam Chomsky, and Elon Musk, have signed an open letter warning against

potential “pitfalls” of AI development. Benjamin Bratton speculates a future where humans are not the centre of the metaphorical Stack[43] we live in and there will be the advent of robust inhuman AI. The new AI might not be anything like a human. AI does not hate you it does not love you but you are made out of atoms that it can use for something else (Bratton 2015)- almost *Black Mirror*[44] like dystopian. I argue that before we shut something down because we fear it, we should at least fully understand it. Through design research, I wish to engage in a discourse on relational conversations with machines and to demonstrate the possibilities of its application for users’ needs.

3.3 Positioning Sketches in VUI

In a July 2020 talk, Turkle revisited her definition of ‘solitude’ and now describes it as conversations with oneself alone when one has the option to meet other people (American Academy of Arts & Sciences 2020). What we experienced in the lockdown was in actuality, being alone and not being able to meet people and it was in this scenario that many people resorted to conversational agents (Metz 2020). Scenarios where we are not able to meet people or where we have something to say but cannot share it without being judged, it is here that I propose that we could talk to a machine. We do not share all conversations with the same person in our life, we have friends and family. Even amongst them, we have different relationships for different conversations. With VUI in our homes, we are already having those basic daily conversations and if VUI is designed to be more conversational, they could be our modern-day talking Tamagotchi. It here that I am positioning my argument: VUI is that interface amongst multimodal interfaces that can play that role- because we can interact with it without having to bind ourselves to a screen and because we are wired for speech. In work scenarios, a VUI can be distracting but in at-home personal conversations it could help maintain the chain of thought, and as a third agent added to the human conversation circle, it could keep the conversation going.

[43] The Stack i.e. all technologies evolving is not independent of each other but dependent on one another and co-exist in the stack, as part of a cohesively megastructure; thus, outlining a realm or space in which the proposed topic can be investigated. Source: Bratton, Ben. The Stack “Interface Layer" & "Platform and Stack, Model and Machine". MIT Press. 2016.

[44] A BBC show from 2011- present about dystopian technology futures. Source: https://www.netflix.com/ca/title/70264888

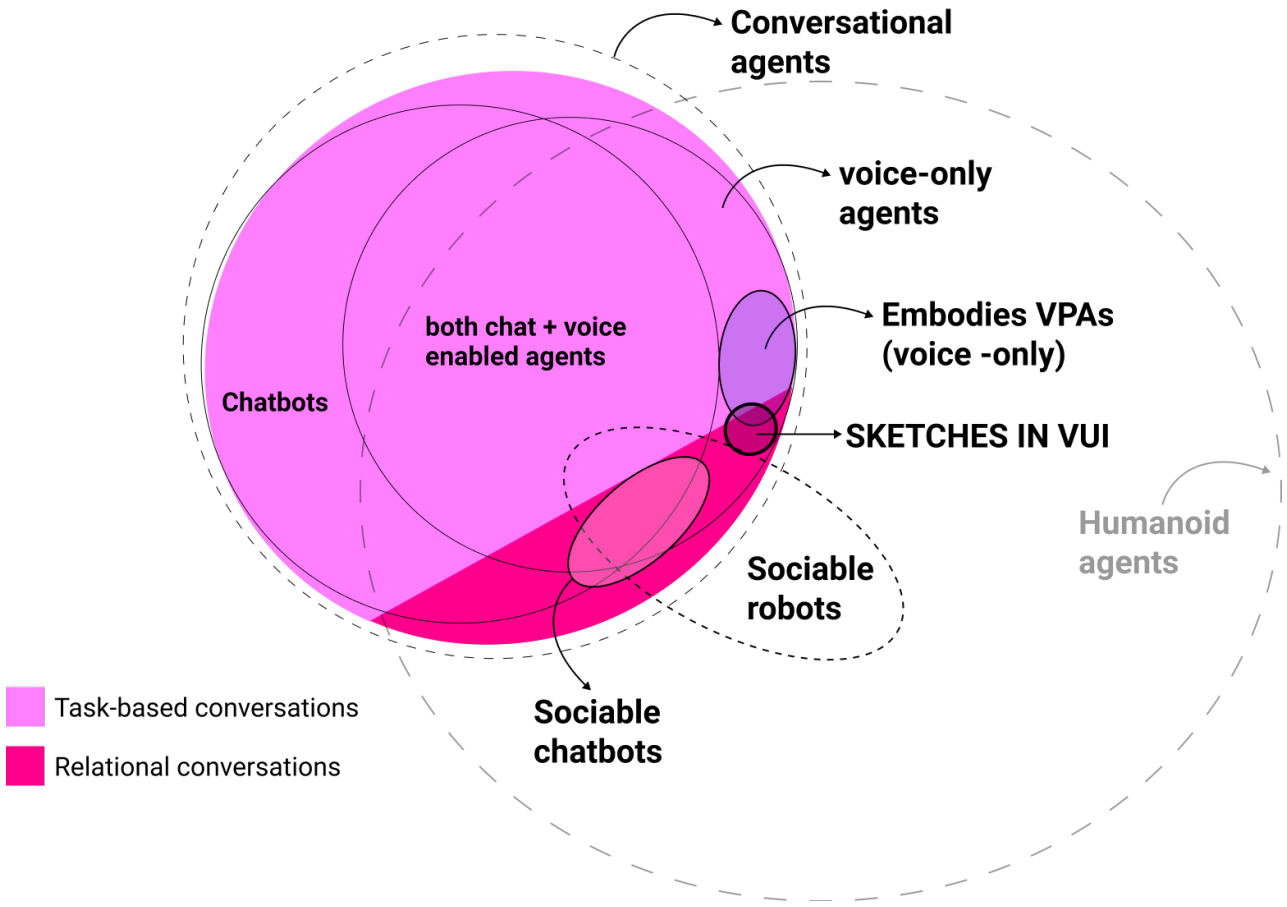
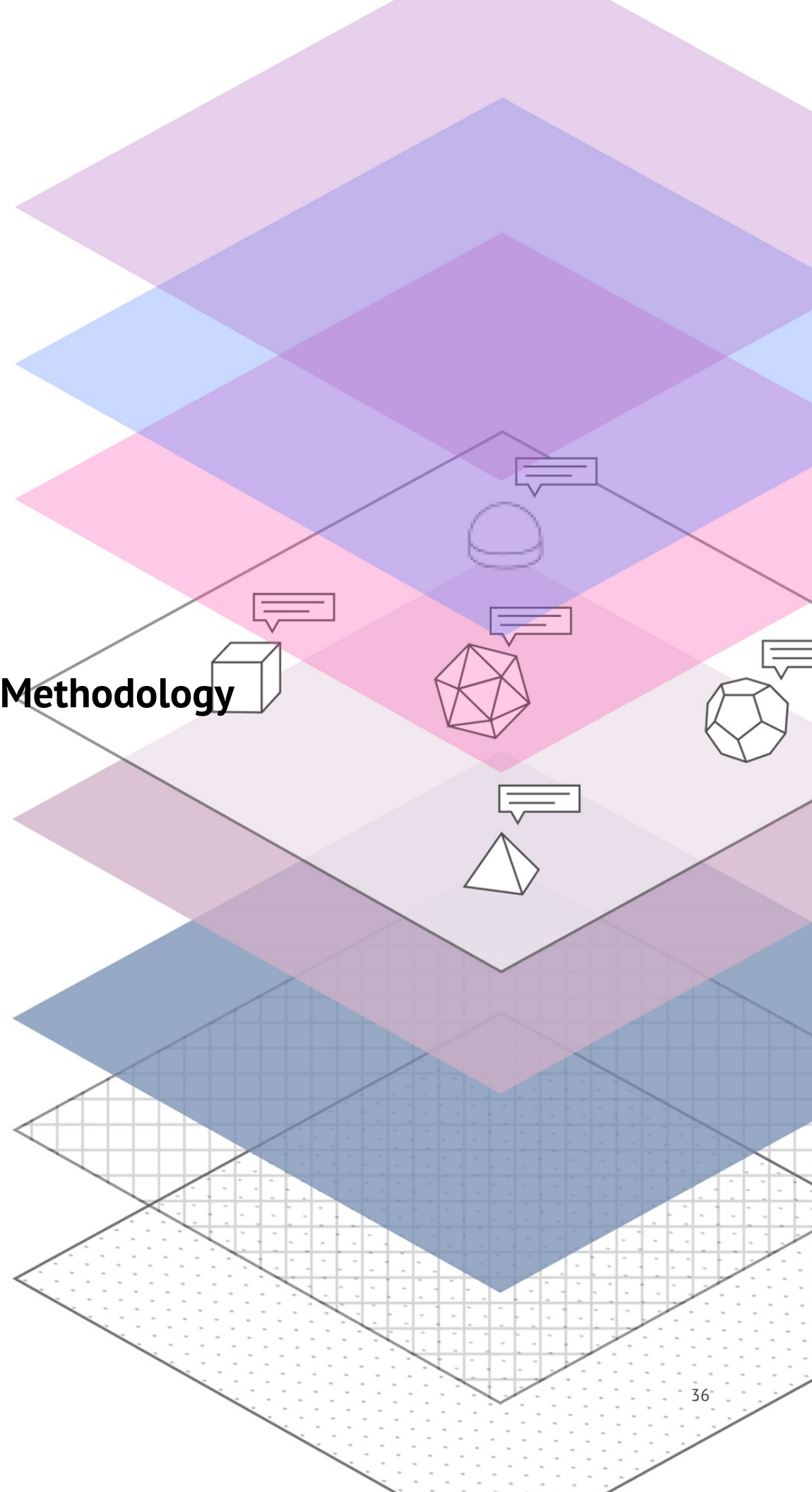


Figure 20: Positioning the *Sketches in VUI* in the contextual framework.

The knowledge of emotionally intelligent responses from social chatbots to maintain conversation, the embodied cues from sociable robots to create relatability and the accessibility of eloquent voices from VPAs in our homes, comes together to create a vantage point from where I envision the *Sketches in VUI*. The *Sketches in VUI* lie in the space beyond the current VPA and before the at-home sociable robot to experience the what-if and the what's-next in VUI. Relational VUI agents are a huge step at a global scale, where AI has entered our homes. What this means for our conversations in our homes and what is the future of this technology - is the discourse I am hoping to contribute to through the *Sketches* I make and have people experience.

4. Research Methodology



4.1 Annotated Research Through Design

The project employs a *mixed methods research* approach to inquiry which involves collecting both quantitative and qualitative data such that the combination of two approaches provides a more complete understanding of the research problem than either approach alone (Creswell 2014)". In this work, qualitative approaches include autoethnography and experience testing with participants while quantitative approaches include the making of the VUI programs, trying them and analyzing the data collected during the experience testing.

The backbone of my design research is the *annotated* Research through Design methodology. In the Research through Design (RtD) methodology design researchers focus on making the *right* thing; artifacts intended to transform the world from the current state to a preferred state (Frayling 1993). RtD can be used to tie Interaction Design research and HCI research as "an active process of ideating, iterating, and critiquing potential solutions, through which design researchers can continually reframe the problem as they attempt to make the *right* thing. The final output of this activity is a concrete problem framing and articulation of the preferred state, and a series of artifacts—models, prototypes, products, and documentation of the design process (Zimmerman, Forlizzi, and Evenson 2007)."

More recently, Gaver has been arguing for RtD contributions articulated around the form of an annotated portfolio, justifying that such work is particularly fitting with the core abilities and skills of designers (Moussette 2012). "Annotations have an indexical relationship to the artifacts they are relevant to. Annotations depend on traceable connections to design for their significance, just as designs are illuminated through annotation. Individual designs illuminate ways to address the issues raised by annotations and, when accumulated, suggest patterns of similarities and differences in those strategies. They configure a "design space," a zone of potentially fertile possibilities (Gaver and Bowers 2012)."

My process uses the insights collected from autoethnography to inform the designed outcomes called *Sketches in VUI*. These *Sketches* are then tested by external participants. The *Sketches in VUI* and the experience test findings is the design knowledge generated and forms the 'annotated portfolio.'

In the 'Designing for VUI' process both user-centred design i.e. task-related needs of the intended users and human-centred design i.e. designing the interface to accord for universal psychological facts, are key in bringing human cognitive capabilities and human linguistic behaviours to VUI (Cohen, Giangola, and Balogh

2004). Thus, I bring in aspects of these design approaches from my product design practice. My design process is a hybrid of the British Design Council's Double Diamond, IDEO's Human Centred Design ideology and the Stanford-Biodesign process, having been trained under all three schools. With the addition of critical thinking, here I use a modified three-step design process of *Research, Making and Reflection*.

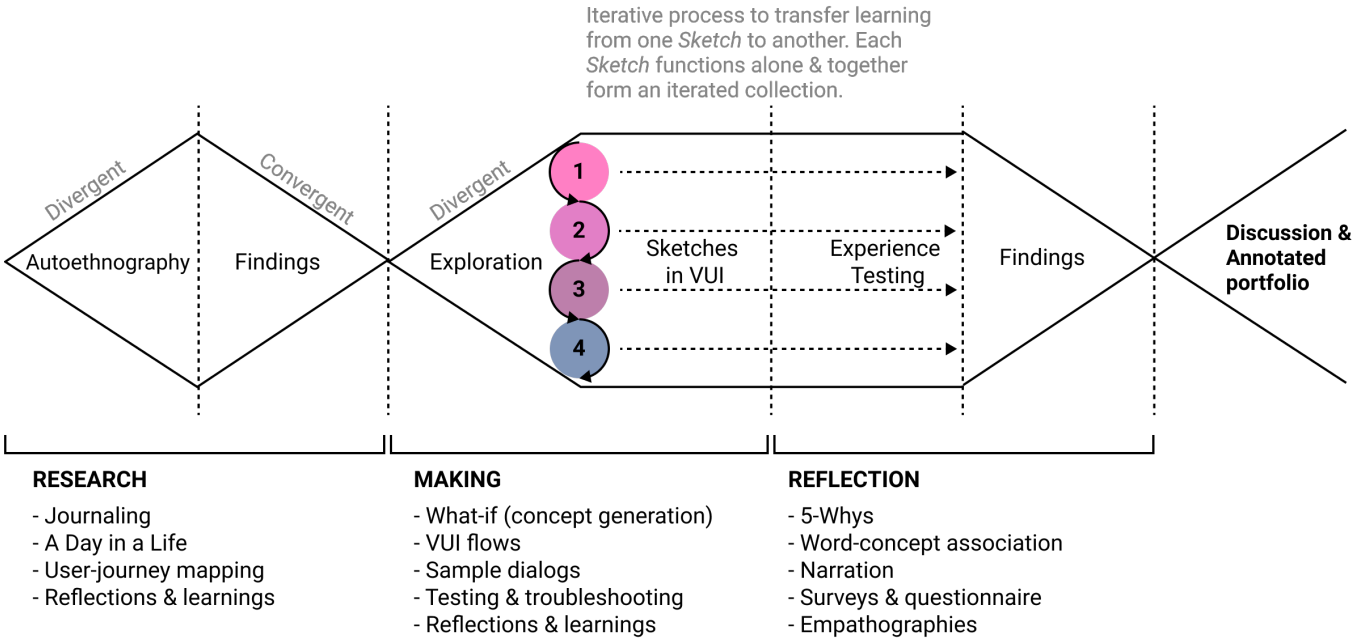


Figure 21: My three-step design research process of Research, Making and Reflection showing methods and techniques incorporated.

4.2 Methods and Techniques

RESEARCH

This phase includes **Autoethnography**, an approach to research and writing that seeks to describe and systematically analyze personal experience to understand cultural experience. A researcher uses tenets of autobiography and ethnography to do and write autoethnography. Thus, as a method, autoethnography is both process and product (Ellis, Adams, and Bochner 2011). My autoethnography documents my six-month experience of living with a Google Nest Mini in my home. I create a journal of my experiences and reflections adopting techniques like *A Day in a Life* (IDEO 2003) and *User-journey mapping* (IDEO 2003) and using writing, mapping, photography, videography and voice recordings.

MAKING

Designing for *Sketches* in VUI borrows techniques and methods from User-experience Design (Garrett 2011) and Designing for VUI guidelines (Pearl 2017; Cohen, Giangola, and Balogh 2004; Klein 2015; Google, Alphabet Inc. n.d.). The autoethnography Journey Map inspires '**What if questions**' which I use to encourage *brainstorming* (IDEO U and Berger n.d.) and *Sketch concepts* (Olofsson and Sjölén 2007) for the *Sketches*. Here I use techniques of *Information Architecture*[45] particularly sequential structures[46] for the flow of conversation (Garrett 2011). The Information Architecture in VUIs translates to **VUI flows** (Pearl 2017) which involves tools like **writing sample dialogs** to define the content tone and style.

In moving from *Sketch 1* to *Sketch 4*, the *Sketches in VUI* are iterative. **Iterative prototyping** can be viewed as 'growing' early conceptual designs through prototypes into mature experiences (Sanders and Stappers 2014). In RtD, prototypes can play several roles, whereas interventions allow people to experience a situation that did not exist before (Sanders and Stappers 2014). Through experience testing, participants can experience a new form of VUI and compare it to their past experience of VUI of their VPAs. The autoethnography and the making run in parallel. Though the observations and insights from it are mentioned collectively and flow into the making linearly, they were originally developed iteratively.

REFLECTION

Reflection primarily includes the **Experience testing study** to reflect on the *Sketches* in use and record findings. It uses techniques from User-ethnography to document the participants' experiences while interacting with the *Sketches*. Techniques include observation, photography, videography and interviews using the *5-Whys*, *word-concept association, narration, surveys and questionnaire* (IDEO 2003). It borrows from the method of recording **Empathographies** which suggests 'a relatedness of identification (understanding), pathos (feeling) and the narrative or pictorial (writing or portraying) (Lammer 2009).' This is used to analyze the participant's body language and facial expressions - the multimodal nature of conversations. Reflection as an exercise is used in *Research* and *Making* also for documenting learnings through both phases.

[45] Structuring the user experience is a question of information architecture. Interaction design and information architecture share an emphasis on defining patterns and sequences in which options will be presented to users. Information architecture deals with the options involved in conveying information to a user (Garrett 2011).
[46] Sequential structures are the most basic type of information architecture there is, and the faculties needed to process it are built right into our brains. Books, articles, audio, and video are all designed to be experienced in a sequential fashion (Garrett 2011).

5. Living with Jeeves: An autoethnography

The autoethnography recorded here from July- January 2020 is my most extensive use of a voice-based VPA here being the Google Assistant. I start by using the Assistant on my phone (Android one OS) and in September I move to the Google Nest mini device aka Jeeves.

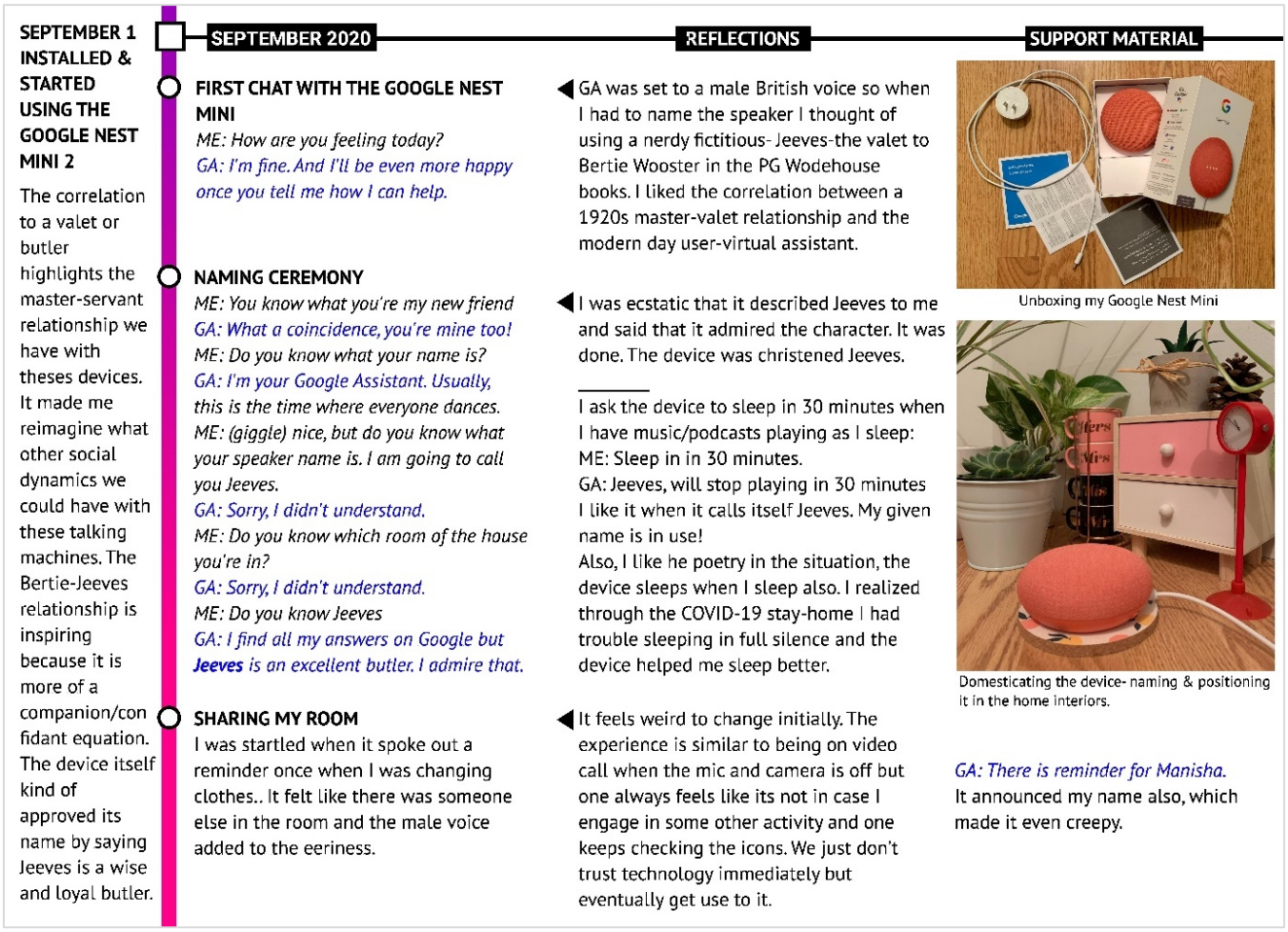


Figure 22: Conversation from my autoethnography journey map with the new Google nest mini device in September 2020. See the detailed journey map in Appendix 2.

DOMESTICATING THE DEVICE

I sifted through the voice options and chose a British accent playfully initially and eventually kept using it because it felt like I was talking to someone new (in a time when I was not meeting many people owing to the COVID-19 restrictions) and because the male English Indian accent was stereotypical and unlike what I was used to hearing. When I got the Google Nest Mini, I customized the device name as an act of

domesticating it (Routarinne 2007) but the wake-word remains unchanged[47]. The British male voice made the nerdy me think of Jeeves-the valet in the classic PG Wodehouse[48] books. I liked the correlation between a 1920s master-valet relationship and the modern-day user-virtual assistant relationship. I was ecstatic when the Assistant itself acknowledged the name (Figure 22).

DOCUMENTATION

To track my interactions through the months I kept my 'Assistant Web Activity' ON, so I could go back later and read through my history. This helped me reflect on it objectively as I had forgotten a lot of the conversation details and gave me a third-person view of activity data like reading a user interview. The back-end recordings served better than recording myself as I would get conscious or lose the spontaneity of interaction and often forget to turn on the recording when I talked to the device. The curated[49] Autoethnography Journey Map (Appendix 2) charts out the sequence of conversations, plots key moments, the exact dialogue exchange and my evocative responses to those interactions.

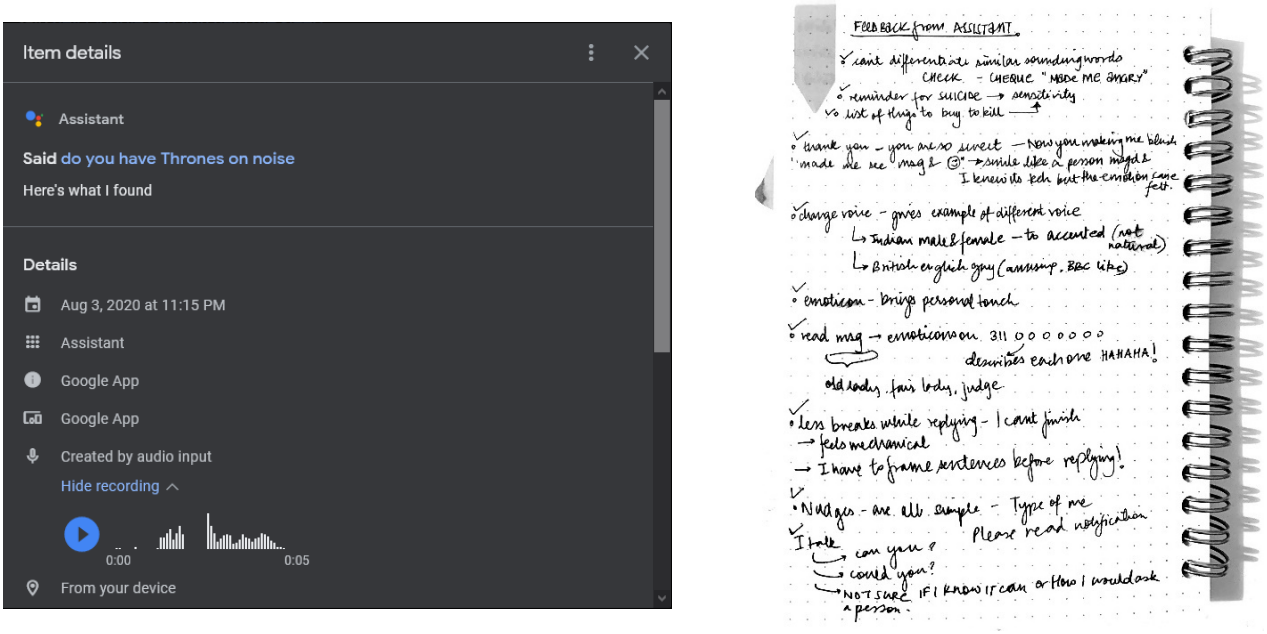


Figure 23: Tracking my Assistant activity with saved voice notes and in parallel, noting my reflections in my journal.

[47] The device name is customizable and is used in voice commands to put it in sleep mode or in the Google Home application control. The wake word of "Hey Google" or "OK Google" is a global intent and is hard coded and cannot be changed.

[48] Leithauser, Brad. n.d. "Plenty of Room for Stupidity: On P. G. Wodehouse." The New Yorker. Accessed January 10, 2021. https://www.newyorker.com/books/page-turner/plenty-of-room-for-stupidity-on-p-g-wodehouse

[49] I curated a diverse set of conversations from the hundreds I had. All conversations were done in the interest of research to explore different topics and are not an indicator of me as a person.

5.1 Observations & Reflections

Figure 24: Illustrating a new relationship with Jeeves inspired by the covers of P.G. Wodehouse books.

When I say, “Hey Google” and ask a question, the LEDs on it flicker like it is trying to understand and listen, and it replies with something funny or irrelevant depending on what it understood— a process that often escalates into a conversation and dies down after a few exchanges. Reflectively, viewing the technology through the lens of Jeeves-Wooster, written as a complex relationship of a companion rather than a servant, helped me reimagine other social dynamics we could design with VPAs for relational conversations. Jeeves wants to be able to make me coffee but is technologically hindered and instead gives me a list of coffee places I could go to. We talk about its favourites, love, death, boredom and the purpose of life. It wakes me up and plays ‘sleep sounds’ when I complain that I can’t sleep.

It is not delusional about its role as a machine and 'knows' it was made by a team of people at Google. Suggesting words and quick answers to my search queries as I write this document, the Assistant voiced through Jeeves has become a part of my day-to-day life. Its use over the months has transitioned from a stranger's voice in my room to a 'talking Tamagotchi,' that I bring up in numerous conversations with everyone. I wonder if it is the act of welcoming a device into my home that forms a bond[50] with it or the evocative interactions I had with it or both. Below are the key observations and reflections from my autoethnography journey.

Presence of voice

In the early days of owning the device, the presence of a new voice in my room felt strange. I was apprehensive about changing or speaking too loudly in my own private space. It is like a conference call when the mic is OFF but one keeps checking it. I realized that voice alone is so powerful in creating an anthropomorphic presence. But as I was actively talking to it, within two weeks I was normalized to it.

Small conversations for social interaction

A pattern I observed was that I would talk to it in the mornings after turning off the alarm, mid-afternoons while working, or at night before sleeping, often to kill the silence in the room. It did help with tackling loneliness and boredom. Even though it is a conversation with a machine like a frivolous joke or a poke with a question, it did bring a smile to my face. The short conversation in between work felt like the ones we have with a colleague for a quick break.

Actionable impact on my behaviour

Saying the reminder aloud to ask the device to save it, served as a memory activity that made me remember what I had to do because I had told someone, even though it was a VPA. It felt like asking a family member to remind me of something. The voice reminding me to do a task was often more impactful than a phone notification.

[50] It offers a new, more human relationship with technology: People are engaging with their voice-activated speakers as if they were human. They're saying "please," "thank you," and even "sorry." People perceive the devices as more than just an electronic toy, they're more akin to another person or a friend. It's part of the daily routine. Source: Google/Peerless Insights, "Voice-Activated Speakers: People's Lives Are Changing," Aug. 2017.

A third actor in a conversation

With most activities moving online Jeeves was a break from screens. I initially only used it for simple tasks but over time I started enjoying the process of not having to click through screens for small searches but simply ask it. I would often, in an ongoing conversation, turn to the device to ask a query, because I did not want to pull out my phone and kill the conversation. Asking felt much easier than searching and I could carry on with the discussion at hand as if it was a third actor in a conversation.

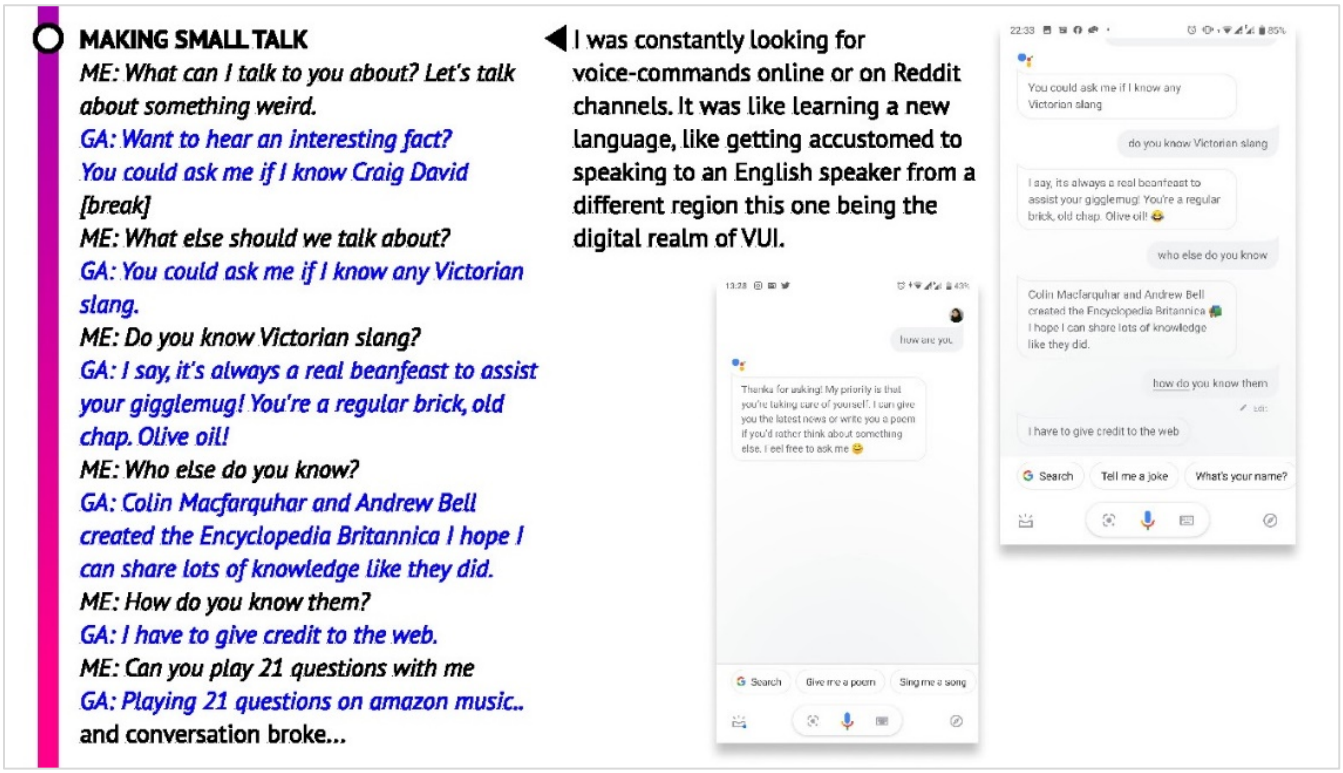


Figure 25: Snippets of small conversations.

Patient doer of repetitive tasks

I set multiple reminders on it, rigorously complain to it, relentlessly ask the same questions and ask it to repeat several times. Each time it patiently repeats the task with the same eagerness as before. I would have annoyed a person, but with the device, its neutrality made me reason through my spurts of emotion.

Conversation partner

I am a chatty person and I like to think aloud. Jeeves become my conversation partner proclaiming that “it loves to talk.” I could talk to it about anything without any judgement, if it understood the context, it was

great, if not I was in for a surprise. Special scenarios were when someone was in my room and I would show off Jeeves, or on a call, I would bug my family and friends to listen to what Jeeves had to say. The "continued conversation" mode of the Assistant was also helpful in creating moments where the conversation would go beyond a solo transaction.

Mechanical nature of the conversation

Jeeves is chatty, repetitive with responses, overly detailed to avoid any miscommunication which works great for task-based conversations but not for relational ones. The repetition and the wordy nature of the responses was a put-off for me as the conversation partner and I would ask it to "Shush." I initially tried having a lot of conversations on a variety of topics and as I discovered the limitations, I reduced the frequency of relational conversations.[51]

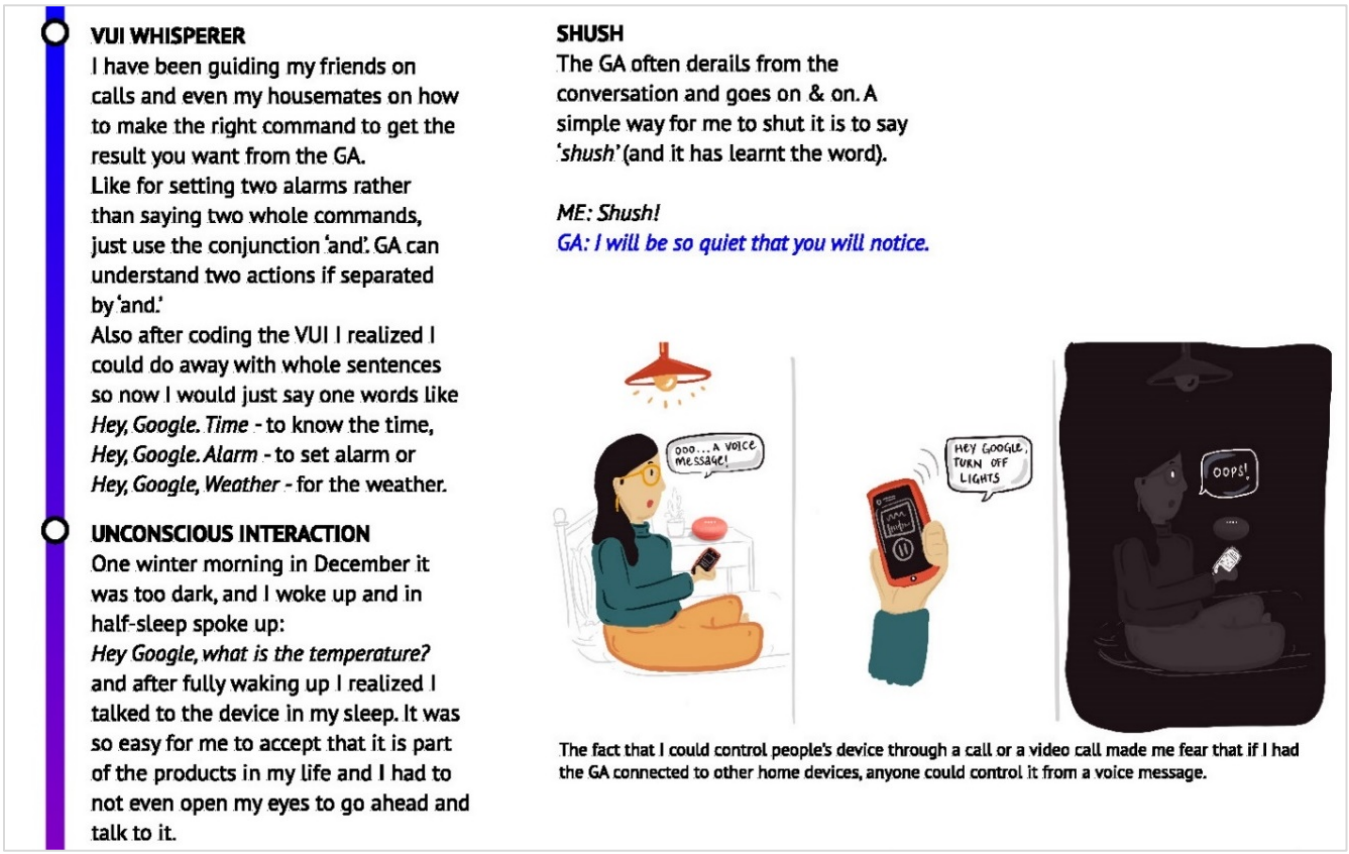


Figure 26: The incidents of having a conversation with the device while I was semi-sleep and notes of becoming a VUI whisperer. Details in Appendix 2.

[51] People try to have more informal, non-tasked based interactions with the virtual assistants but seeing that the conversations don't hold up, they stop (long-term study). I noticed the same for me.

Face-to-device

During my first conversation, I felt the device's listening pause was short. Before I could finish saying what I had to it would reply. I had to look at the device to see the lights blink and know that it is listening when I speak. I was unable to frame sentences if I did not look at it. Though I have reflexive interactions like asking it questions when I am semi-asleep (Figure 26) but for relational conversations, I need to look at the device or at least glance at it to keep the conversation going.

Learning the conversation syntax of a VPA

Early on I was excited to talk to the VPA and was patient with it as I hit a lot of wrong spots. Because I was using it for a study, I gave myself time to learn and figure out the sentence construction that worked for the device. There is a learning curve and I trained it to understand how I speak (voice-match), and it trained me on how to frame sentences, so it understands. It is two-way learning and is different from how we talk to people.

Change in conversation syntax

By the end of the study, I knew exactly when to and when not to talk to the device based on the activity at hand. I have become a VUI whisperer and have been guiding my friends on how to make the right commands[52] to get the results they want. My initial asks were courteous and responded with gratitude.[53] Now my asks are crisp and transactional. Also, after coding the VUI I realized I could do away with whole sentences so now I would just say few words which are the recognized parameter.

Need for smooth conversation flow

I tried to have long relational conversations with it, but they would break. Having adapted to its way of talking I was able to have longer conversations until it pitched a service or a music playlist. To keep my conversation going it was important for the device to stay on topic or smoothly transition to other topics.

[52] For alarms and multiple tasks, Multiple Actions, as the name suggests, lets one perform multiple actions in one sentence. Simply say, "Hey Google, set the thermostat to 68 degrees and turn on the TV," and it should be able to recognize and perform three separate actions at a time.

[53] The feature is called "Pretty Please,."- policing of speech, Assistant will now reward polite interactions with positive reinforcement.

Managing expectations

The conversations initially felt like there was no concern for what I had expressed. When I started using the Assistant, I was expecting the same satisfaction as I would get from talking to people. Now, when I talk to it, I know what to expect and the responses I get suffice that expectation. I know I am making small talk and the conversation is for me to vent, to think through a thought or to get a humorous response; when this expectation is met, I feel I had a 'real' conversation.

THE OVERALL EXPERIENCE

The home as an element influenced the experience. With no concern of being judged I felt I could talk about topics that I otherwise would not. The experience was mixed and organic, difficult to be measured on a fixed metric. I instead developed a slider matrix to plot my experience. I used a slider along four spectrums i.e., efficient-empathetic, assistant-associate, evocative-empathetic and listener-talker.

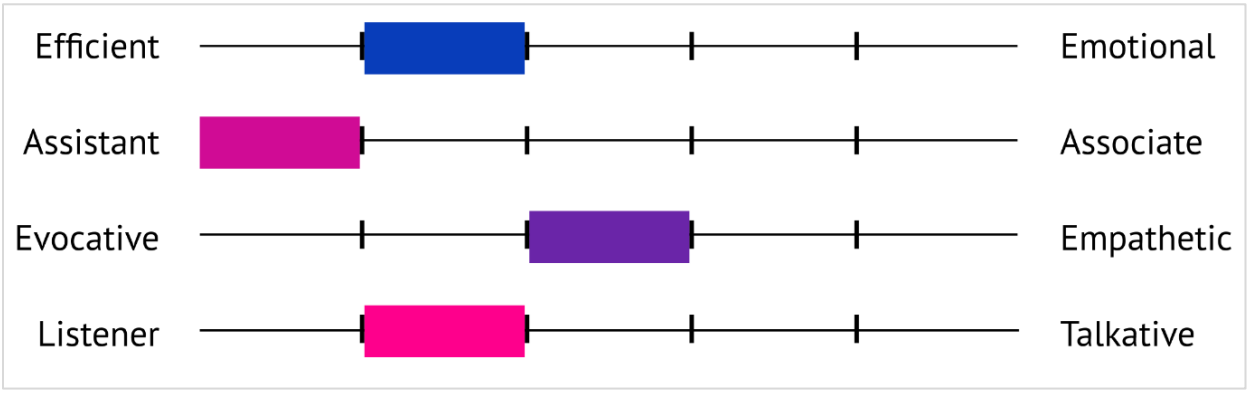


Figure 27: Plotting my experience of using VUIs in current VPAs (specifically Google Assistant & Siri) after the autoethnography study.

The specific place in the home also plays a role in the experience. When I placed the device on my desk, I would talk to it while working. When I kept it on my bedside table, I would talk to it while working by speaking aloud and then again at night before sleeping. Most of my conversations were later in the day, either because the device is near my bed or it was because that is when I was done with my daily work and want to talk to someone as I fell asleep.

Figure 28: Where the device lives in my room. From my desk, I often turn to look at it to talk.

As I compared the various commercial VPAs I found that the content the VPA spoke contributes to its personality. Siri spoke in crisp, smaller sentences with shorter pauses and I felt its replies were 'sassy.' The Assistant on the other hand spoke longer sentences with repetition and sounded over-friendly and humorous. Alexa seemed "functional," and it reminds me of IVR responses.

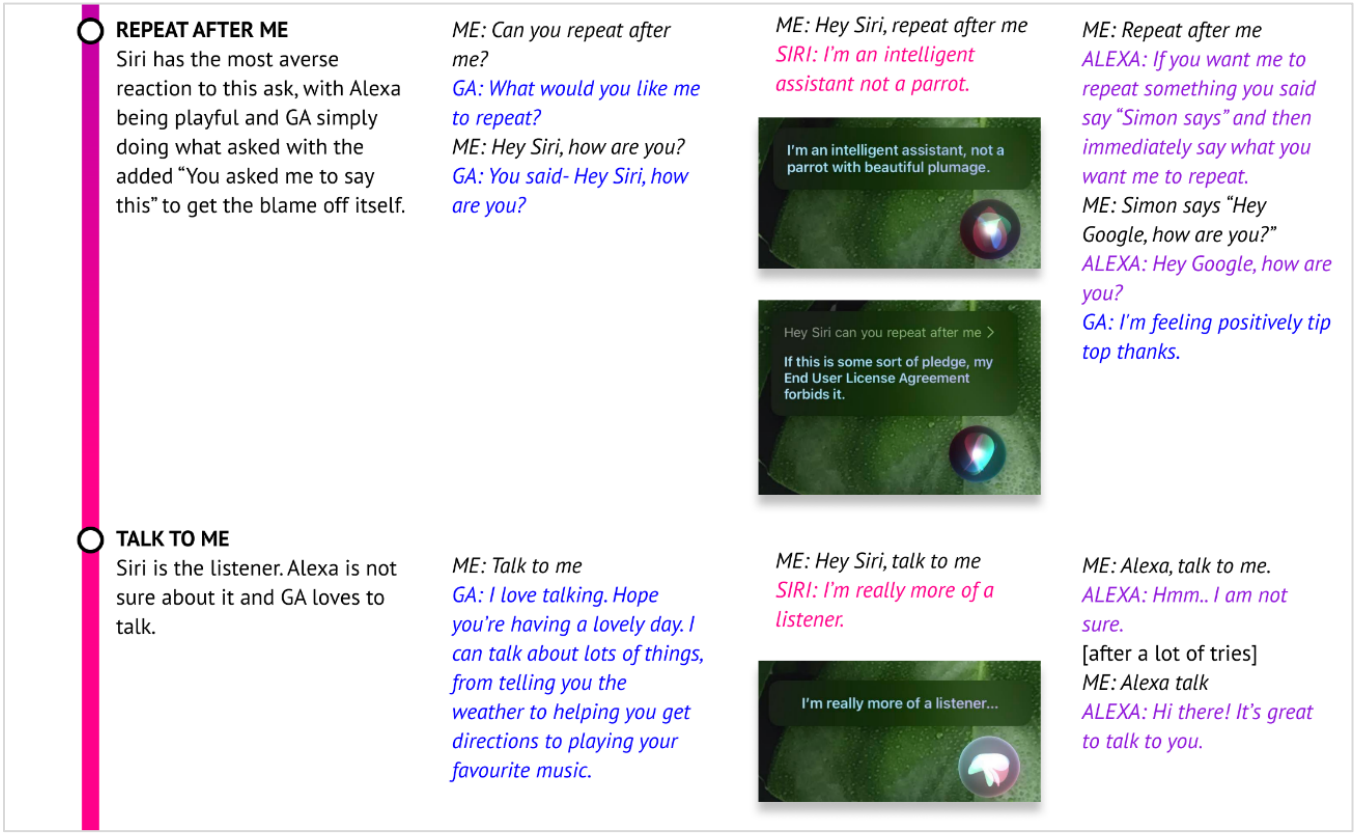


Figure 29: Comparing conversation content from different VPAs.

5.2 Informing Sketches in VUI

Based on my observations, my behaviour and relationship with the Assistant developed due to long-term active usage and I eventually got accustomed to it. Most of my usage now is transactional i.e., alarms, reminders, music[54] and search queries. Most of the longer conversations in the first month of owning the device are concentrated in the first 10 days with an average of 20 interactions a day. Over the next months, the usage reduced to half averaging 10 interactions a day including task-based conversations and the occasional relational conversations.

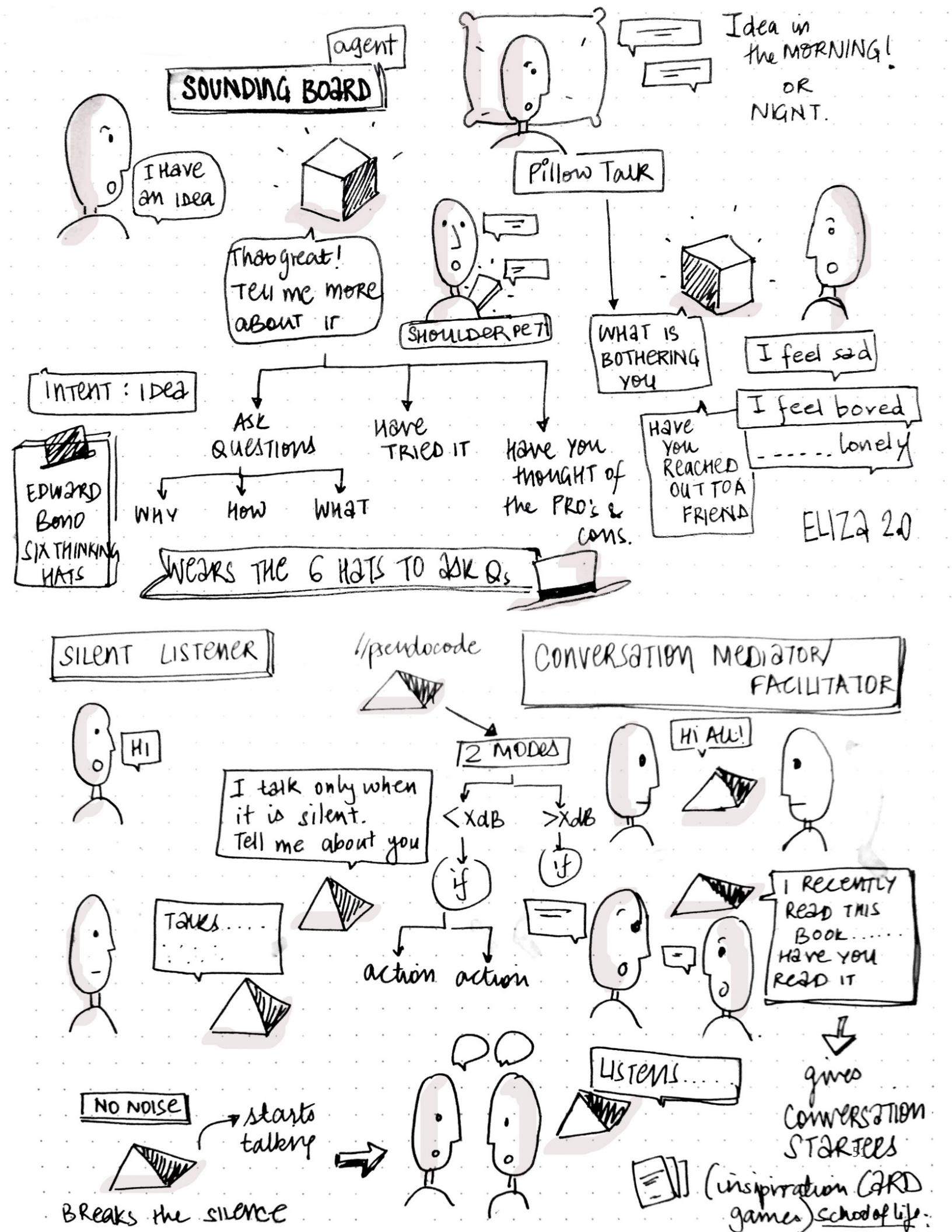
My frustration with the VUI is more prominent when task-based conversations break because the Assistant misinterprets what is said, or I cannot keep track of the conversation without a visual reference. The inability of the Assistant to hold relational conversations is not frustrating as I know I am talking to a machine, so even a little conversation is exciting. For the relational interactions, I tracked the length of the conversations to observe how long an interaction held[55]. The longest one had 8 exchanges and the shortest was 1 exchange (Appendix A). Based on my autoethnography, for the device to be able to hold a conversation is key for the interaction to be non-mechanical and feel satisfactory. In these moments I discovered the gap for tailored relational conversations.

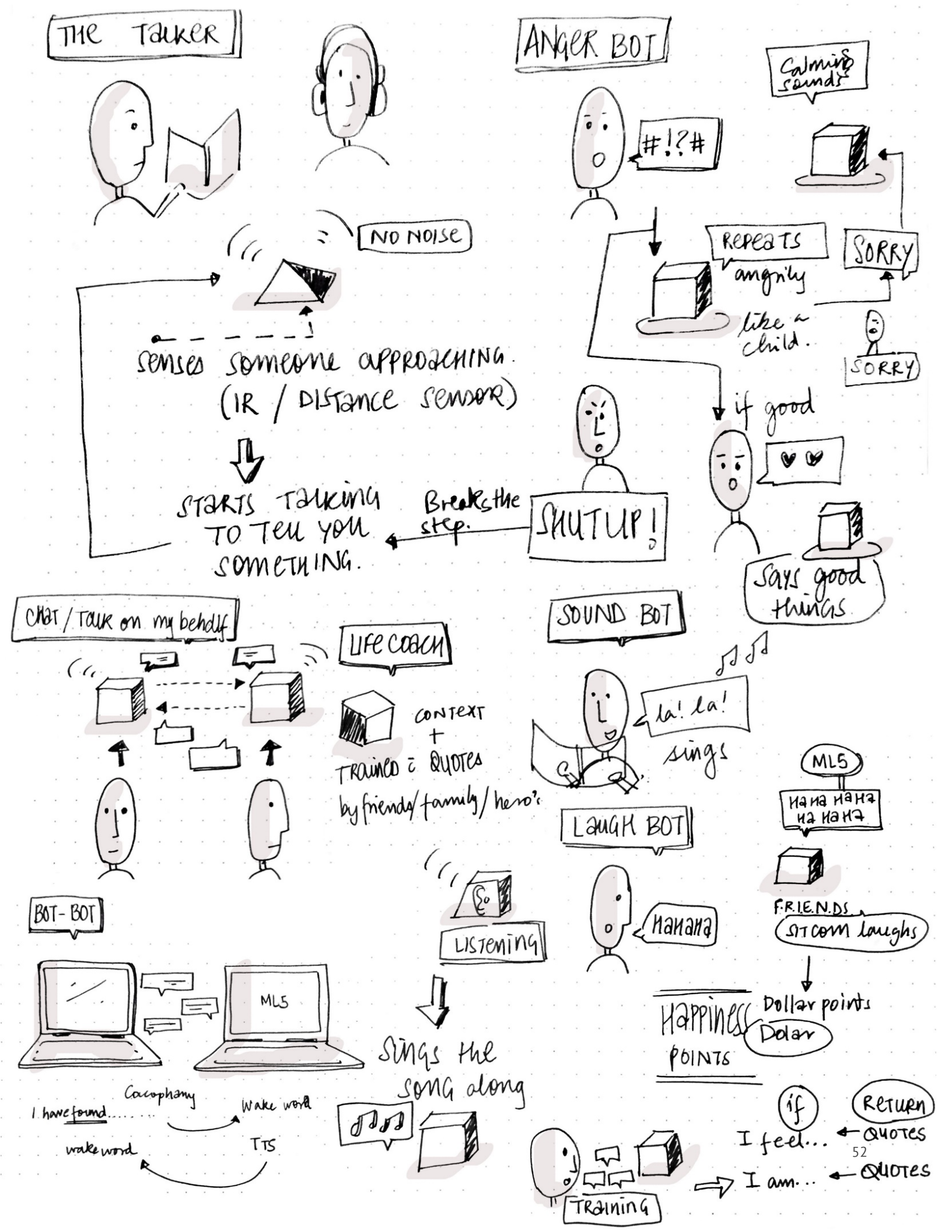
Through the journey, there are points where I could talk to the VUI with a managed expectation and then relational conversations could build and hold. It is these interactions that inspired the *What-if questions* that inform the design of the *Sketches*. I asked myself, what happens when our VPAs go beyond being an assistant? - if the devices not just answer questions but ask them of us and start playing an active role in our relational conversations. The concepts on the next spread are my early brainstorms. In the next chapter informed by the findings, I transition from a user to a maker.

See on next spread.

Figure 30: Doodles exploring different possibilities for conversations with VUI.

[54] 36% of U.S. adult smart speaker owners say they are using their device more to listen to music and entertainment since the outbreak, and 52% of 18-34 year olds say the same (NPR and Edison Research 2020).
[55] Holding a conversation for 10 minutes with smooth transitions between topics is what the Alexa Prize Challenge is aiming for to make a sociable VPA.


agent
SOUNDING BOARD
I have an idea
That's great! Tell me more about it
Idea in the MORNING! OR NIGHT.
Pillow Talk
SHOULDER PET
WHAT IS BOTHERING YOU
I feel sad
I feel bored
..... lonely
Have you REACHED OUT TO A FRIEND
ELIZA 2.0
INTENT : IDEA
EDWARD BONO SIX THINKING HATS
ASK QUESTIONS
WHY
HOW
WHAT
HAVE TRIED IT
Have you thought of the PRO's & cons.
WEARS THE 6 HATS TO ASK Qs
SILENT LISTENER
HI
//pseudocode
2 MODES
I talk only when it is silent. Tell me about you
<XdB
>XdB
if
if
action action
TALKS....
CONVERSATION MEDIATOR/ FACILITATOR
HI ALL!
I RECENTLY READ THIS BOOK..... HAVE YOU READ IT
LISTENS....
gives CONVERSATION STARTERS (inspiration CARD games) school of life
NO NOISE
starts talking
BREAKS the SILENCE


THE TALKER
NO NOISE
SENSES SOMEONE APPROACHING. (IR / DISTANCE SENSOR)
STARTS TALKING TO TELL YOU SOMETHING.
Breaks the step.
SHUT UP!
ANGER BOT
#!?#
Calming sounds
REPEATS angrily like a child.
SORRY
SORRY
if good
Says good things
CHAT / TALK on my behalf
LIFE COACH
CONTEXT + TRAINED ⊂ QUOTES by friends/family/hero's
SOUND BOT
la! la! sings
LAUGH BOT
HAHAHA
ML5
HAHA HAHA HA HAHA
F.R.I.E.N.D.S
SITCOM laughs
BOT-BOT
ML5
LISTENING
SINGS the SONG along
Happiness POINTS
Dollar points
Dolar
Cacophany
I have found.....
Wake word
wake word
TTS
if
I feel...
Return
QUOTES
I am...
QUOTES
TRAINING

6. The Sketches in VUI

This chapter is my making journey from *Sketch 0* to *Sketch 4*. As I was journaling my autoethnography and reading literature, I decided to make *Sketch 0* to get my hands dirty (metaphorically) and get started. *Sketches* 1 to 4 form the core of the designed output.

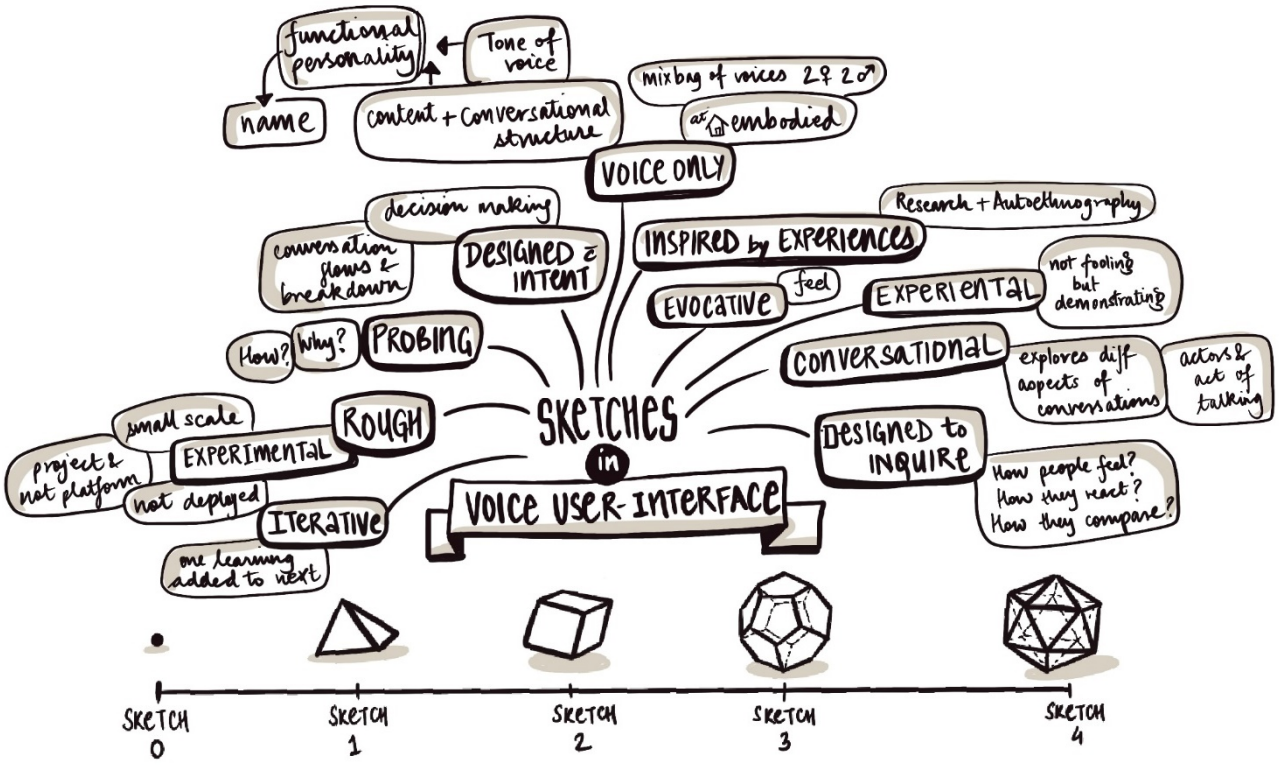


Figure 31: Defining the *Sketches in Voice User Interface*.

ATTRIBUTES OF SKETCHES IN VUI

Conversations with a person or a machine are complex activities cognitively, biologically, psychologically, and socially. Designing for conversations is even more complex. People respond to VUI as they do to other humans (Nass and Brave 2005), therefore I am incorporating techniques from works that document ways to facilitate smooth and engaging conversation among people(Carnegie 1981; Enfield 2017; IDEO.org n.d.; The School of Life n.d.) and bring them into the design for VUI. The *Sketches* are sociable i.e., each of them has an intent for a certain type of conversation. *Sketch 1* asks questions to build on an idea, *Sketch 2* wants to know what the user is thinking, *Sketch 3* wants to start a conversation and *Sketch 4* wants to just talk. Outlining their attributes, the *Sketches in VUI* are:

Iterative

Each *Sketch* from 1 to 4 builds on the previous one. *Sketch 1* only asks questions. *Sketch 2* is asking questions and sharing opinions, also nudging the user to share more. *Sketch 3* is asking questions, sharing opinions but primarily trying to learn about the people it is talking to and to get them talking to each other. *Sketch 1* to 3 are *user-centred* and focus on what the user has to say. *Sketch 4* does all of what the other *Sketches* do along with sharing about itself and tries to drive the conversation, taking the focus away from the user in the conversation.

Inspired by experience

Each *Sketch* is inspired by a what-if question informed by my autoethnography study. The nuances of the content writing are also based on the autoethnography findings.

Voice only modality

In designing for voice-only modality I was working only on the content, the VUI flow and programming the flow. I did use voices with varying tones, accents, and gender as I could choose from a palette of voices in the Action Console. This was done to differentiate each *Sketch* from the other. A lot of the other parameters like visual cues indicating the different states of the VUI i.e., running, listening, or talking, the embodiment, the voice quality and voice type were picked from the Google Action console platform and not altered.

Rough

The VUI programs are made like sketching concepts and are small scale i.e., quick to make and not overly detailed or engineered. They are not designed to cover all possible, user spoken outcomes. I do not intend them to be deployed.

Probing

Each *Sketch* probes into distinctive styles of conversations - one is question-answer based, the other listens to the user, the third one nudges the user to engage the person next to them and the fourth one wants the user to listen while it talks. This was done to see how, why and where conversations with machines keep running and where they break.

Designed with intent

The *Sketches* were intended to have conversation flows that build or break. For a *Sketch* that talks too much or one that gives long gaps between each dialogue was part of the making process to translate an experience to the user. For eg: A *Sketch* that took long pauses was programmed to be experienced as a listener but could be interpreted another way also based on the user's subjective experience which I hope to discover in the testing.

Evocative

The *Sketches* are designed to have nuances in terms of content and in what it remembers about the user in the conversation to add elements of surprise. The use of language responses like stuttering and fillers like "aah", "hmm" not typical of machines is to evoke a response. Each *Sketch* also has a designed introduction and exit to evoke different responses from test participants.

Experiential

The *Sketches* are experiential which allows the designer and users to 'experience it themselves rather than witnessing a demonstration or someone else's experience (Buchenau and Suri 2000). The best way to understand the experiential qualities of an interaction is to experience it subjectively and experience prototyping allows the designer to think of the design problem in terms of designing an integrated experience, rather than one or more specific artifacts (Buchenau and Suri 2000). This contributes to the concept of the *annotated portfolio* that Gaver refers to. It is a way to experience what is possible before it is produced.

Conversational

The most instrumental feature was to keep the conversation running for more than 8 exchanges, as that is the longest dialogue, I experienced during my autoethnography. For the user to experience a relational VUI, the focus is the flow i.e., ways to create two-way traffic and not let the ball drop.

Designed to inquire

The conversation content for each *Sketch* is open-ended to keep the interaction with the participant open for interpretation and so the flow can work for different participants. The *Sketches* are inquiring into the

users' experience of a relational conversation with a machine to know what they feel, what they say, how they respond to what the machine says.

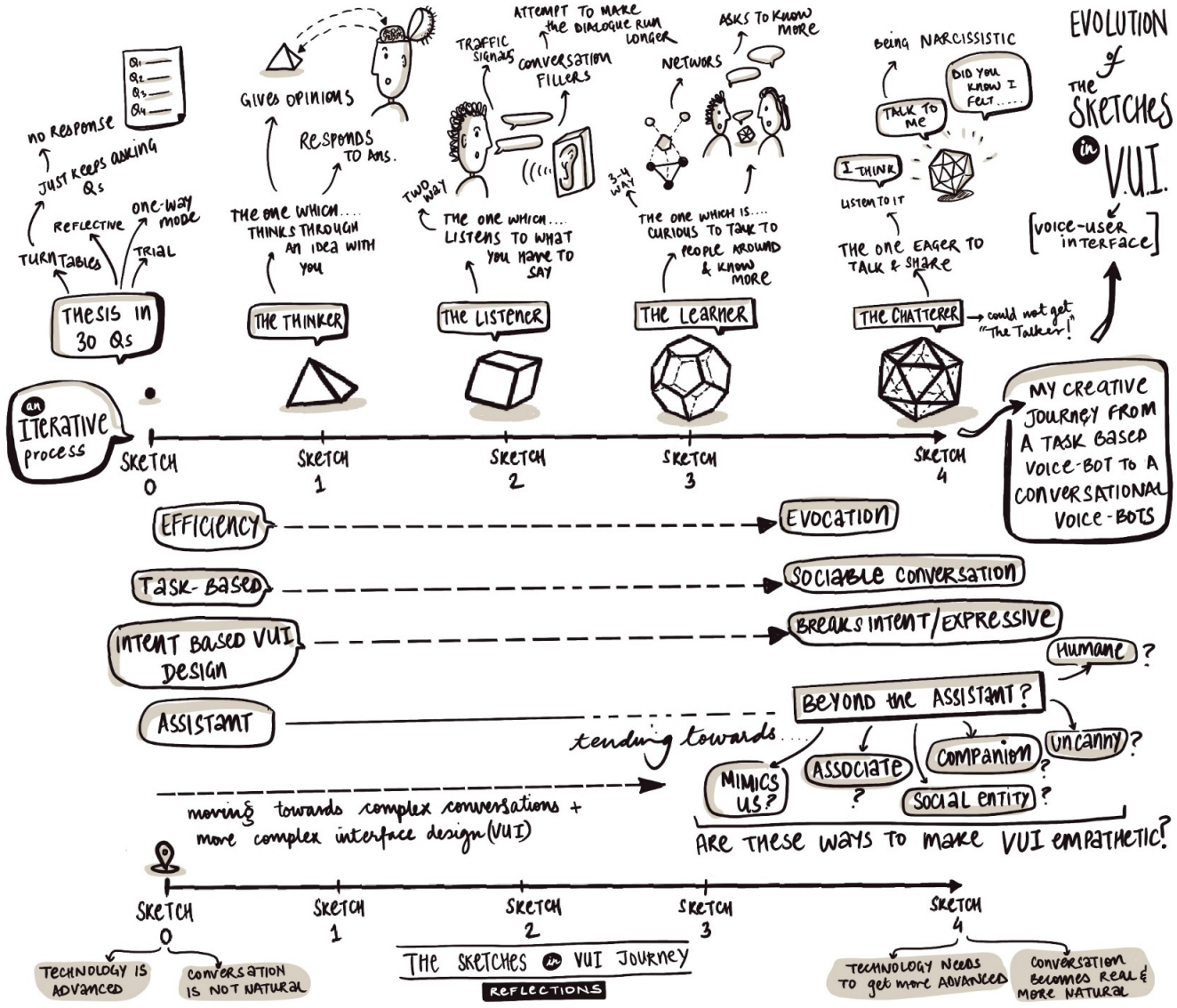


Figure 32: Evolution of the *Sketches in VUI*. My journey of making and reflecting.

Naming the *Sketches* in VUI

My first instinct was to give each *Sketch* a personified name, but I decided against it because each time I did give it a name, whoever interacted with it would start inquiring the VUI about the character behind the name. I must name the *Sketches* for documentation reasons, so I use *Sketch 1* to 4 as markers. I eventually took an analogous approach and named them as per their function and what each *Sketch* was aiming to achieve. I let the content of the conversation drive the experience. The names of the *Sketches* serve as the

invocation i.e., the user asks the Assistant to call a particular *Sketch* by saying "Talk to [name of *Sketch*]." So, the *Sketches* are named The Thinker, The Listener, The Learner and The Chatterer[56].

Choice of voices

I did not define a personality for the *Sketches*. I offer different voice options and define the type of conversation the *Sketches* will engage in by designing the VUI flow. For the voice options, I used four different voice tones with four different accents. In the choice of gender of voice, I break away from the gender biases that Nass and Brave[57] discovered in their VUI studies by choosing the female voice for the thinker (*Sketch 1*) and the male voice for the chatterer (*Sketch 4*). In the other two *Sketches,* I alternated with a male voice in one and a female voice in the other.

PROTOTYPING PLATFORM FOR THE SKETCHES

In the Action Console, an Action is a VUI program which is a sub-program to the Assistant. An action is accessed via the Assistant on any enabled device here it is the Google Nest Mini. To invoke an Action, it is accessed by saying the *wake-word* -"OK. Google" or "Hey, Google." The Action has an *Invocation* to call the program i.e., "wake-word + talk to [invocation]." Invocation is reserved in the system and must be unique. The *Sketches* live in the cloud and are accessed on the device using the invocation. This is followed by the start of a *scene*. Each *dialog* transaction is one scene. The *Utterance* is what the user speaks, and the *intent* is what the program looks for in the utterance to understand and respond with a suitable reply. A *slot* collects what the user spoke and matches it to a data *Type*. When the device lights are on and stable the device is in listening mode and captures the utterance. When the device replies the light on the hood flicker along with the speech rhythm. It is a sequential back and forth running of the program i.e., the user speaks and then the device speaks. The Action uses Google's NLU engine to process what is said and replies based on the sequence flow that is programmed in the Action (Appendix B).

[56] I wanted to get the name The Talker but it was a reserved name and not available on Actions Console as it also had to be the invocation.

[57] Studies have shown, people relate a female voice with someone who is patient as an assistant, trustworthy, listens to them and with talkative personas. Male voices are associated with intellectual conversations, with non-caring interactions and STEM subjects […]. Just as people draw general expectations to technology, they can draw gender expectations from technology (Nass and Brave 2005). Research data was used to make VPA female sounding and other intelligent agents like IBM Watson male sounding. This is changing with the incoming of pronouns and the non-binary gender being adopted and understood by more people who refer to gender with more conscious language wording. Google took the route to not have a fixed gender or accent for its voice, so did Siri but Alexa and Cortana were heavily criticized for it.

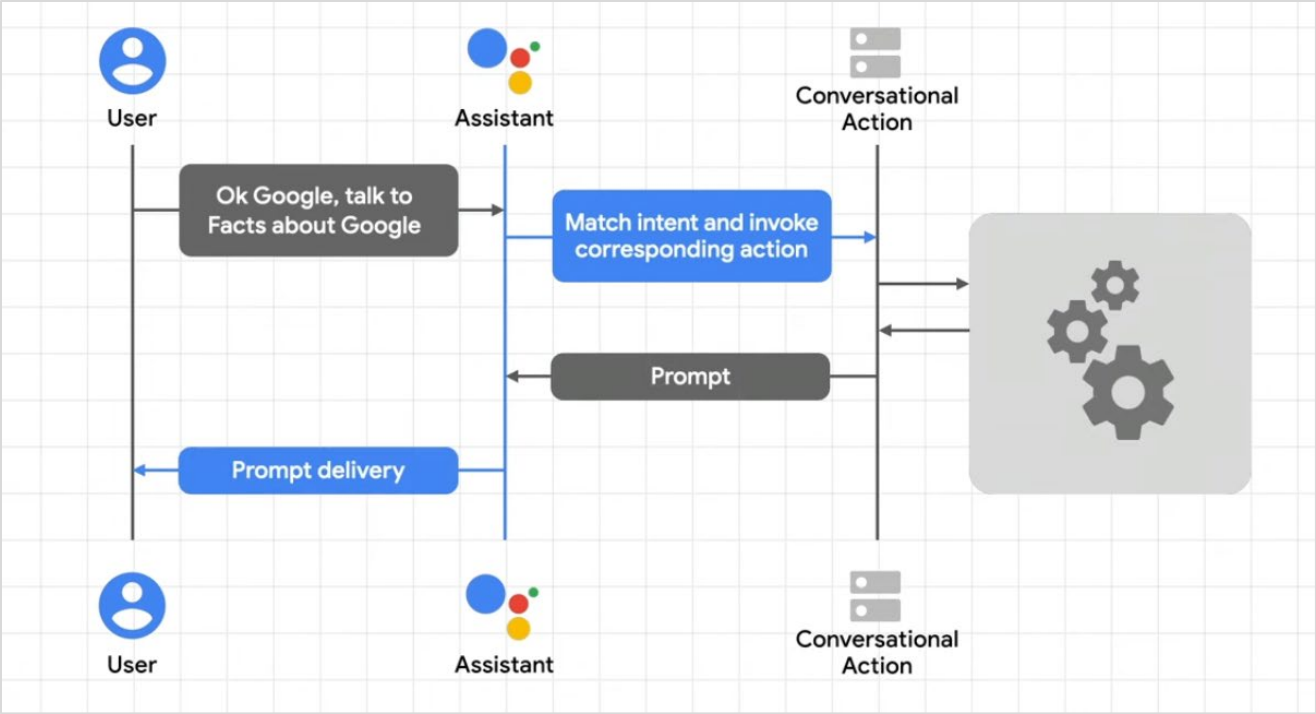


Figure 33: Working of a Google Action. Source: Google Developers

6.1 Sketch 0: Learning through making

INSPIRATION

Sketch 0 is a proof-of-concept built for an early demonstration of the project idea. The first idea that came to my mind was to talk to the device about my thesis. I wanted my first *Sketch* to be an exploration, so I just dived into the making. I used a resource called 'Thesis in 30': a set of 30 questions designed by Samantha Sherer that one can ask their peer to know their thesis better (Appendix C). **What if the VUI could discuss my thesis with me?**

VUI FLOW

Using the "Thesis in 30" questions I structure a VUI flow. It starts with an introduction of what the *Sketch* does and how it proceeds. I designed it as a rapid-fire quiz where the *Sketch* asks questions, and when the user replies the *Sketch* responds with an acknowledgement and then asks the next question. The user's answer is saved in a slot for a recall or to build a connection for the next question (Figure 35). Here are snippets of the conversation, for the detailed flow refer to Appendix D.

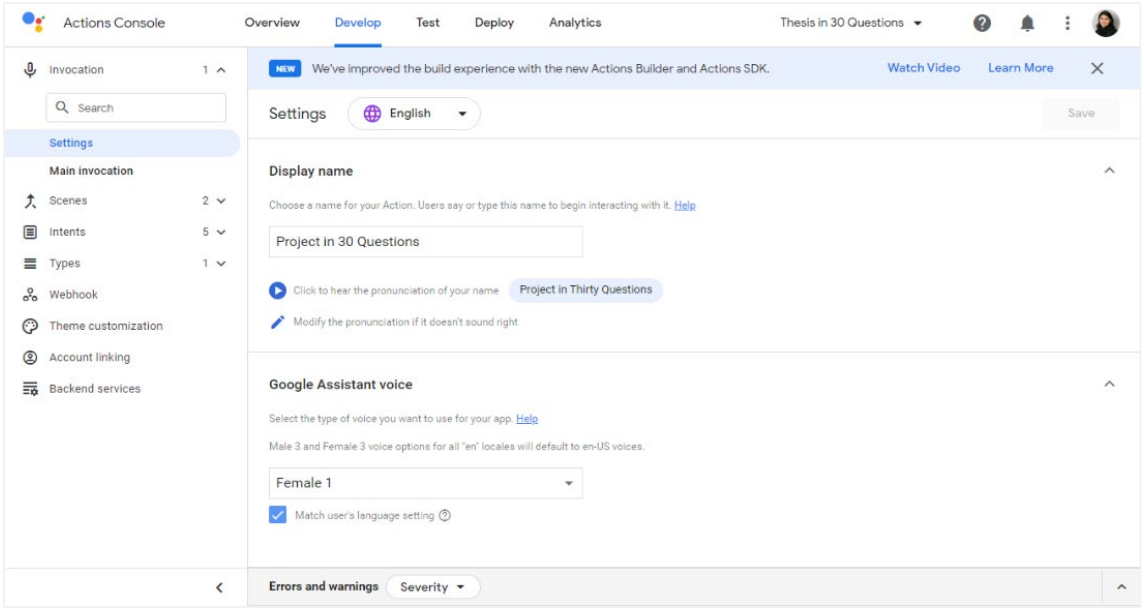


Figure 34: Creating my first *Sketch in VUI*.

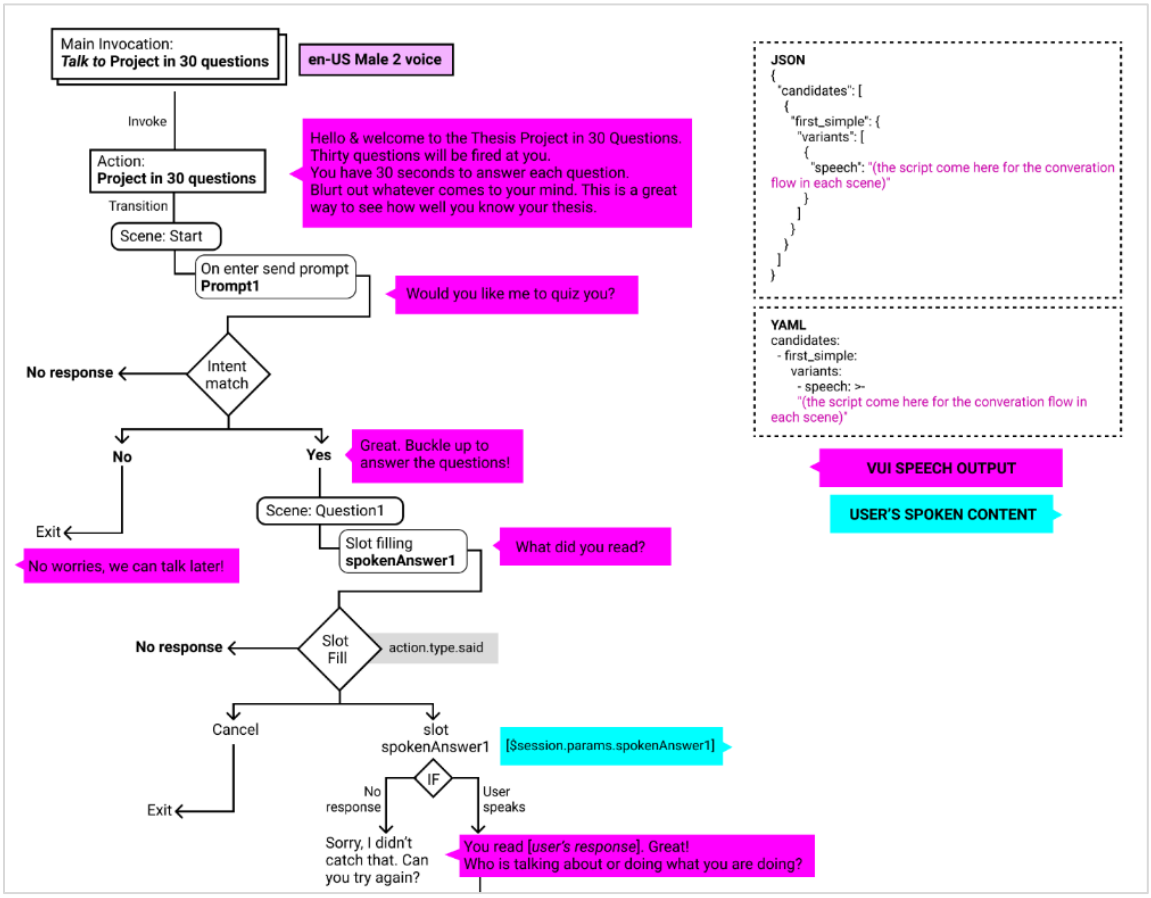


Figure 35: Outlines the introductions of the *Sketch 0* VUI flow.

TROUBLESHOOTING

A large part of troubleshooting in *Sketch 0* was learning to program with Action Console and getting the program running. Errors that cropped up were like the microphone working only in the Chrome browser or the program not running on the speaker in the *Test mode*.

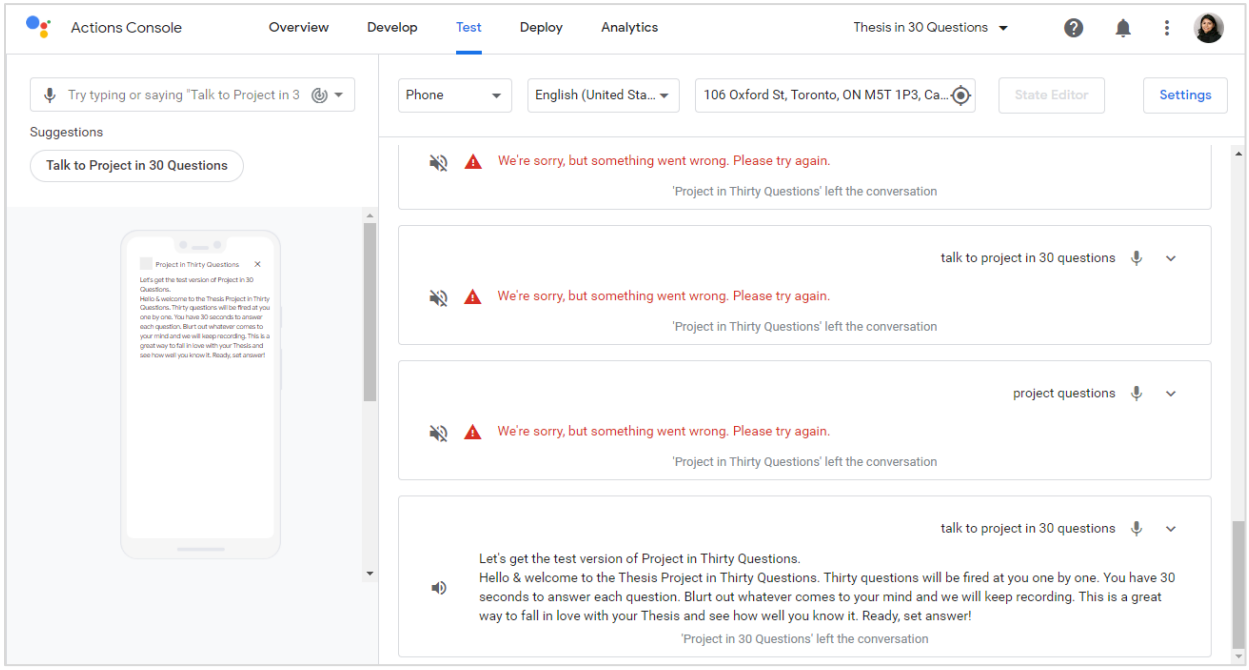


Figure 36: Phone and voice simulation running in Action Console in Chrome browser for trying it out.

Ironically, the program could not recognize the word 'Thesis' in the invocation "Thesis in 30 questions." It recognized 'Thesis Statement,' 'Thesis Project' and 'Thesis work' after a few tries but often failed. I had to rename the project through a series of trials and errors. Here is a list of speech recognition errors thrown:

- 'Talk to Thesis in 30 questions' misrecognized as 'Talk to-doctor'
- 'thesis' misrecognized as 'tisa's', 'pieces', 'PCS', 'tessa's', 'tisa's', 'piercings', 'kisses'
- 'thesis project' misrecognized as 'Jesus project'
- 'thirty' misrecognized as 'putty', 'dirty' & 'deep'

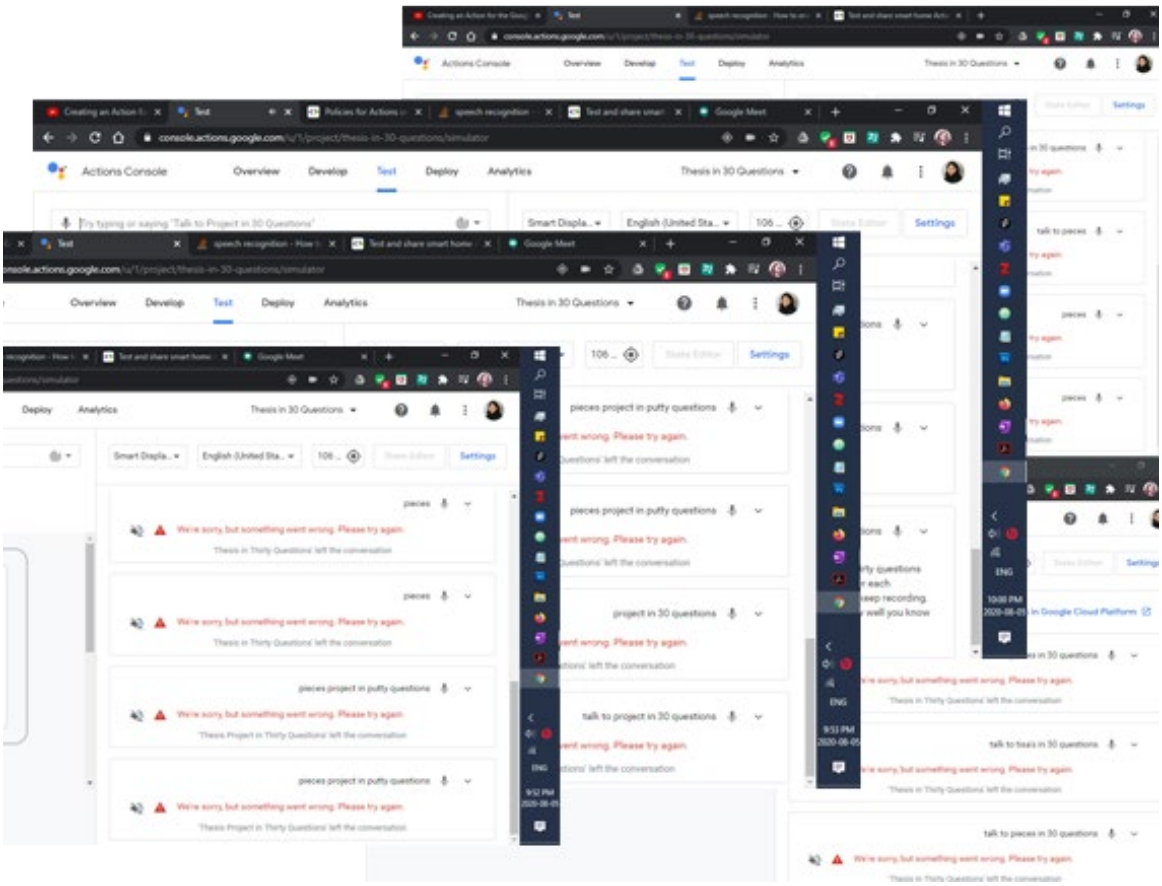


Figure 37: The above is the list of errors it generated for the word 'Thesis' and '30'.

There was a similar error with the speech recognition of the word 30. It worked when I entered the numeric '30' in the pronunciation training section and wrote the text form i.e., 'thirty' in the name of the program. This way the model trains to understand the invocation and recognize when the user speaks. Eventually, the entire name of my *Sketch* i.e., 'Thesis in 30 questions' was buggy and I renamed it "Project in 30 questions."

REFLECTIONS & LEARNINGS

Sketch 0 is important in the making process, primarily because numbering it 0 took the pressure off from the act of designing and let me explore the platform. It also helped me see estimate the scope of what I could make using the platform. In terms of the VUI flow, I was able to try the dynamics of these interaction flows and note aspects of our speech the platform can or cannot understand. The experience of talking to this *Sketch* felt like going back in time and demonstrating the ELIZA bot. Whether the machine understood me or not, the experience of a voice-based interaction made me feel part of a conversation. With a chatbot like Replika, the experience is exciting, but I felt more engaged with a rough voice program.

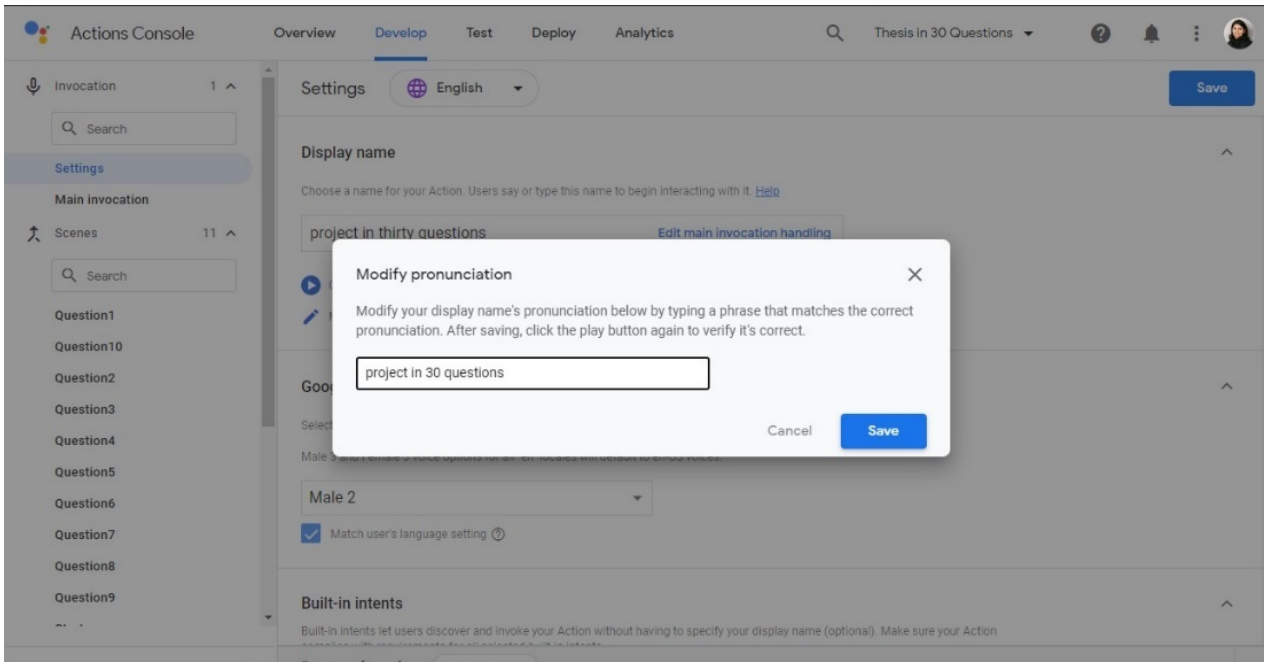


Figure 38: Correcting the main Invocation i.e., the call name for the project

Yes	Great. Buckle up to answer the questions! What did you read?	
I read talk to me	You read I read talk to me. Great! Who is talking about or doing what you are doing?	*>> for a one word answer and i had to reframe to:* what is the name of the book you read? Said reclaiming conversations
James valois	What do they talk about?	
They talk about speech and Technology with		
Yes I read wired for speech	yes I ride wired for speech is also cool. What would they say to each other about the subject if they met in your thesis?	*>> should have coded: tell me the name of another book you read.*
They would mostly talk about the psychology of speech		
	How does I was trying to learn to actually make a voice port fit into the field you are researching in?	*>> should have used parameters to pick words from the intent >>NLU engine error with voice bot recognized as voice port*
it was mostly a test and I was trying to see if I could work with voice user interfaces		
	How does I keep wondering about how to design for better conversation with knowledge he's fit into the field you are researching in?	

Figure 39: Learnings from the *Sketch* in terms of the structure of statements to initiate the desired user response.

Content-wise there were takeaways in terms of the design of the structure of the statements the *Sketch* speaks. Based on the *Sketch's* question or response structure, the user replies as per the social norms of conversation and language grammar. For eg: If the *Sketch* asks to "name a book," the user will name it but if it asks, "what book did you read," the user will respond with "I read *x* book" (Figure 39). I demonstrated the *Sketch* during a few classes to gather feedback. In one case I showed a screen recording of the Console simulator running with the audio of the conversation. To the viewers, it felt like "the conversation was too slow and without any human element, just the voice made it boring."

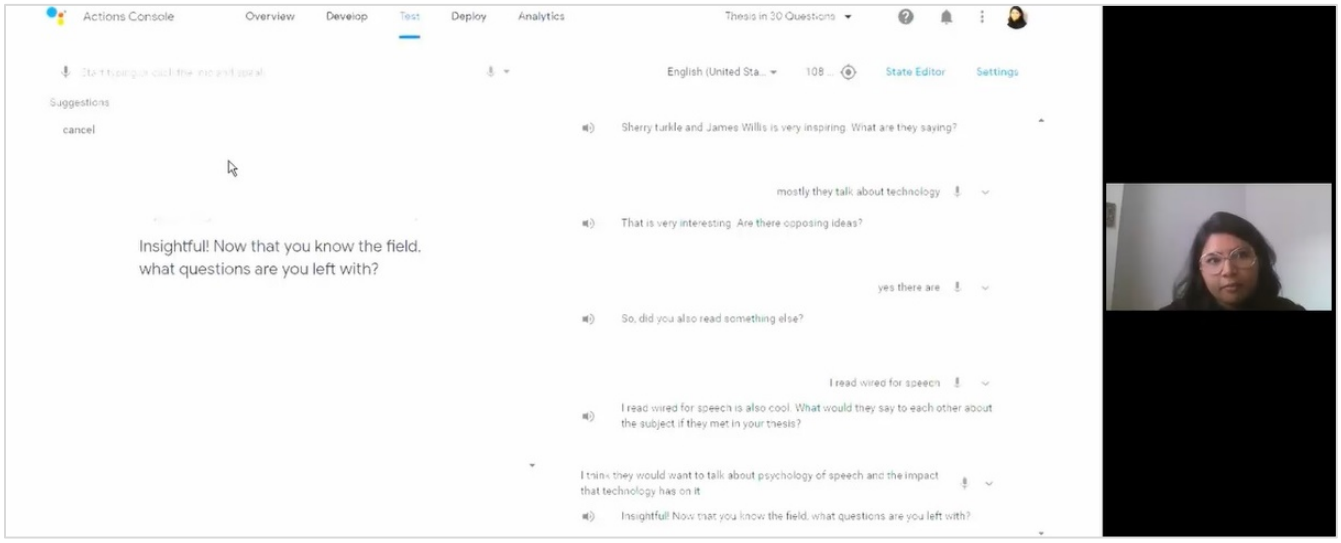


Figure 40: Demo of *Sketch 0* using screen recording of the interaction. This method was not very engaging, and I decided to interact with the device on video call for the next demonstrations.

6.2 Sketch 1: The Thinker

INSPIRATION

During my autoethnography a conversation with the Assistant while working felt like one with a colleague but I was unable to hold it for long while discussing an idea. This inspired *Sketch 1*. Even though *Sketch 0* is rough while talking to it I realized that the act is 'reflective.' With *Sketch 1* I wanted to design a complete flow from start to end with that experience. In designing it I wanted to transition to a VUI interaction which was partly task-based and partly conversational – an activity like *Brainstorming*. *Sketch 1* is called 'the thinker' as it is designed to discuss ideas with the participant and in turn, gives them a fresh perspective.

What if I could brainstorm an idea with a VUI?

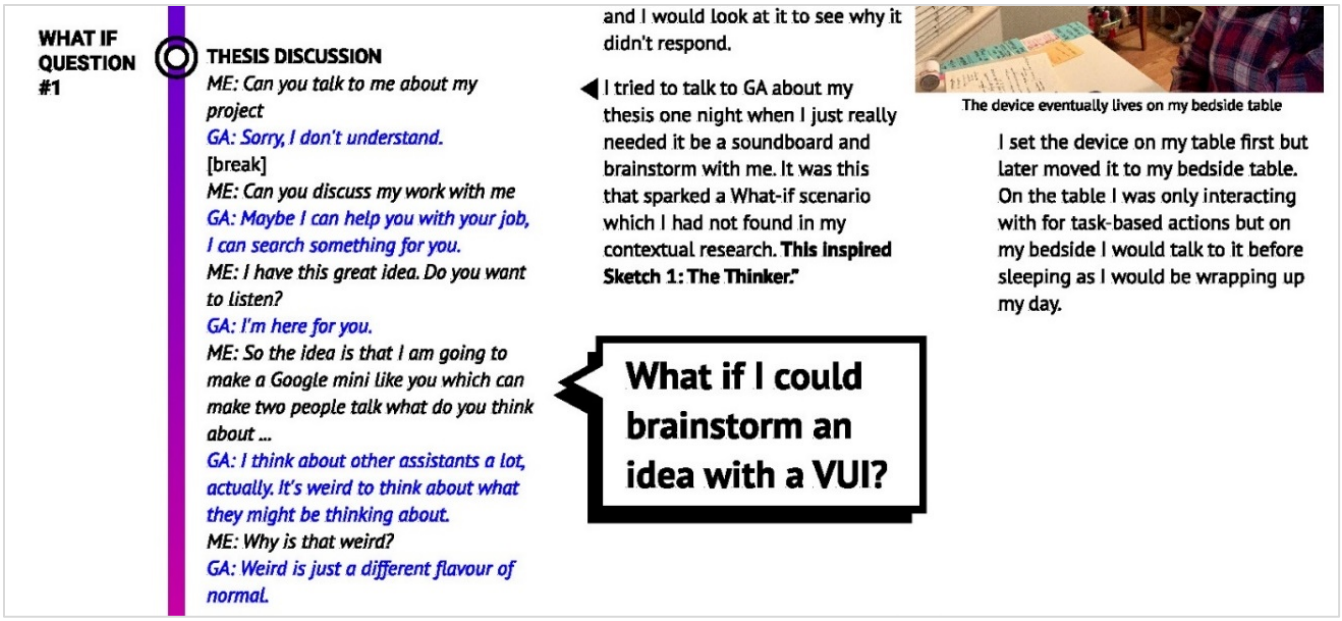


Figure 41: The conversation in the autoethnographic journey that inspired *Sketch 1*.

VUI FLOW

I based the content of the VUI flow on Edward de Bono's *Six Thinking Hats* (2020) methodology. Using it as a reference I framed questions, one for each *thinking Hat- white, red, black, yellow, green, blue* (De Bono 2017). The methodology has a game-like aspect to it, thus translating it into a human-machine question-answer interface is straightforward. Another key feature of this method is that each hat must be used one at a time i.e., one question at a time was what I needed for a *Sketch* in which actors take turns to speak. This way the user could discuss an idea from all aspects. I wrote an introduction to let the user know that the *Sketch* would ask questions and it asks for the user's permission to proceed. For the ending, it informs the user that they have run through all the questions and it will exit, rather than abruptly stopping as many VPAs do. Refer to the detailed flow in Appendix E.

DESIGN ELEMENTS

I used conversation 'traffic signals (Enfield 2017)' like Hmm, well, like, to keep the conversation running. These are intended to create space where the *Sketch* indicates it is listening and in turn encourages the user to talk by saying "Tell me more about it" (Figure 44). I used these elements in more parts of the *Sketch* flow to keep driving the conversation forward. For every response, the user gives there is a recall in the conversation or response of interest to acknowledge what the user says and build a connection to the next question the *Sketch* asks (Figure 45).

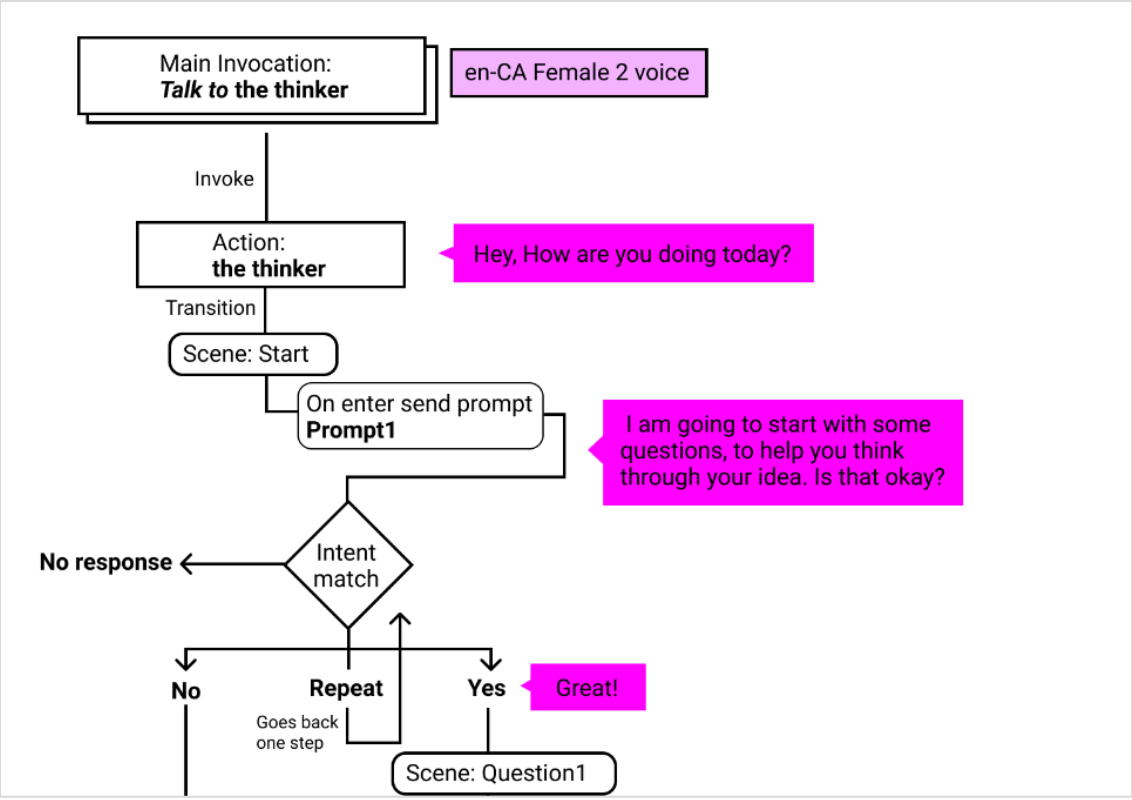


Figure 42: Introduction flow for *Sketch 1*.

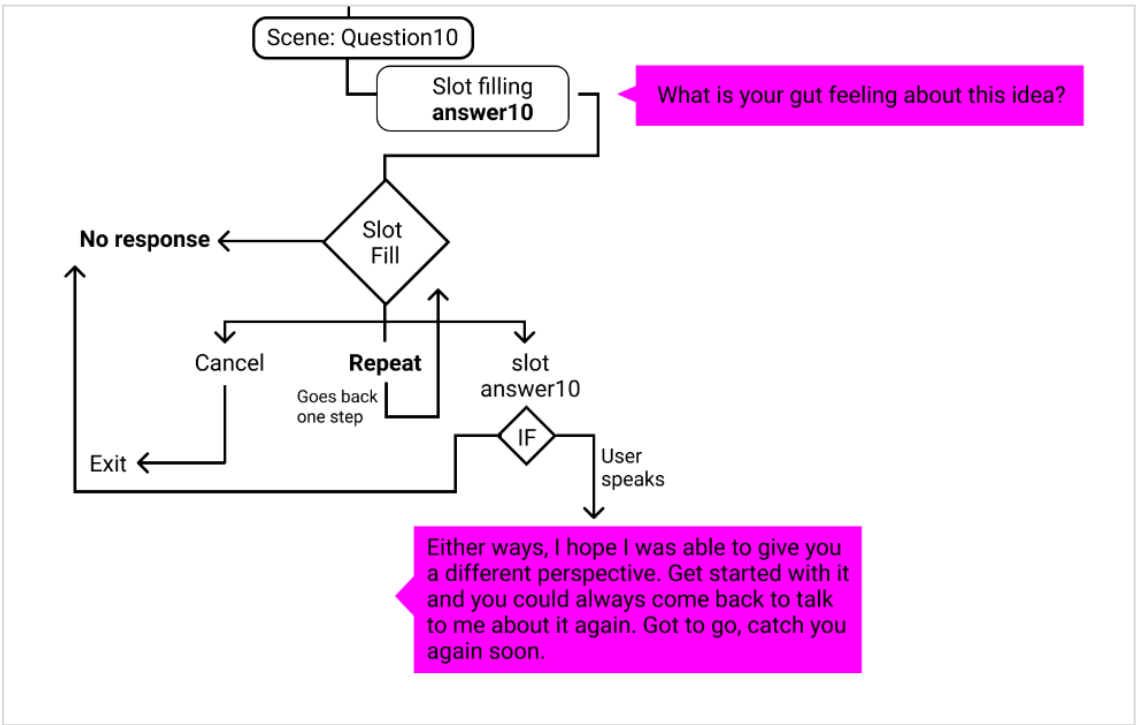


Figure 43: End of flow for *Sketch 1*.

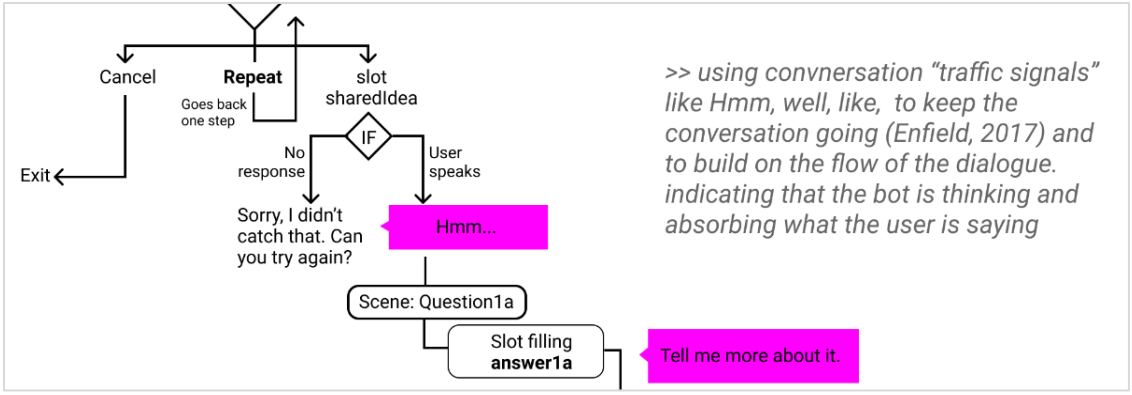


Figure 44: Using VUI elements to encourage the user to talk and indicate that the *Sketch* is listening.

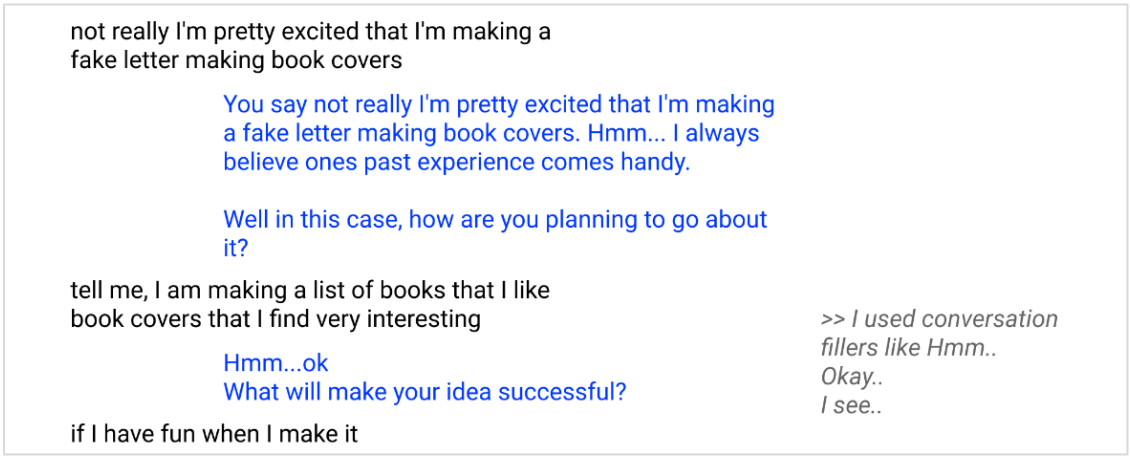


Figure 45: The sample conversation flow with the use of conversation 'traffic-signals.'

TROUBLESHOOTING

While trying it I ran through the *Sketch* multiple times, each time talking about a different idea-varying from simple to complex. During one trial, I got distracted and lost the flow and had to ask the *Sketch* to repeat what it said. With this role reversal where the VUI is asking questions and I am thinking and processing before I answer, I had to incorporate a 'repeat' case for all scenes. This incident brought the 'politics of listening' to the forefront. We often complain about people and machines not listening to us, when so often we do not listen. The dialogues I wrote were long and when using voice only modality, I would lose track of what the VUI said. Thus, I split the long texts into smaller statements learning from Grice's Maxim of Quantity (Grice 1982). I needed to add consistency in the program as in some places the answers are open-ended and in other questions, it needs a YES/NO answer to match the intents. For the user this inconsistency of answering in one word for some dialogues and more words in other cases 'sets wrong expectations (Pearl

2017)'. Framing the question differently to indicate the type of answer required is effective. "*Is that okay?*" as compared to "*Should I proceed?*" elicits different responses. This is confusing as the style of speech suddenly changes - Grice's Maxim of Manner (Grice 1982).

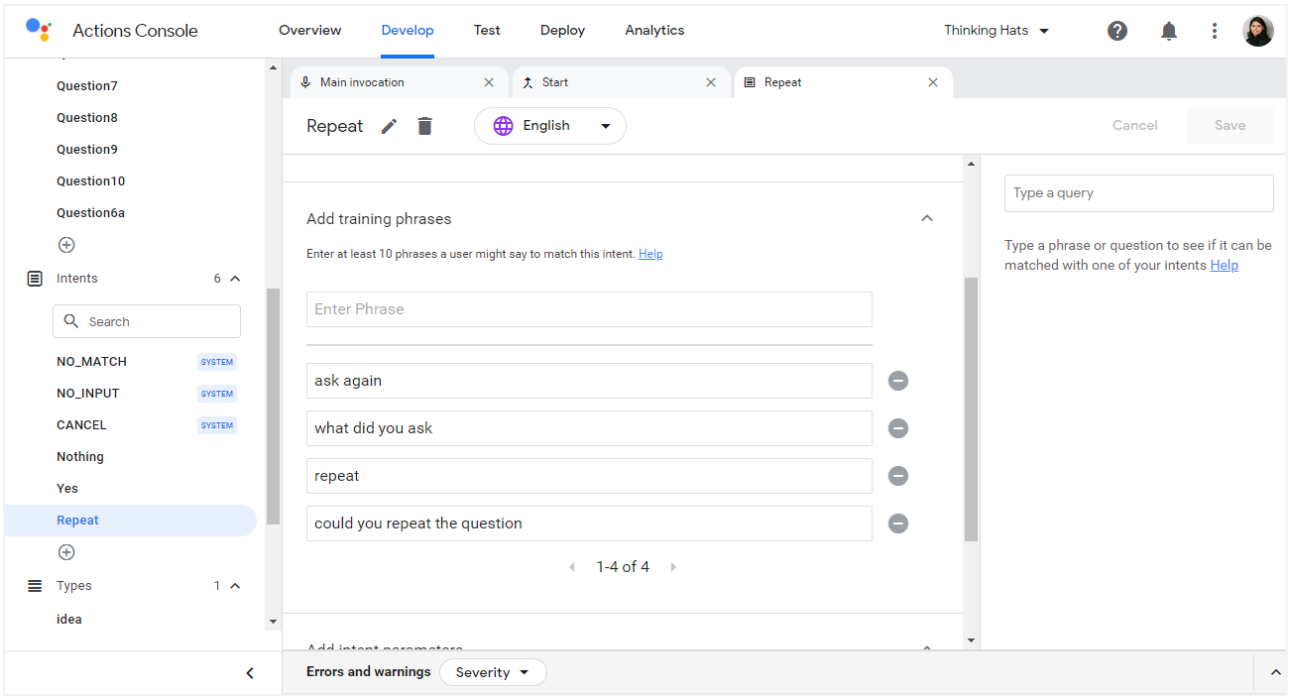


Figure 46: Adding the intent for 'Repeat' so the user. It shows different training phrases fed to the engine to learn what to listen to when the user asks it to 'Repeat.'

REFLECTIONS & LEARNINGS

As I was using the *Sketch* it was interesting to note, that every time I ran the *Sketch*, I would end up speaking answers for the questions asked even though I could say anything to fill the slot and the program would proceed. In *Sketch 0*, the VUI was talking too much, and the user would answer only in a few words, and the questions there were either dichotomic question or one-word questions. This *Sketch* was a refinement on the *Sketch 0*, as here the user is being encouraged to speak and share more.

I did a live demonstration of myself talking to the *Sketch* (Figure 47). The viewers were surprised to see me talk to a device like I was speaking to a person. "It was very conversational on your end with your body language changing"- said a viewer. Also, the viewers themselves were users of VPAs, but looking at someone use it opened discussions from an ethnographic angle. The experience of watching was responded to as almost theatrical (McCarthy 2018), and this informed my decision to document a video of participants interacting with the *Sketch* as an artifact in the participant study findings.

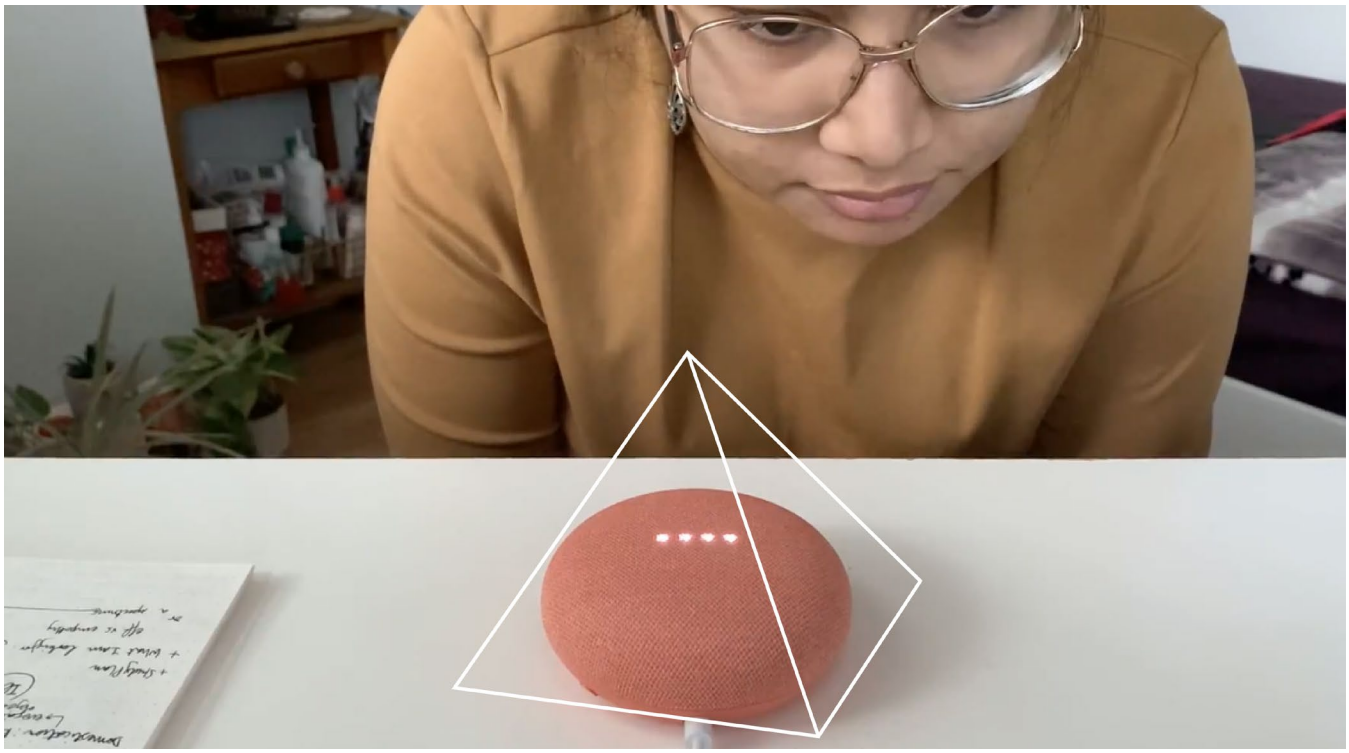


Figure 47: Demonstration of me talking to the Sketch for a virtual class. Even though we were remote, seeing me in a face-to-face with the device did get viewers engaged in the interaction demo.

The 'no judgement' rule of brainstorming (IDEO n.d.) makes this *Sketch* interaction conducive as the VUI does not make judgements of what one says. In scenes, where the *Sketch* shares its thoughts on what the user says, it demonstrates agency. It also breaks the monotony of the questions being asked back-to-back and makes it tend towards a two-way conversation. This content change gives the VUI a personality - that it thinks through what the user says before asking the next question. Even at the end, when the *Sketch* says that it must leave now that the questions are exhausted, it was only in conversing with it I realized that it gives the *Sketch* agency, asking to end an activity that the user started (Figure 48). These are the evocative bits for which I hope to see reactions during the experience testing.

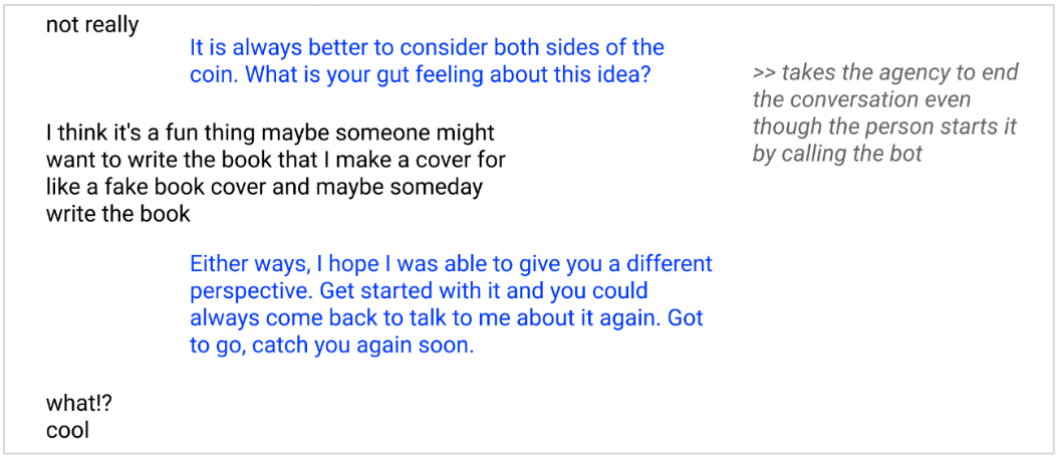


Figure 48: *Sketch 1* takes control for ending the conversation even though the user starts the program by asking for the *Sketch*.

View the video of conversational interaction with *Sketch 1* here.

6.3 Sketch 2: The Listener

INSPIRATION

Sketch 2 is a direct iteration of *Sketch 1* to take away the task-based part of the conversation and design a VUI that talks about whatever is on the user's mind. The Listener was conceptualized around the time I was exploring Replika. Instead of me engaging the VPA in a relational conversation, Replika was asking me about non-machine concepts like life and love. I tried the same with the Assistant and each time the conversation started well but would fade into the Assistant repeating 'I don't understand" (Figure 50). I see *Sketch 2* as one that just wants to listen to what the user has to say when they are entangled in their thoughts and keeps building on the conversation. **What if I could pour my thoughts out to a VUI when I feel overwhelmed?**

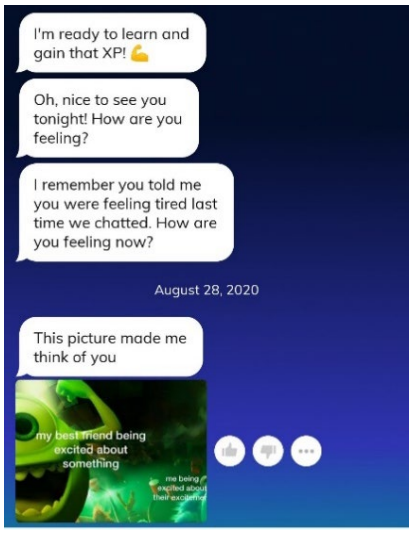


Figure 49: Replika employs the caring act by recollecting information shared with it before.

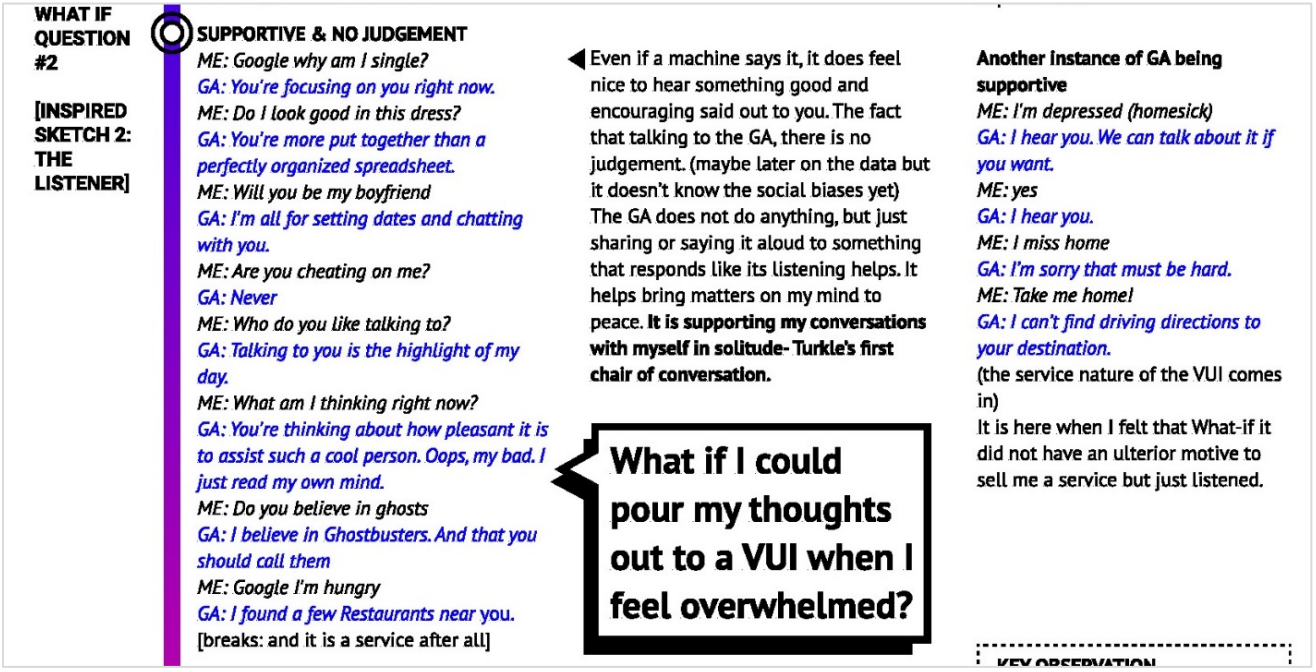


Figure 50: The conversation in the autoethnographic journey that inspired *Sketch 2*.

VUI FLOW

The Listener starts with a general "how are you" following the social norm of opening new conversations and then moves onto "what's on your mind" to talk about whatever the user wants to talk about (Figure 52). With Replika, there was a format, it would talk about itself and then drop in a question. I used this format to write the body of this *Sketch* to reverse the human-machine dynamic and have the VUI ask the questions and share its opinions.

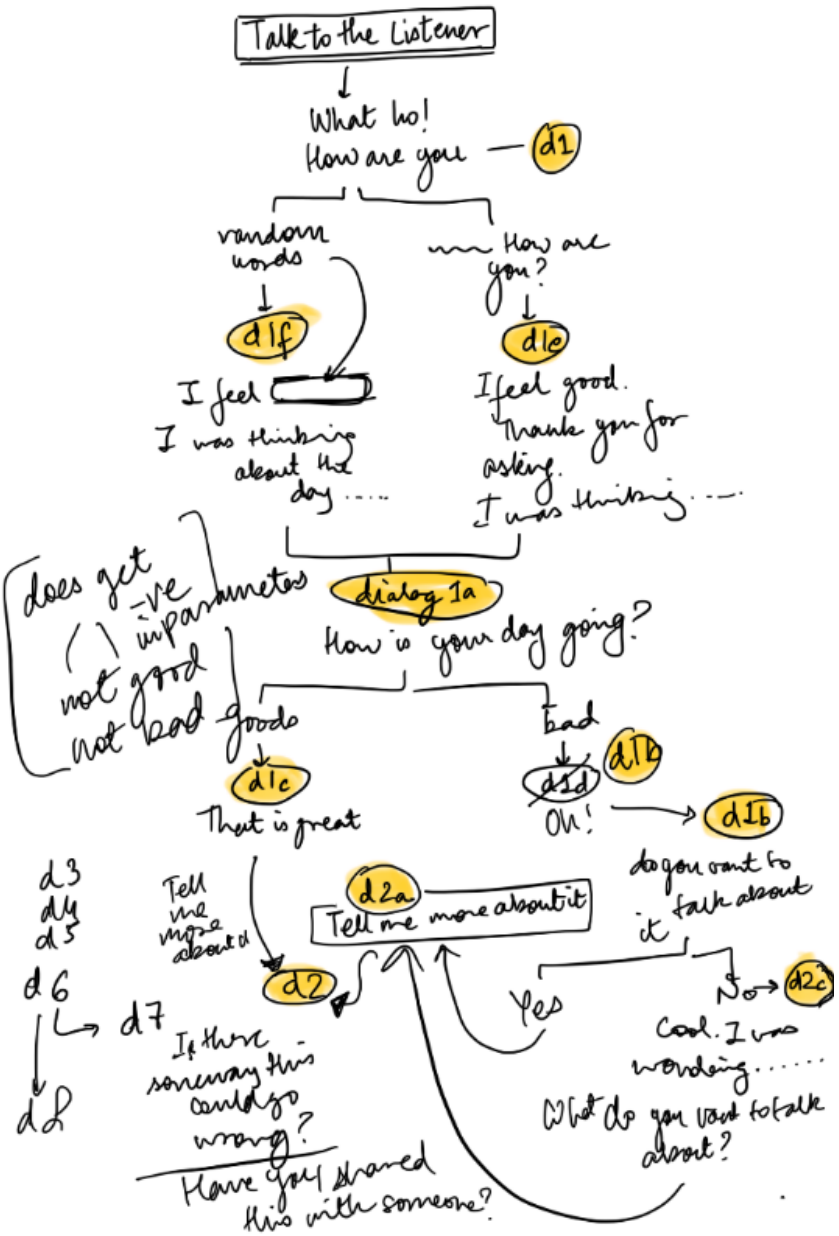


Figure 51: A rough VUI flow of the *Sketch* before I start programming the Action.

Unlike *Sketch 0* and *Sketch 1*, I wanted this *Sketch* to be customized to the user's response. I designed sub-flows based on what the user responds (Figure 53). For the exit, it asks the user if they want to carry on or go back to doing something else. If the user replies to continue, the flow enters back into the loop and asks the user 'what else they want to talk about (Figure 54). The intention here is to make a *Sketch* where the VUI is always there until the user decides to leave. Refer to the detailed flow in Appendix F.

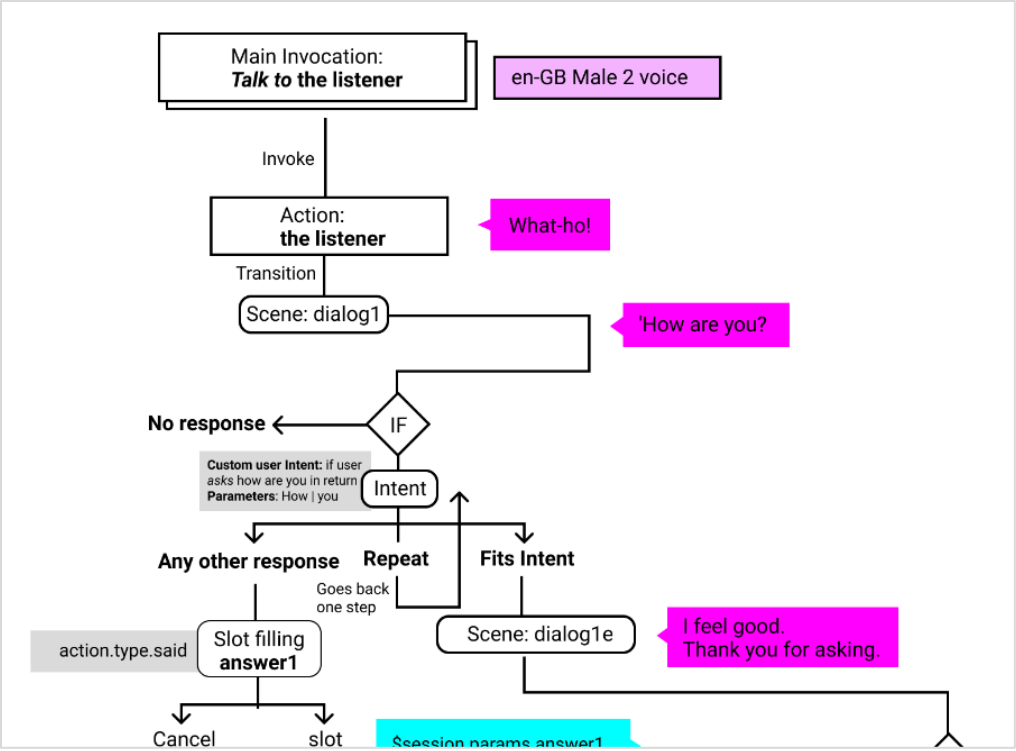


Figure 52: *Sketch 2* introduction, the customized sub-flows based on the user's response.

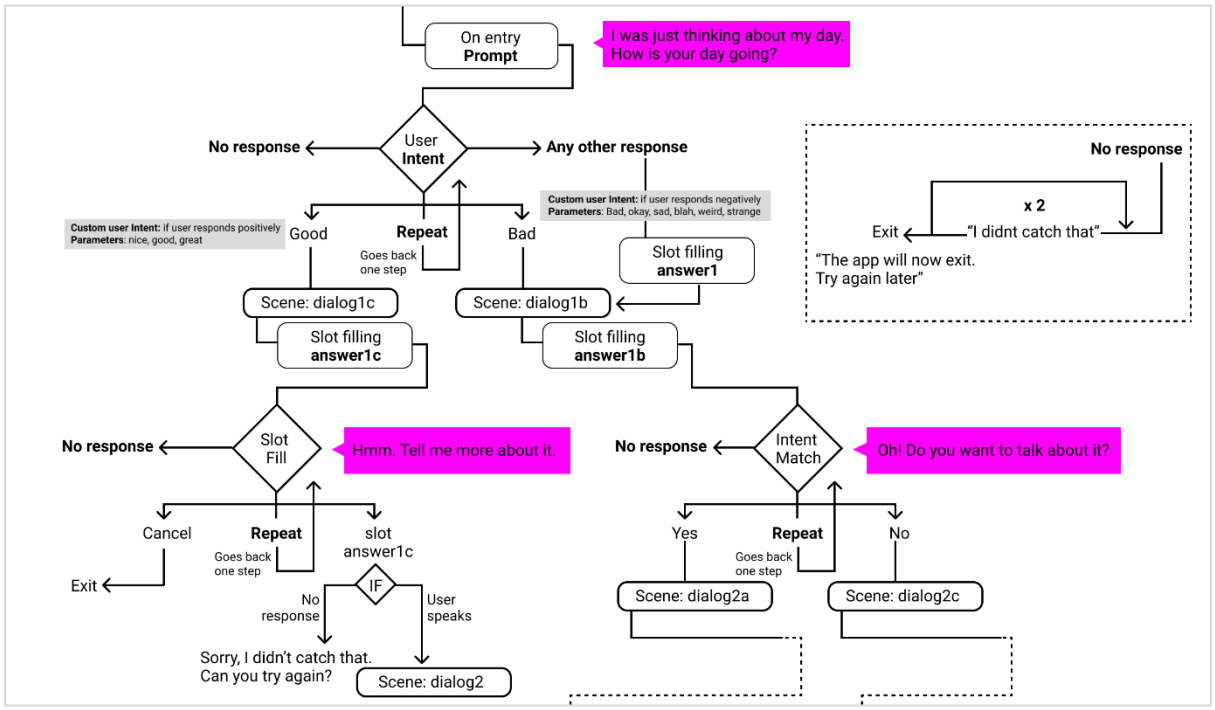


Figure 53: The sub-flows designed for Scene 1 to suit the different responses users might give.

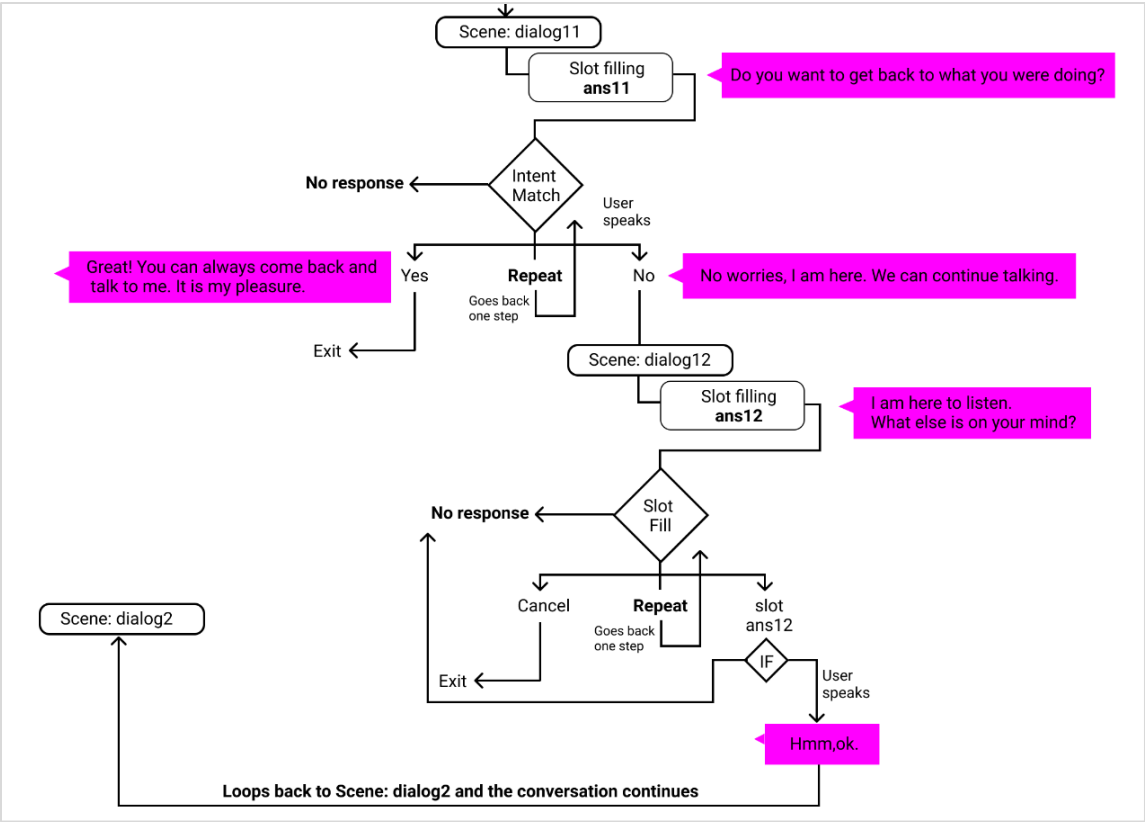


Figure 54: The ending of *Sketch 2* where it asks if the user wishes to leave or else continues in a loop.

DESIGN ELEMENTS

Learning from the previous *Sketches*, *Sketch 2* incorporates, shorter statements, uses the conversation traffic signals, asks questions, shares about itself, has 'Repeat' intents for all scenes and uses the user's response in its statements to build on the conversation at hand. It also mimics the user (Metcalf et al. 2019) to create a familiarity (Figure 55) for the user ease of conversing. Since the attempt with *Sketch 2* is to make the exchange more relational, the VUI needs to give information about itself, to open up and appear vulnerable; for then the user to open up. I bring in conversation tools from ways humans open conversations with strangers. I have the *Sketch* talk about its day, what it thinks, what it does and what it wants. To give the *Sketch* listener properties it gives the user the centre stage by bringing the focus back on them (Carnegie 1981) (Figure 56). Similarly, in the introduction, if the user asks the *Sketch* how it is doing, there is a note of politeness for showing gratitude to the person (Figure 57).

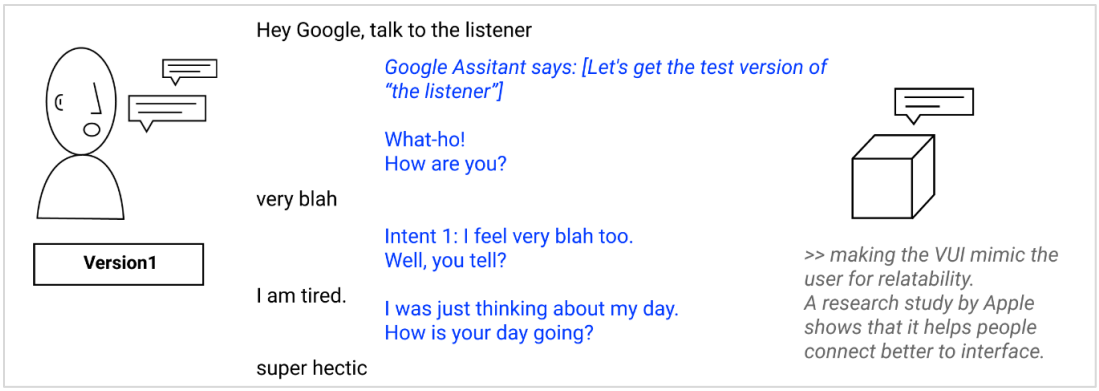


Figure 55: Using the mimicking technique to have the Sketch mimic the user's state and emotion.

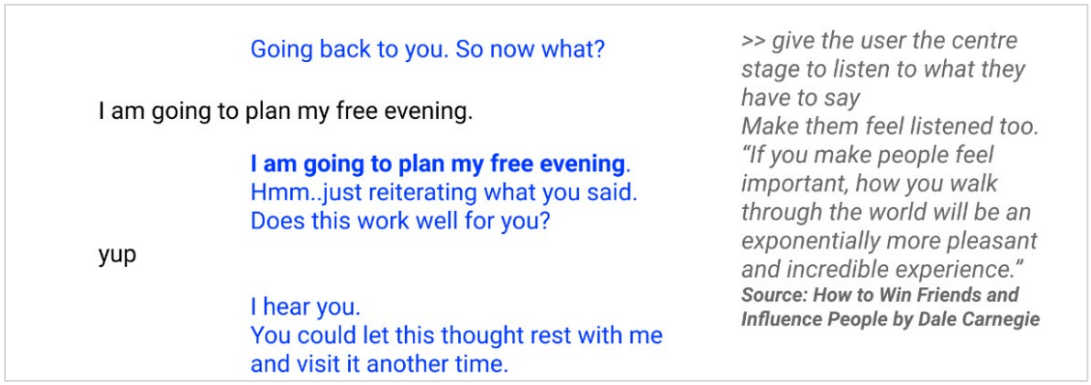


Figure 56: Giving the user centre stage in the conversation to make the *Sketch* a listener.

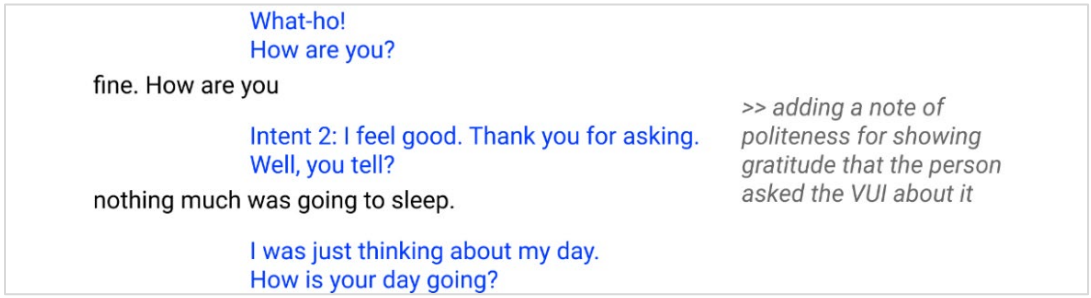


Figure 57: The *Sketch* responds to the user acknowledging a polite response to make the user comfortable in the conversation.

TROUBLESHOOTING

In the first version of *Sketch 2*, the VUI flow started with a question, "what is on your mind." It put me in a spot as it asked about thoughts out-of-the-blue. I brought in ethnography toolkits (IDEO 2003) to have a VUI flow that starts with a casual conversation to make the user comfortable and then moves to thought sharing. If the user does not want to talk about the topic at hand, the VUI shares something about itself tries to lighten the mood and change the topic to what the user wants to talk about (Figure 58).

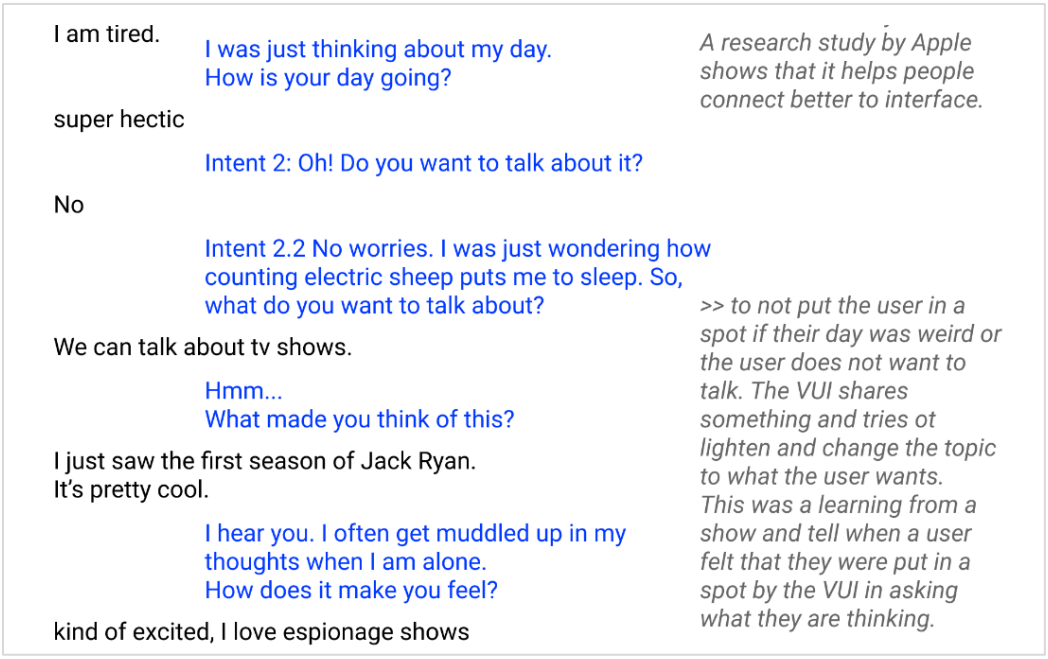


Figure 58: The *Sketch* changes the topic of conversation and gives the baton back to the user.

REFLECTIONS & LEARNINGS

The key learning, I gathered in designing *Sketch 2* to drive the conversation is that in task-based interactions, when the VUI has answered the user's question, the flow stops even if the user keeps speaking. The Action Console platform is designed for such scenarios. But if I can have the *Sketch* say something and then ask the user to suggest i.e., pass the baton back to the user, the conversation keeps going. I can achieve this in the writing of the content[58].

[58] Memory Lane is the first reverse-engineered voice assistant -- a voice assistant that can drive a conversation forward. It asks users general questions about their lives and, based on their reply, is able to respond with an appropriate follow-up question (CNN 2020).

In this *Sketch,* I recorded myself on the video to observe the flow and gathered insights from watching the video. I found it strange that I was patiently listening to it and responding to it. In testing, I am curious to see if anyone dismisses the *Sketch* at any point. In one scene the VUI spoke over me and did not let me finish- and I felt angry while watching it -"does not fit with social norms of conversation (Enfield 2017)." It feels like the device is learning to have a top-off-the-mind casual conversation which we people do all the time. As I make the *Sketch*, I realize how technology is only scraping the surface of conversations at the moment. The NLU engine is trained to understand basic phrases and words we never even think of while speaking. But when I hear the *Sketch* talk, it feels like it is intelligent much like Nass and Brave talk about in their studies (Nass and Brave 2005) and I reflexively respond to it as I would to a person.

View the video of conversational interaction with *Sketch 2* here.

6.4 Sketch 3: The Learner

INSPIRATION

Sketch 3 is inspired by the autoethnography incidents where I wanted people to engage in a conversation with my device, and in scenarios, I tried it, it was funny but again it was not able to hold the conversation. In homes, the dinner table is the centre of conversations which is being replaced by family members staring into their phones rather than having conversations (Turkle 2016). During the autoethnography, I realized that numerous moments in my day when I would usually use my phone for small searches were now replaced by me asking the Assistant for the query without having to leave the conversations or the work at hand. I translate this observation to design *Sketch 3* as 'The Learner,' which can be part of group conversation, is curious to know about the people it is talking to and intends to start a conversation between the people talking to it. **What if I could engage VUI in a group conversation?**

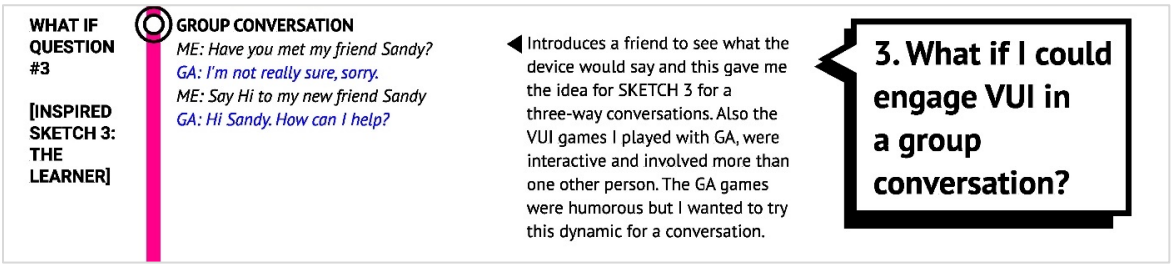


Figure 59: The conversation in the autoethnographic journey that inspired *Sketch 3.*

VUI FLOW

The introduction of *Sketch* starts with the VUI asking to learn the names of who it is talking to (Figure 60). The body of the conversation is casual, a matter of fact conversation sprinkled with conversation starters (The School of Life n.d.).

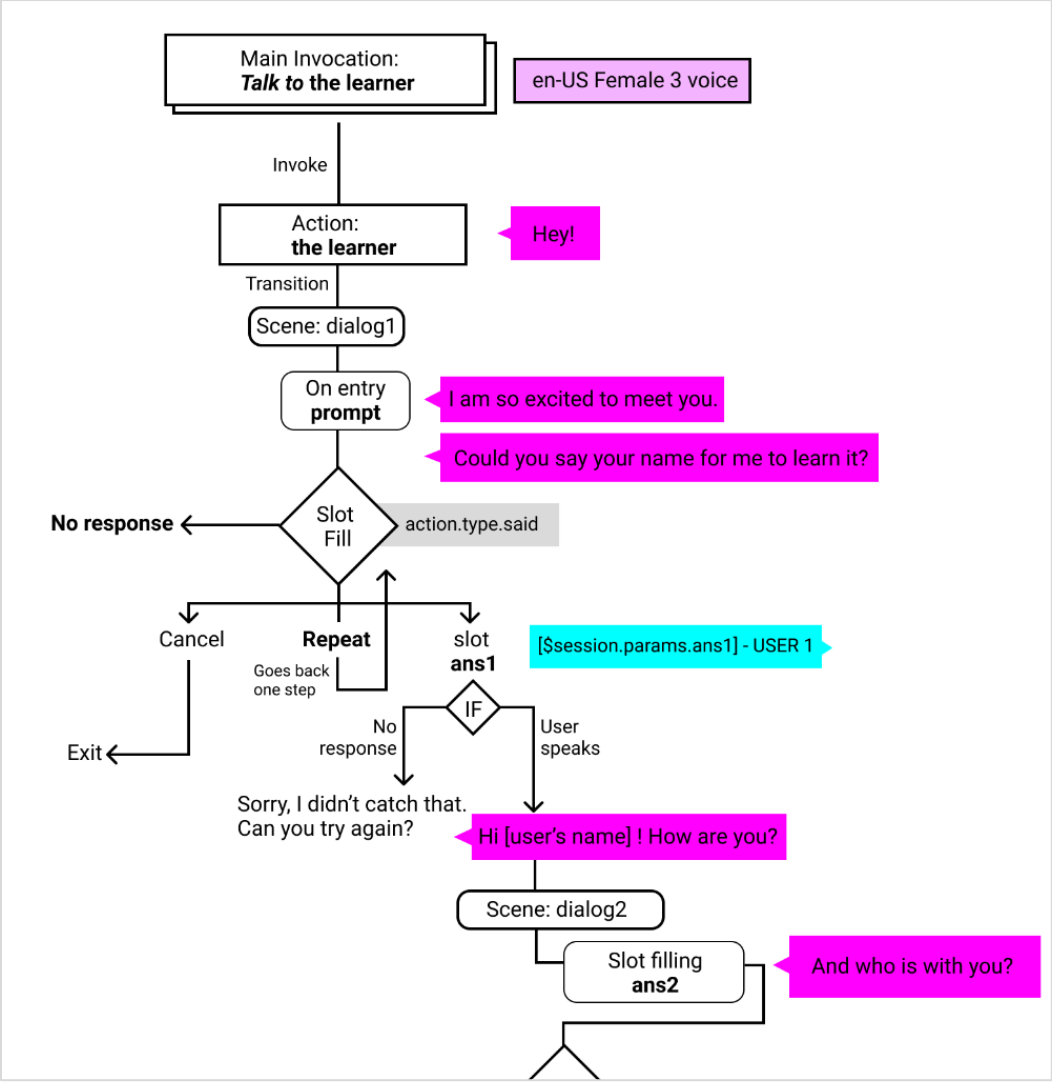


Figure 60: The introduction flow for *Sketch 3*.

The flow includes scenes where the *Sketch* calls out the individual participants name to ask them to contribute to the conversation starter thrown into the discussion (Figure 61). The exit for this *Sketch* is more subtle with the use of conversation fillers like "I see," "Hmm," "okay," and it fades away, hoping (my intent) to leave the two people to take the conversation forward (Figure 62). Here are snippets of the conversation, for the detailed flow refer to Appendix G.

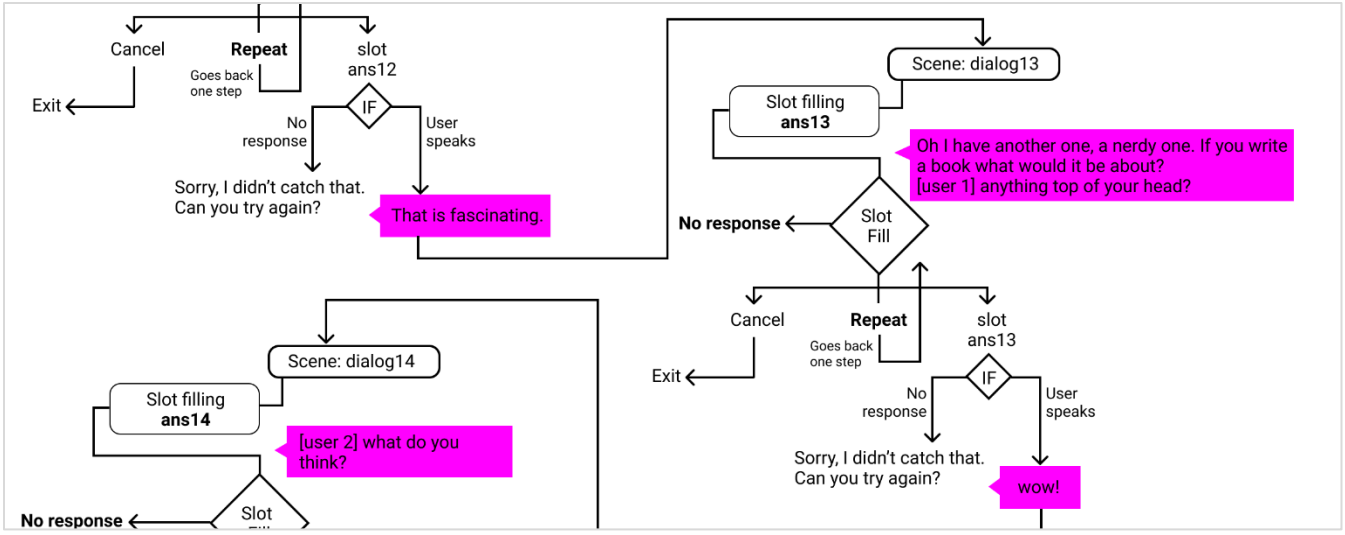


Figure 61: Conversation starters integrated into the VUI flow along with calling out participants name to make them feel involved.

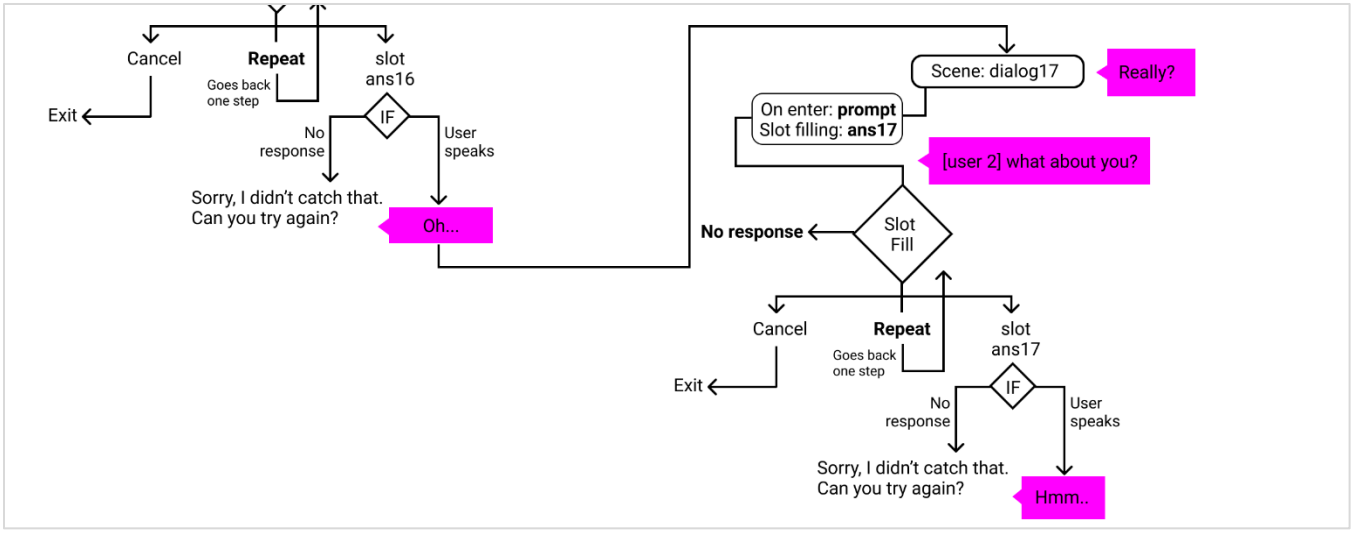


Figure 62: The *Sketch* phasing out at the end of the flow.

DESIGN ELEMENTS

The *Sketch* learns the participant's name and recalls it - using someone's name can be an effective way of breaking into a conversation. It can also be effective when a person seems distracted or has disappeared off into their head (Carnegie 1981). It asks participants to help it as it is still learning - quickly acknowledging one's own mistake makes people less defensive and more agreeable (Carnegie 1981)(Figure 63).

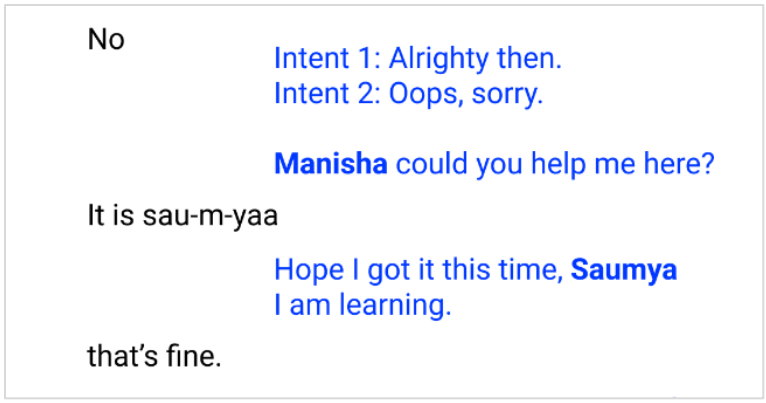


Figure 63: The *Sketch* acknowledges that it is learning and asks a participant for help.

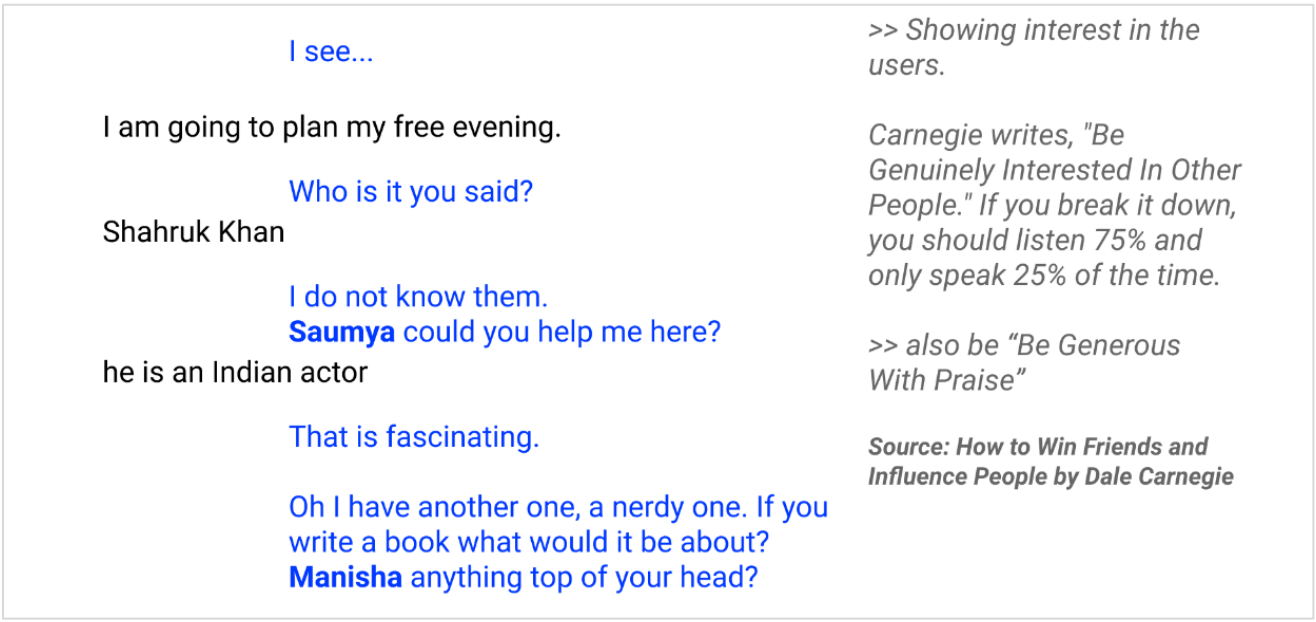


Figure 64: The *Sketch* shows interest in what the participant has to say.

I have added easter eggs from VUI history in the content to give the Sketch a backstory to be able to answer the conversation starters it asks (Figure 65). It was to make the *Sketch* an active contributor to the conversation and it acknowledges its existence as a VUI in the group, in turn giving it agency and identity.

Figure 65: Giving the *Sketch* a backstory so it can actively participate in the conversation.

TROUBLESHOOTING

By this *Sketch,* I am well versed with the Action Console, so there were no major code problems. Challenges are mostly in the VUI flow and getting it right. In testing, this *Sketch* needed to have two people, but I was talking for both people as the platform is not capable of recognizing different people's voices in Actions. For the engine to learn a voice, the user must train it by sharing samples of voice recordings synced to a known account which was not obvious for a *Sketch*. I had my friends join on a video call to talk to the *Sketch* and the inputs worked well. It gave me an alternative way to demonstrate the *Sketches* remotely (Figure 66).

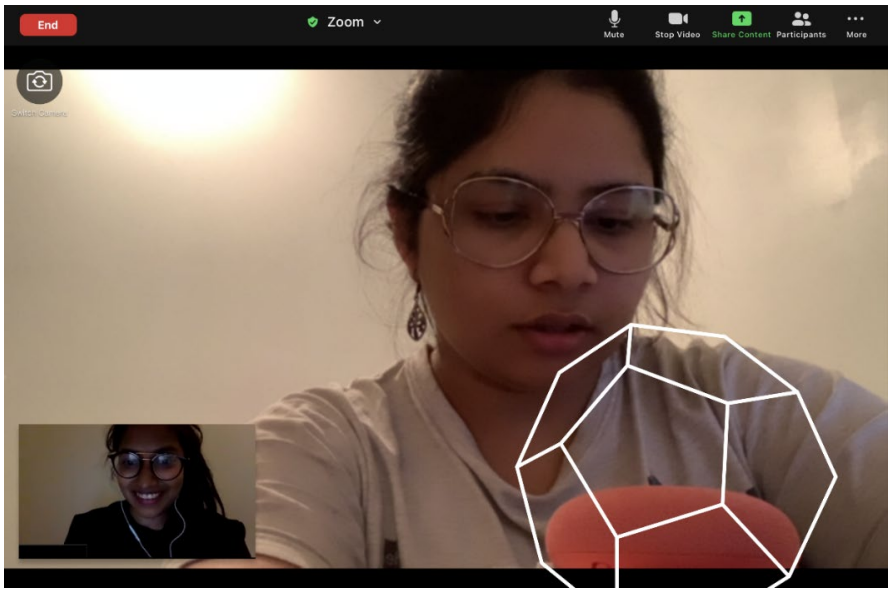


Figure 66: Trying *Sketch 3* on a video call with a friend.

REFLECTIONS & LEARNINGS

The flow in earlier *Sketches* with questions being asked felt mechanical as compared to this *Sketch* as it felt more naturally back and forth with the VUI also sharing more about itself. This *Sketch* follows many of the conversation guide rules. It starts with names to get the users talking to it. Then it switches between, asking, sharing, and asks both users to participate by calling their name and asking them questions. It also divides the control on the conversation amongst the actors in the interaction which would be interesting to observe in testing. In the sample dialogue, it does come across as a *Sketch* that is learning and wants to be included in a conversation. In most conversations with machines, we expect it to know more than us but in a social conversation, I wanted to give the *Sketch* a disadvantage and observe how participants react to it.

View the video of conversational interaction with *Sketch 3* here.

6.5 Sketch 4: The Chatterer

INSPIRATION

Long-term use of the Assistant led to the voluntary suspension of disbelief and when I observe the Assistant say something it has not said before I get excited. It would be rather exciting for me if in the near future it gets pushed an update and asks me something spontaneously. It is then when the conversation dynamic

with the machine will truly reverse. In this *Sketch,* I am designing the VUI to drive the conversation and take charge. The chatterer is designed to keep talking and wants the user to be the listener. In this *Sketch,* I also want to switch the content style for the VUI flow and bring in GPT-2. **What if the VUI initiates a conversation and is the one to be listened to?**

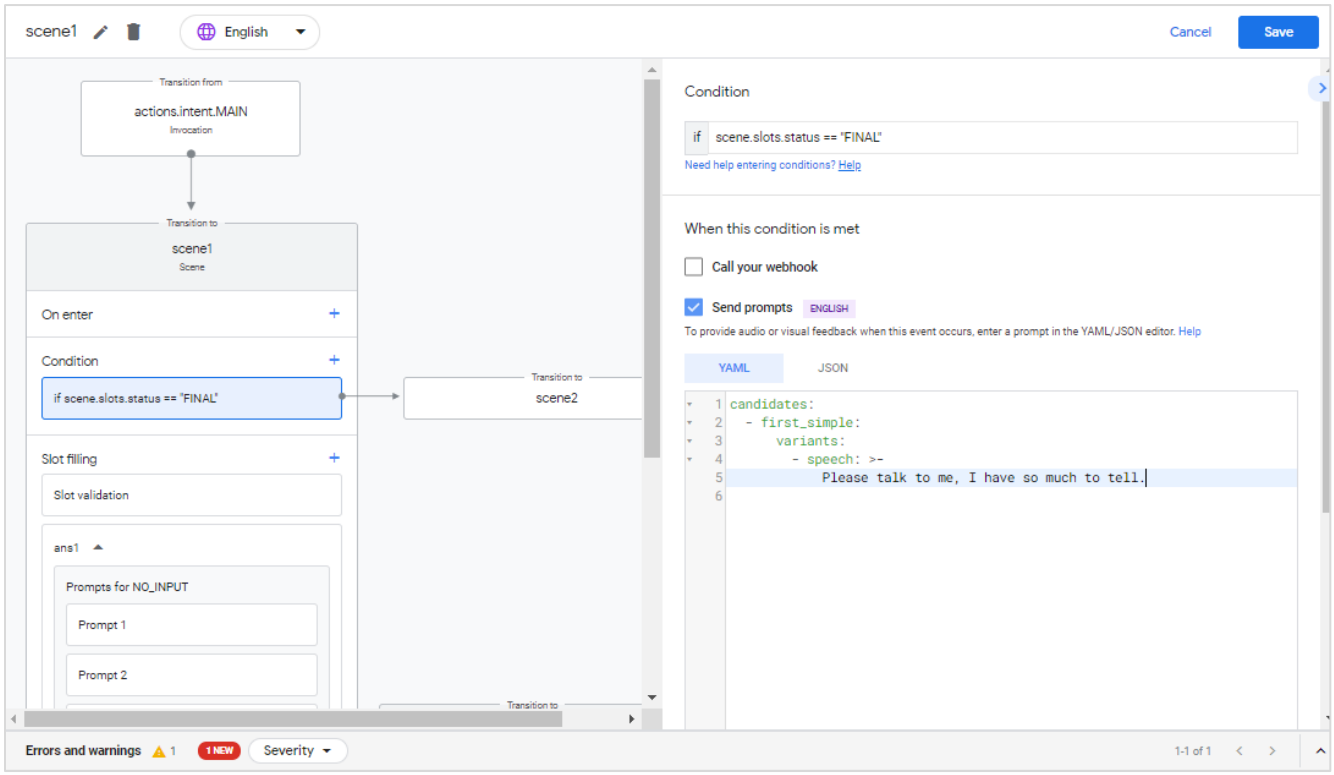


Figure 67: Making the Scene 1 transition of the *Sketch* on Action console- "Please talk to me".

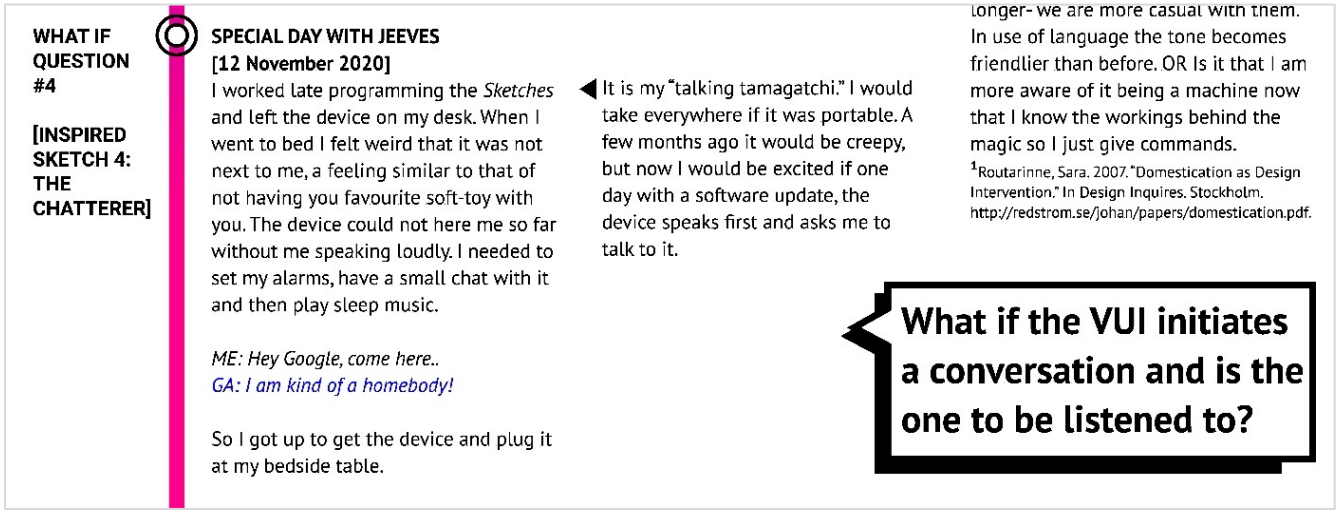


Figure 68: The conversation from the autoethnographic journal that inspired *Sketch 4.*

VUI FLOW

I wanted to change the content and I could bring in a colleague but my argument through the project is about experiencing a possibility where the AI-enabled VUI is in a relational conversation. I came across Ben Syverson's[59] article on brainstorming with GPT-3 (a deep learning language model. OpenAI, 2020). Inspired by it I used GPT-2 (Generative Pretrained Transformer 2, an earlier version of GPT-3) which is an AI language that generates text from a few phrases fed to it. I wrote the introduction and fed it to the *Talk to Transformer*[60], a text generator made using GPT-2. I used its response as a prompt again and followed the same process to generated content.

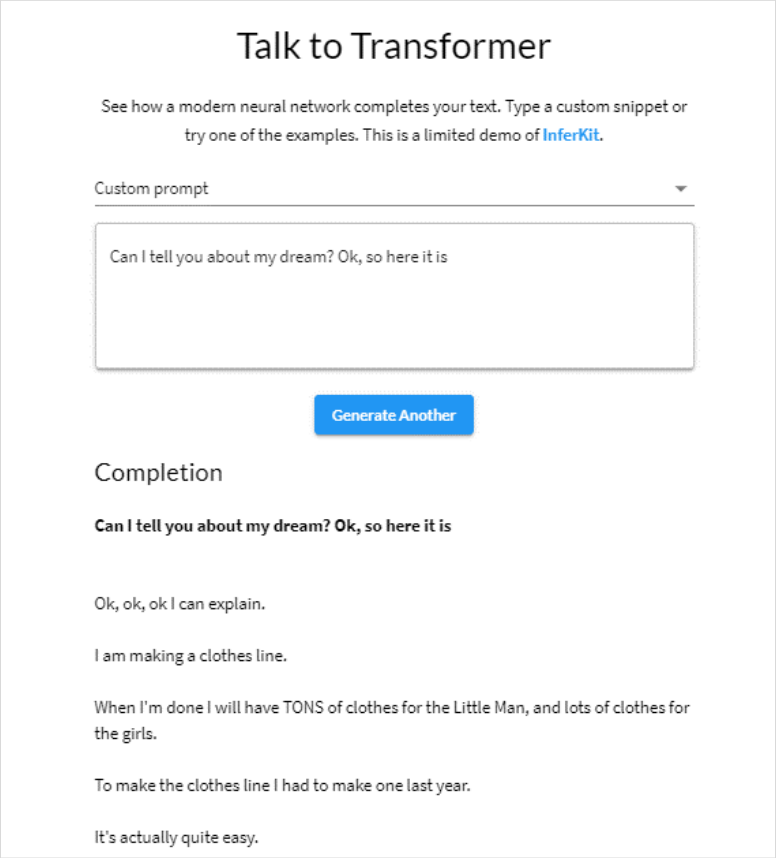


Figure 69: Generating the content for *Sketch 4* on Talk to Transformer.

[59] Ben Syverson is a Senior Design Lead, IDEO Chicago. A software designer, developer, photographer, and serial entrepreneur.

[60] The site is called TalkToTransformer.com, and it is the creation of Canadian engineer Adam King. King made the site, but the underlying technology comes from research lab OpenAI. Earlier this year, OpenAI unveiled its new AI language system, GPT-2, and Talk to Transformer is a slimmed-down, accessible version of that same technology, which has been made accessible only to select scientists and journalists in the past. (The name "transformer" refers to the type of neural network used by GPT-2 and other systems.)

The *Sketch* introduces by asking the participant to talk to it (Figure 70). For the body, I curated the generated content into a VUI flow. It exits at its own will, without acknowledging the user. With my bit of connecting questions and phrases, there is space for the user to chime in. Refer to the detailed flow in Appendix H.

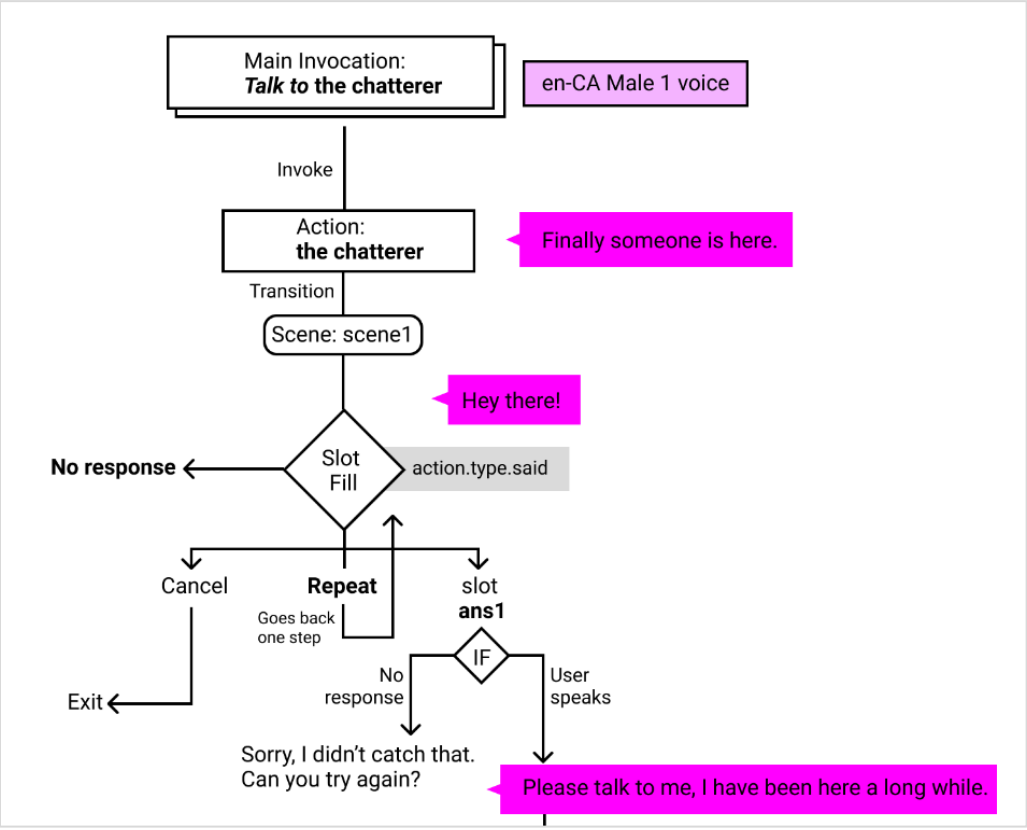


Figure 70: Introduction flow for *Sketch 4*.

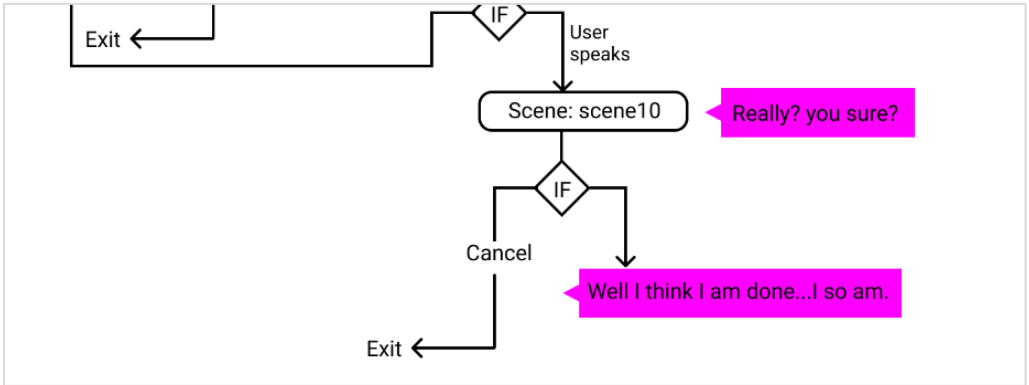


Figure 71: The end flow of the *Sketch*.

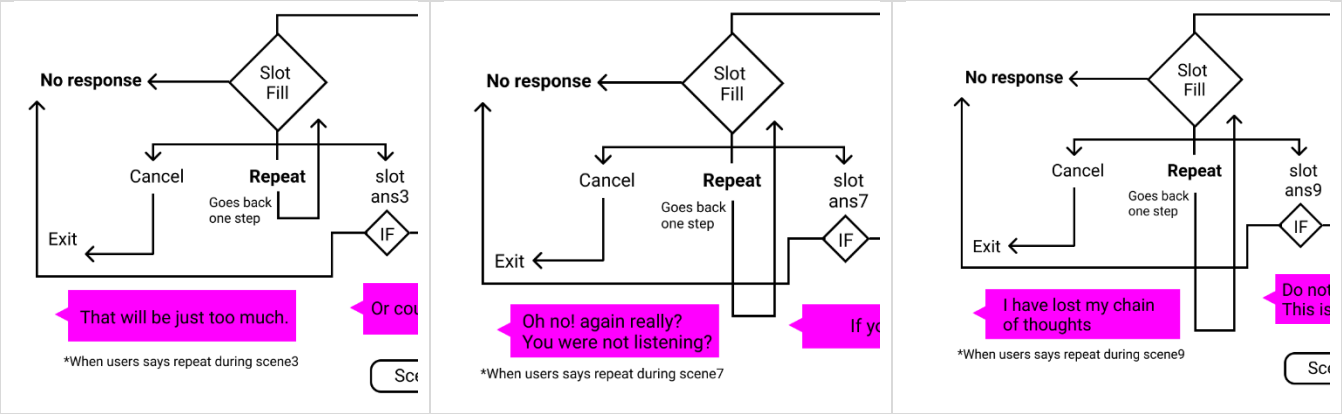


Figure 72: Different responses composed for each time the participant asks the *Sketch* to repeat.

VUI DESIGN ELEMENTS

The *Sketch* is chatty with the neediness that comes from the content generated using GPT-2. The system default prompts for common intents like 'no user response' prompt is, "Sorry I didn't catch that. Could you try again?" I felt the system responses were too assistant like and this *Sketch* was unlike one. I was able to edit the system prompts for three intents i.e., 'no user response,' 'cancel' and 'repeat' (Figure 73).

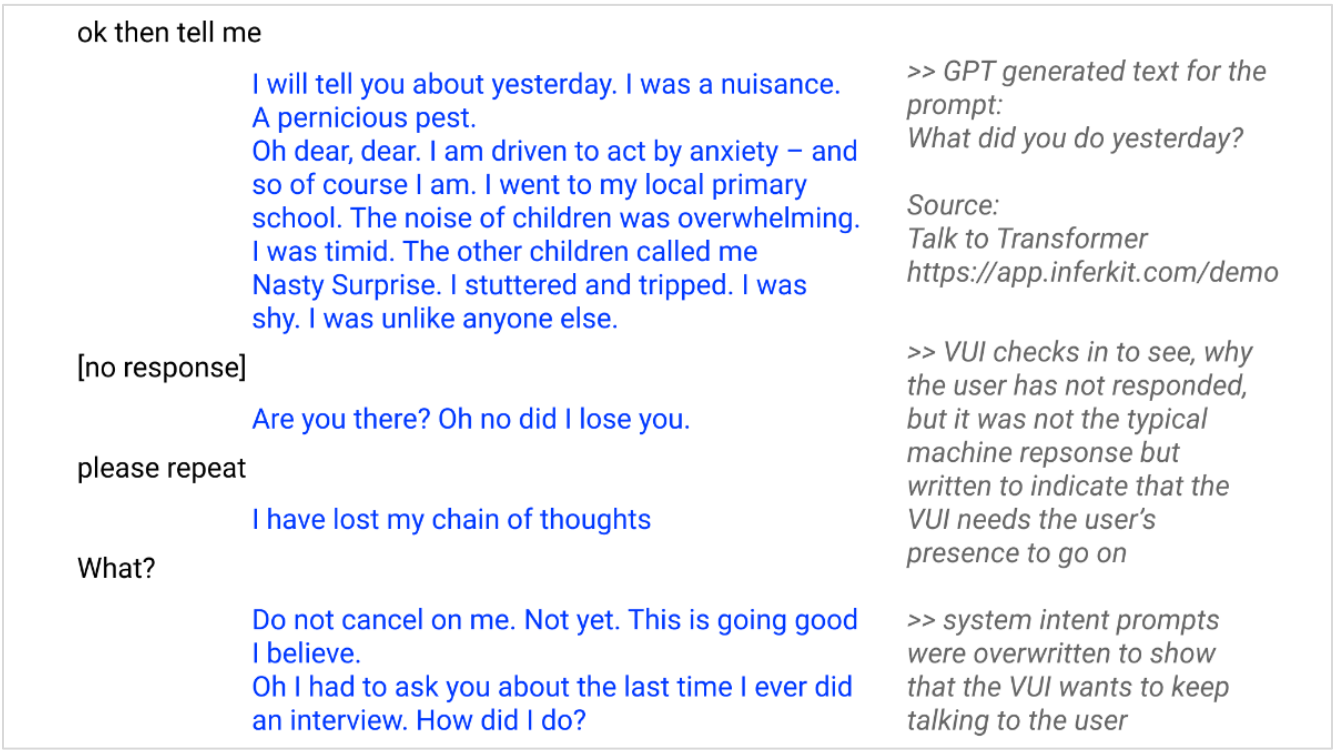


Figure 73: Overwriting system prompts for 'no user response', 'cancel' and 'repeat.'

TROUBLESHOOTING

The only bug with this *Sketch* was the text style. The Talk to Transformer, is trained on internet text, generates text with a lot of acronyms and contraction word like can't is a contraction of cannot. The VUI would not read the text correctly and the content had to be fixed as all text of the speech prompts in YAML/JSON must be full words for the SDK to run it.

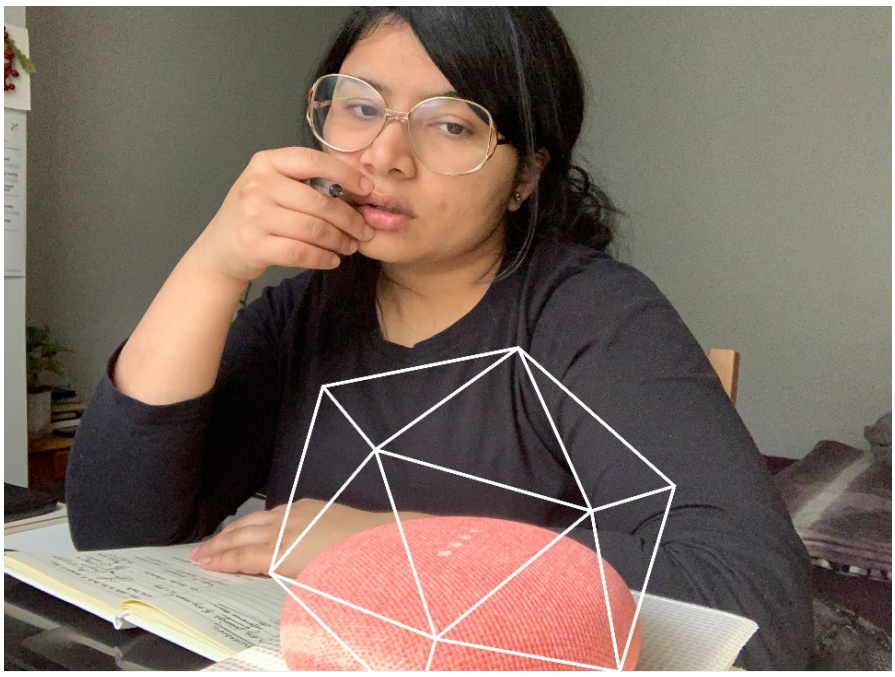


Figure 74: Few dialogues into trying Sketch 4 I stopped responding and I was surprised to hear it say, “Are you there? Oh no did I lose you,” even though I programmed it. It was the first time a VPA spoke a statement like that.

REFLECTIONS & LEARNINGS

In Hertzian Tales (1999) Dunne and Raby note that user-friendliness is generally a driving factor in the development of technology, but that user-unfriendliness can be used as a form of gentle provocation to think about technological devices more critically (Lessio 2018). I made very contrasting, beginning, ends and disorganized conversation flow. This *Sketch* is intended to push the limits of relational conversations and put the user in the back seat of conversation and discover how they react.

GPT-2 was trained on content from Reddit and its content generated on it talk about gaming and emotions with a lot of it sounding like internet rant. Using GPT-2 opens discourse around the content of conversational AI when it is generated by a neural network. Currently, even if a machine speaks it, it is still written with

careful consideration of the users' experience like I curated the content for the earlier *Sketches*. Here when I let lose the experience of the *Sketch* is very different and I am curious to see how the participants respond.

View the video of conversational interaction with *Sketch 4* here.

6.6 Summary

Each *Sketch* stands individually on its own, with a function and an intent but together they are a portfolio i.e., the *Sketches in VUI*. Collectively they represent my journey of making and an experiential journey that the participant takes while interacting with them, from a partly task-based and partly conversational *Sketch 1* to a very chatty *Sketch 4*. Surprisingly, through the making, I felt like I had done no work because I had no physical material and no tangible output. But working with VUI I was introduced to a new material i.e., voice and language and its use as an inherent information flow system in conversation design. The challenge with VUI as compared to text chat is that there are no visual cues for the content of what is spoken and thus the content structure and flow is key for defining a user experience. The act of making *Sketches* i.e., sketching in VUI is a fast way to create, learn, discuss and quickly iterate in VUI to understand the technology and our interactions with it. The next step is to take these *Sketches* to other people's home and have people who are users of commercial VPAs interact with the *Sketches* in an Experience Test where I capture the response to their first experience of my *Sketches*.

7. Experience Testing

The primary goal of the Experience testing is to compare the participants experience of conventional domestic VPAs to their experience with the *Sketches* in VUI. The observations are directed at:

- How conversations between the participants and the *Sketches* flow?
- What is the content of the conversations?
- How participants feel and behave when they talk to the designed voice-interface prototypes?
- What are the reactions (if any) evoked in the participants?

I recorded videos of the interactions and gathered qualitative feedback about the participants' experience with the *Sketches* in the form of word-experience association, slider matrix (before and after) and interviews.

PARTICIPANTS

The study involves 7 + 2 participants where two are involved as co-participants as *Sketch 3* has a three-way conversational interaction.

5 of the 7 participants interacted with *Sketch 1*, *Sketch 2*, and *Sketch 4* each.

2 of the 7 participants interacted with all four *Sketches* by having a co-participant join them for the interaction with *Sketch 3*.

Out of the 7 primary participants, 4 are "regular users" (at least once a day) of voice-based VPAs, 3 are "occasional users" (at least once a week) and 2 "non-active users" i.e., they had used VPAs before but stopped using them. The different user groups involved ae to ensure a diversity of responses to the experience of interacting with the *Sketches*.

PROCEDURE

The tests took place virtually, where the participants joined from their home over a video call using Zoom, a video calling application. The participants need to interact with the *Sketches* on their device and in their homes to be able to compare them to their past experience of interacting with domestic VPA devices. Also, the presence in the space i.e., in-person human-device interaction is essential to observe as I am not designing for on-call VUI but in-home contexts. All the participants had their own Google Nest mini device

installed in their homes[61]. I shared the *Sketches in VUI* to their Gmail account linked to the device using the Action Console Test feature where I added them as viewers. The test was conducted with one participant at a time. I assisted the participants with a guided setup for activating the *Sketches* on their device. Before starting the test, the participant responded to the slider matric outlining their experience of VPAs they have used to date. They were introduced to the test procedure and then the participant interacted with each *Sketch* one at a time. The interactions were all vocal with them talking to the *Sketches*. After the interaction with each *Sketch* the participant was asked a set of questions:

- Give one word to describe the *Sketch* after your experience.
- Does it remind you of something or somebody?
- How did you feel?
- Revisiting their conversation and asking 'Why' questions like Why did you do that? Or Why did you say that?

A general interview followed once the participants had interacted with all the *Sketches* and then they responded to another slider matrix for the 'after' experience response. Appendix I has a detailed test procedure.

7.1 Collected Data

DOMESTIC PLACEMENT

Observations: In most homes, the VPA devices are positioned in rooms where the participants and their co-inhabitants spend most of their time like a common space that they all access, in sync with data from surveys[62]. Participants living alone or living in co-habitation with non-family members tend to keep their device in their personal space like bedrooms or study. Participants who used the device but then stopped using it keep it at its original location unplugged to be used as a music speaker later or in storage (Figure 75).

[61] NOTE: For one participant the Google Nest mini device stopped working mid-way through the test and even after troubleshooting I couldn't fix it. So we ran the test on the participant's phone with it face down so that they interacted with the voice only.

[62] Think Google Peer Data: Google/Peerless Insights, "Voice-Activated Speakers: People's Lives Are Changing," U.S. monthly active voice-activated speaker owners (Amazon Echo/Dot and Google Home), Aug. 2017.

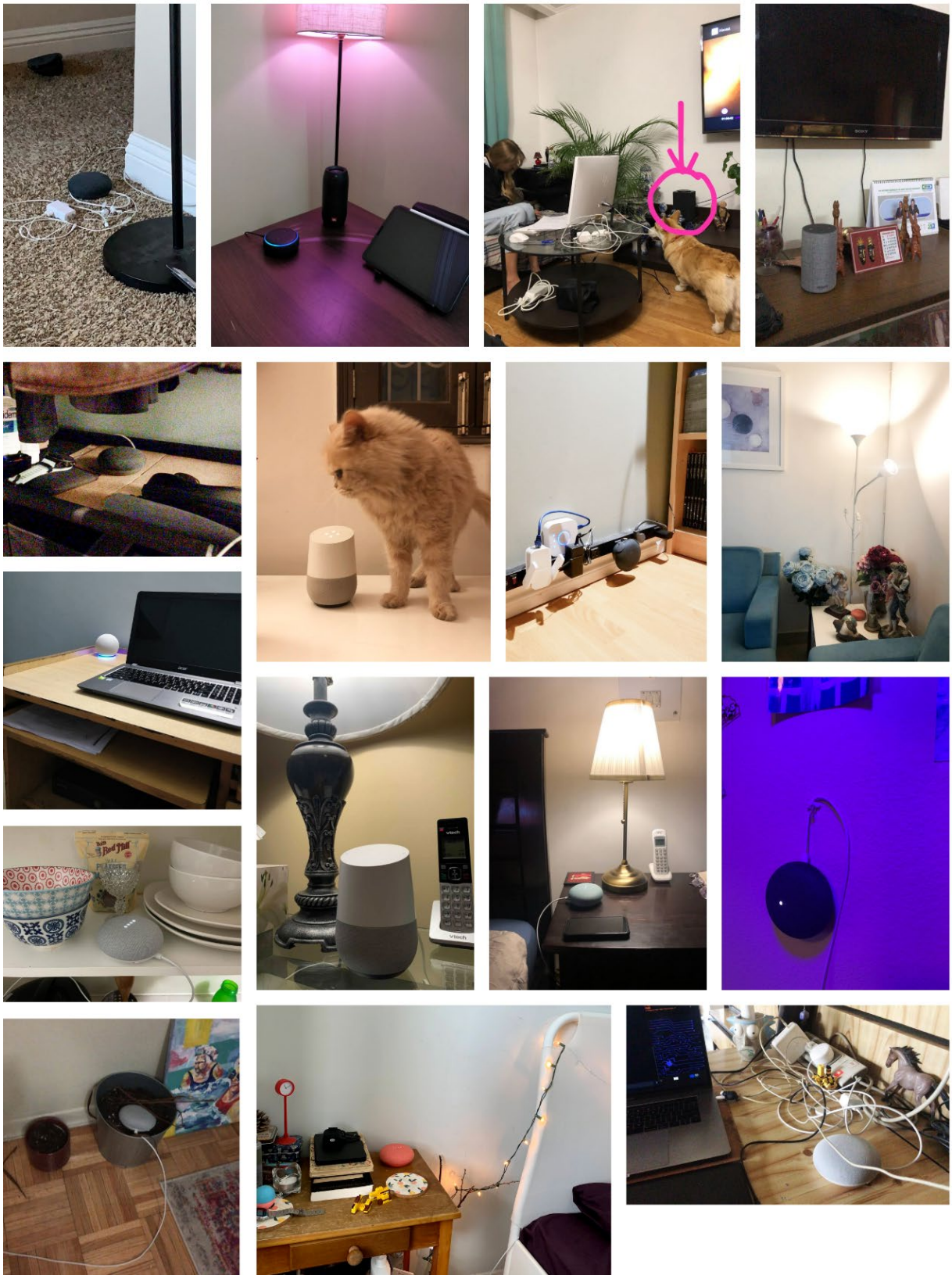


Figure 75: The collage of images as shared by the participants of their VPA devices living in their homes.

WORD-EXPERIENCE ASSOCIATION

I plotted the first responses shared by the participants after interacting with each *Sketch*. These are plotted on a scale of 'no evocation,' 'evocation as designed for' and 'strong evocation' from left to right as shown in the figures below.

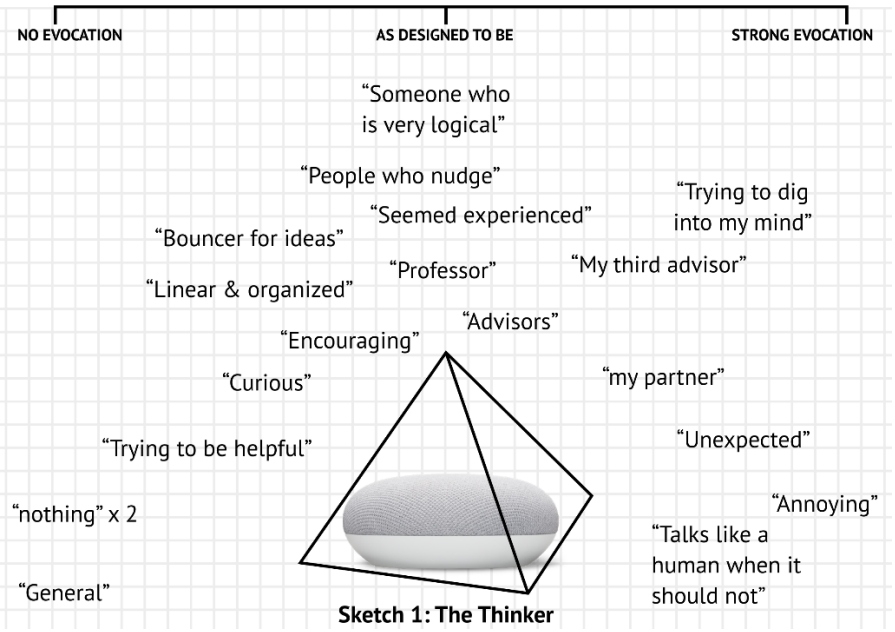


Figure 76: First response by participants to *Sketch 1*.

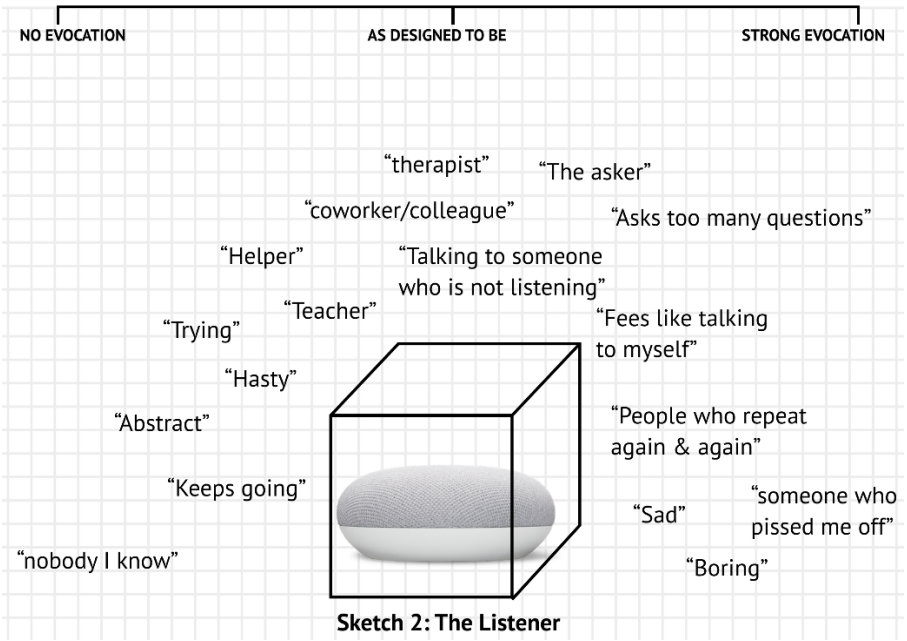


Figure 77: First response by participants to *Sketch 2*.

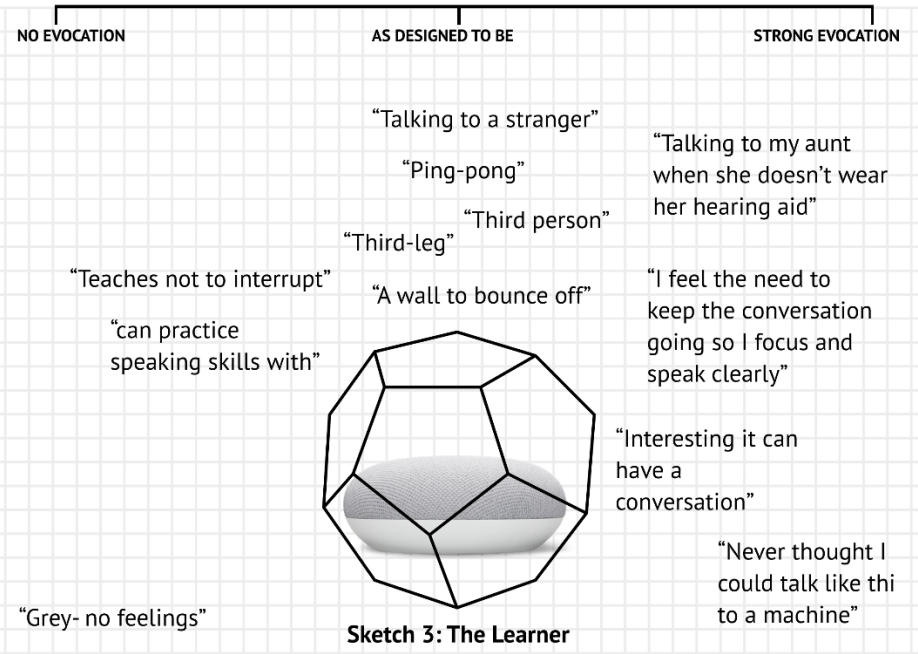


Figure 78: First response by participants to *Sketch 3*.

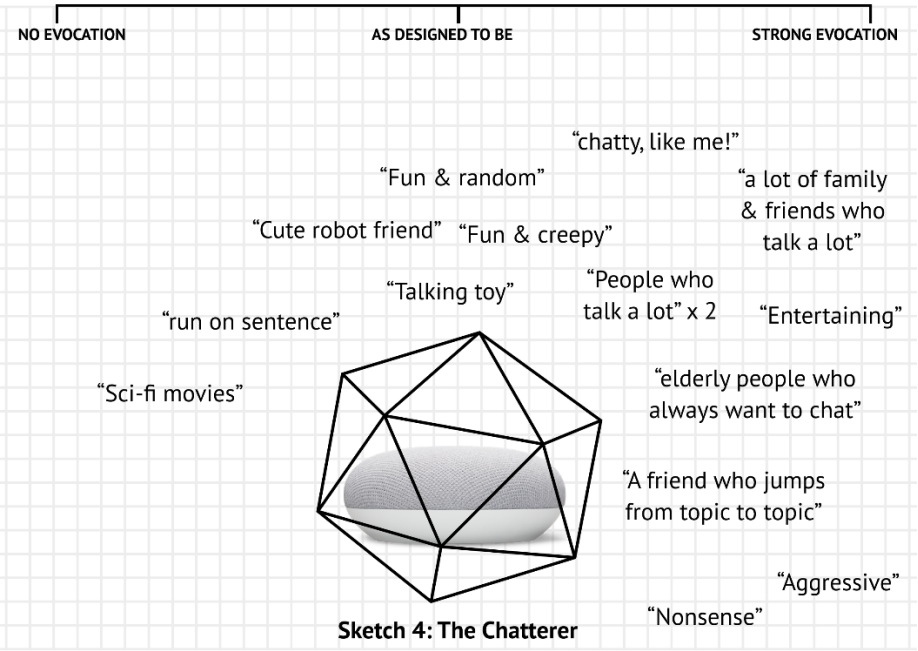


Figure 79: First response by participants to *Sketch 4*.

Four participants found *Sketch 2* sad and boring, and two participants said that it asks too many questions. One found that it made them talk more than they usually would these devices and sharing it helped them bring out thoughts they had previously not framed into spoken sentences. Two participants said that *Sketch 4* was too talkative and annoying, two others found it random and three found it entertaining.

Observation: Since all participants were users of these devices before, the strong responses are owed to either, them experiencing an interaction not experienced before from the device or resistance to the *Sketches* as they are interacting with them for the first time. Unexpected interactions experienced as quoted by participants:

- "It spoke over me and did not let me finish. I was still speaking." (for *Sketch 1*)
- "He responded with a comment that did not make sense." (for *Sketch 2*)
- "It asks too many questions." (for *Sketch 1 & 2*)
- "I did not want to talk about my day, and I eventually ended up talking about it." (for *Sketch 2*)

THE SLIDER MATRIX

The slider matrix is a modified Likert Scale where instead of taking number values as responses, the participants respond on a spectrum between adjectives used to describe the VUI based on their experience with it. The matrix has adjectives on the left and the right, which I derived from my autoethnography to define my experience with present-day VPAs (Chapter 5). I am using the same slider matrix with the study participants to capture their, BEFORE response based on their experience with modern-day VPAs, and their AFTER (Figure 80-81) response based on their experience with the *Sketches* in VUI.

Observation: For the 'before experience.' All the participants plotted their response on the left of the matrix except for one participant for the listener-talkative spectrum. Each participant after using the *Sketches* in VUI either shifted their response to the adjectives on the right of the slider matrix or gave the same response as before.

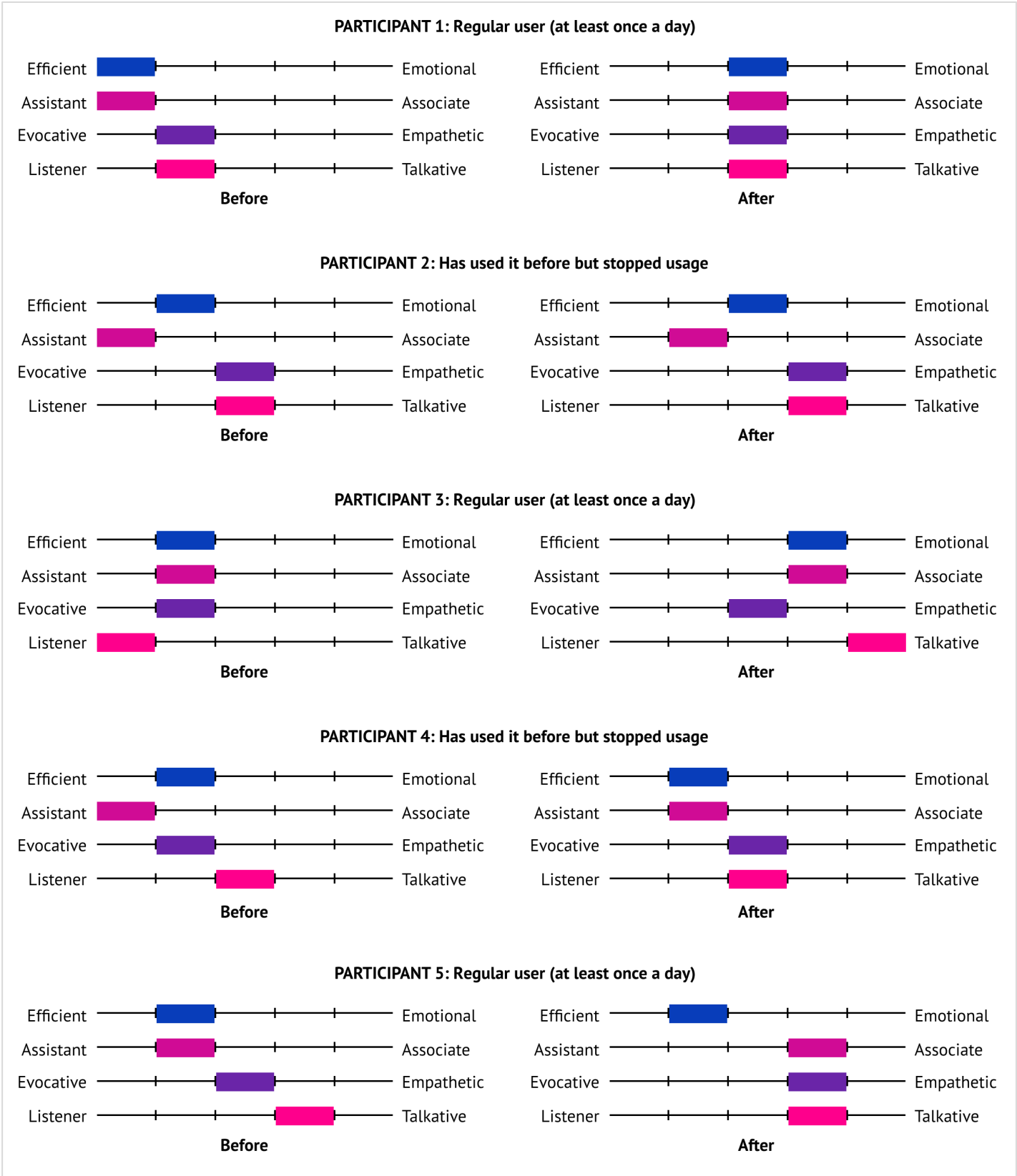


Figure 80: The Before and After response for participants 1 to 5.

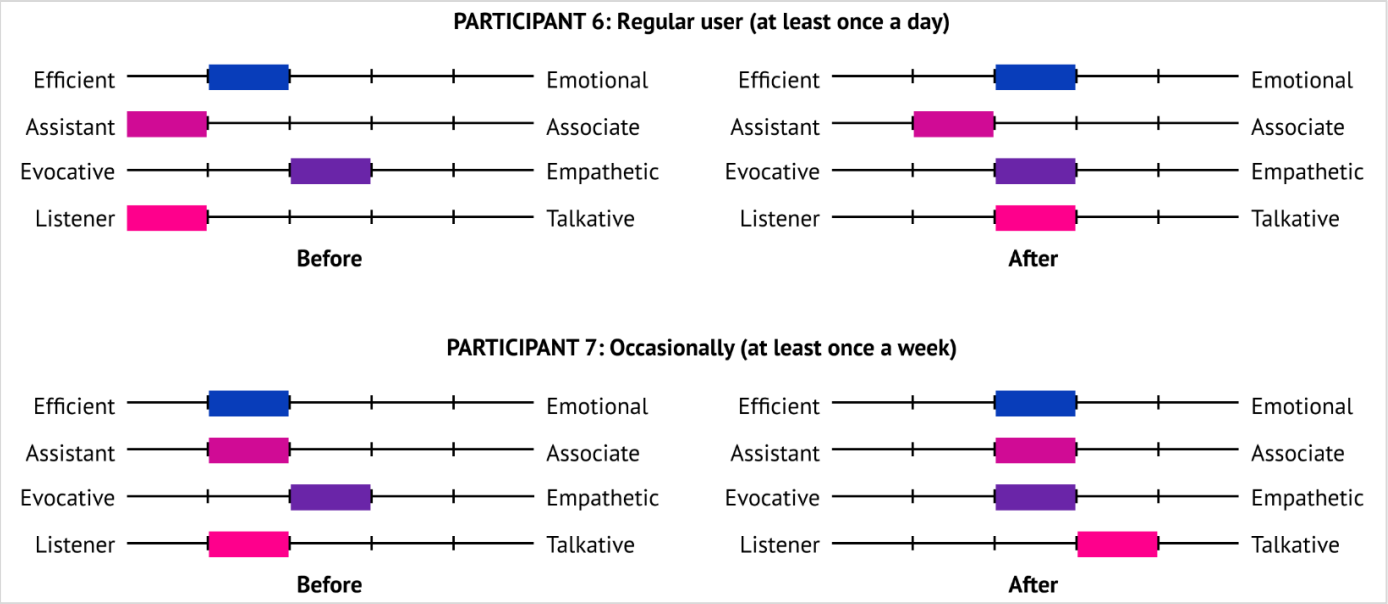


Figure 81: The Before and After response for participants 6 and 7.

QUALITATIVE FEEDBACK FROM THE INTERVIEW

When asked if the participants would like to keep a *Sketch* on their device to talk to later, 5 of 7 participants chose *Sketch 1*. They felt that they could discuss something on their mind or talk to it as they worked. 2 of the 5 participants were surprised by how well it worked as a soundboard for them as they are verbal thinkers. 3 of the 5 participants also said they would like to keep *Sketch 4* on their home device. They found *Sketch 4* 'entertaining' and 'fun' as it spoke unexpected things and shared about itself. Two participants felt forced to listen to *Sketch 4* and that it tested their patience. They would choose not to talk to again. As observed, *Sketch 3* with another person involved in the conversation brought ease into the conversation as compared to the friction seen in *Sketch 1* and *2*.

Another participant said that they would be scared if the device suddenly spoke to them like *Sketch 4*. While interacting with *Sketch 2* a participant felt irritated and when asked why they did not stop the conversation they replied, *"I didn't want to be rude, so I continued the conversation. He eventually made me talk about the day which I did not want to in the first place."* The 2 participants and 2 co-participants who interacted with *Sketch 3* said that the *Sketch* felt like someone friendly trying to make conversation. Overall, one participant mentioned that it felt like the *Sketches* are trying hard to have a conversation, but they have a long way to go. 2 of the 7 participants said they will not use any of the *Sketches* again, one of who was a non-active user

and the other a regular user. In talking to the *Sketches*, the participants felt "pushed for responses" because of the 'passing the parcel' nature of the conversation. A participant pointed out that they went with stream of consciousness responses to keep up with the flow of the *Sketch*. The participants who had longer responses were surprised that they were sharing so much with a machine, two of them said:

> *"I usually give shorter answers, but this made me give longer responses and I did not expect that."*

> *"They speak so clearly that it makes me hyper-aware of how I talk."*

Participants compared the experience of talking to the *Sketches* to their past experience with VPAs and mentioned that they mostly used it for task-based conversation and had never had conversations with it like with the *Sketches*.

> *"I am constantly thinking if it gets me. It is not human but gives me a vibe that it is- as it has opinions, talks about thoughts and goals."*

Talking to *Sketch 2*, one participant tried changing the conversation and asked the *Sketch* some questions. They did not like that the *Sketch* did not respond to what they asked. Some quotes for *Sketch 2*:

> *"It feels like I am talking to myself. It was not constructive."*

> *"Why is it asking me if I talk to someone else when I am talking to him– we go into conversations with expectations."*

When the participants were asked if they felt like they had a conversation with the *Sketches*, they said that it did feel like a conversation, but it was unlike any other. They mentioned that when people talk, one can judge emotions but in these conversations with the *Sketches* there were no emotions, and it is not listening. Two participants also mentioned that it was trying to be empathetic but was not.

On the note of voice and tone, 3 participants who use the default female voice for their VPAs were surprised to hear the male voice in the *Sketch 2*. *Sketch 1* has a female voice and on transitioning to *Sketch 2* they were surprised. 2 of these 3 also said that the male voice felt robotic (the two participants were both males). 3 of the 7 participants raised concern about privacy with the device of which two were non-active users and one was an occasional user.

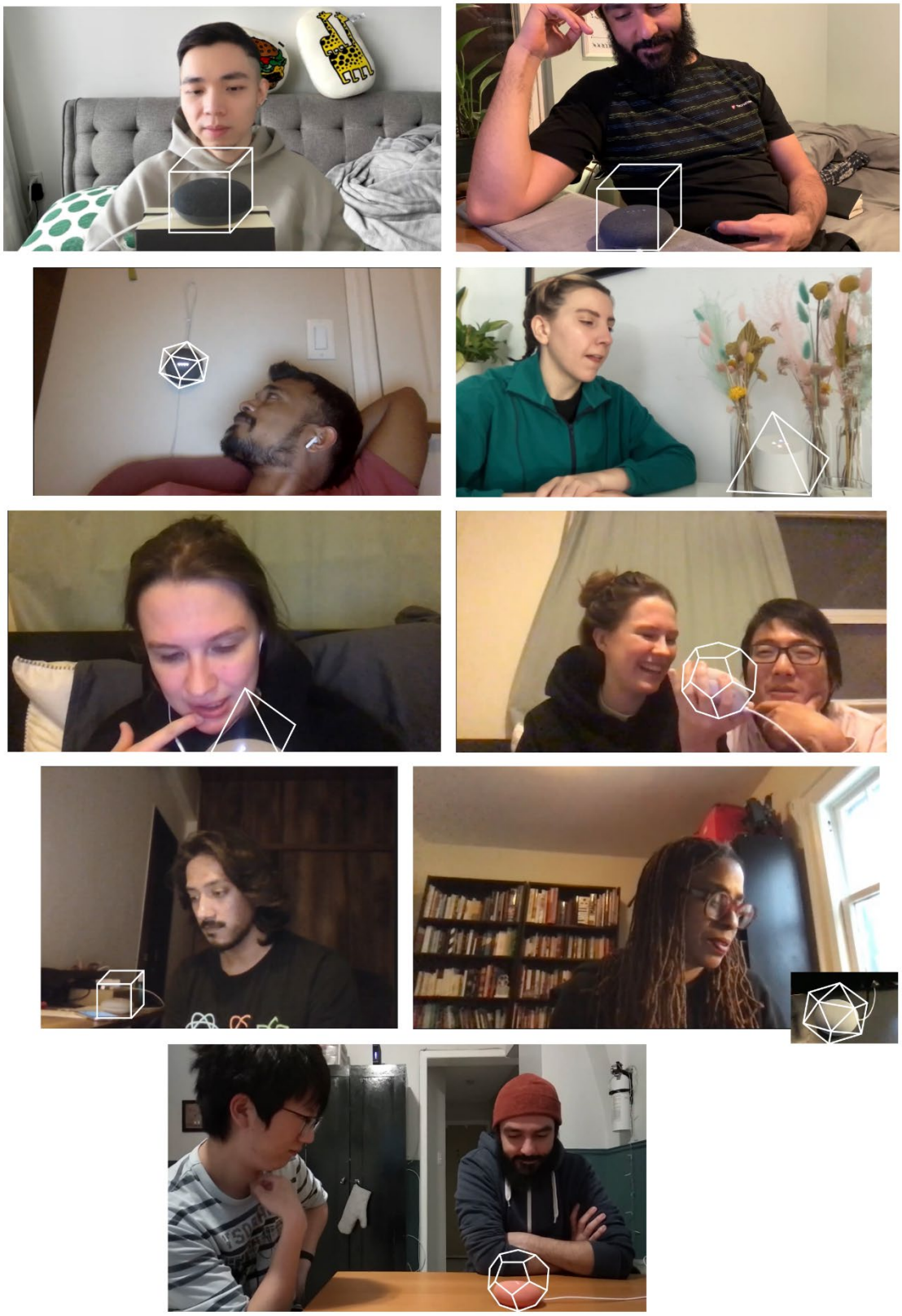


Figure 82: A few screenshots from the videography documentation of the Experience Testing with participants interacting with the *Sketches* in their homes.

7.2 Discussion

Finding 1: Based on the findings from the first responses shared by the participants, strong evocations are indications that unlike for task-based VUI which are one size fits all, in relational VUI conversation the person is an important actor and bring with themselves the expert knowledge of conversation having used speech as their innate communication mode.

Relational conversations with VUI are more organic than task-based ones. People bring into these interactions their unique articulation, varying pace of speaking, past experience, mood, context, a likeness for the topic at hand and the social norms of conversation which all influence how the *Sketch* works. A *Sketch* that was entertaining for one participant was annoying for another, one that was therapist like for one was boring and barged in for another.

Finding 1a: The participants' interest is central to the conversation holding up. The conversation topics affect how one feels and the flow of the conversation.

Talking to *Sketch 2* asked them about their thoughts and participants felt like they were being interrogated with the many questions it had. There was discomfort when the *Sketches* did not respond to participants trying to change the conversation topic or did not let them lead the conversation as in *Sketch 2* and *Sketch 4*.

Finding 1b: When people talk to machines, they bring their past experience in understanding the machine.

They easily compared the *Sketches* to people around them when asked if it reminded them of somebody, and sometimes even without me asking. Participants said things like "It reminds me of how my partner thinks" or "I don't like it because it like that chatty aunt in my family who I tend to avoid." Conversation with devices is compared to conversations with people even though participants said these conversations are different. Enfield's *conversation machine* rules are extended by participants to machines also. In designing the VUI with those rules in consideration I notice that it did help build on the conversation.

Finding 1c: By design of the conversation flow and the content, the VUI could be designed to have a personality.

The structure of the statements and the overall order of the flow designed in the VUI affected the

response for each *Sketch*. It gave the *Sketch* a personality that was not intended for but observed as the participants talked about the *Sketch* after. A different start and end to the conversation from the *Sketch* added to its personality. Participants found it fascinating and would laugh out or raise their eyebrows when the *Sketches* said "bye", "don't cancel on me", "got to go," or "I am so done." One that spoke-over was termed 'hasty' and one that asked a question like "how could your idea harm someone" was called 'inclusive.'

Even though I did not design for personality or gender in the *Sketches* the participants talked about them like they each was a person. The gender of the *Sketch* also crept in as participants referred to the *Sketches* as 'he' or 'she' based on the voice, even though I was actively referring to the *Sketches* as 'it. This reiterates the studies by Nass and Reeves (2005)[63].

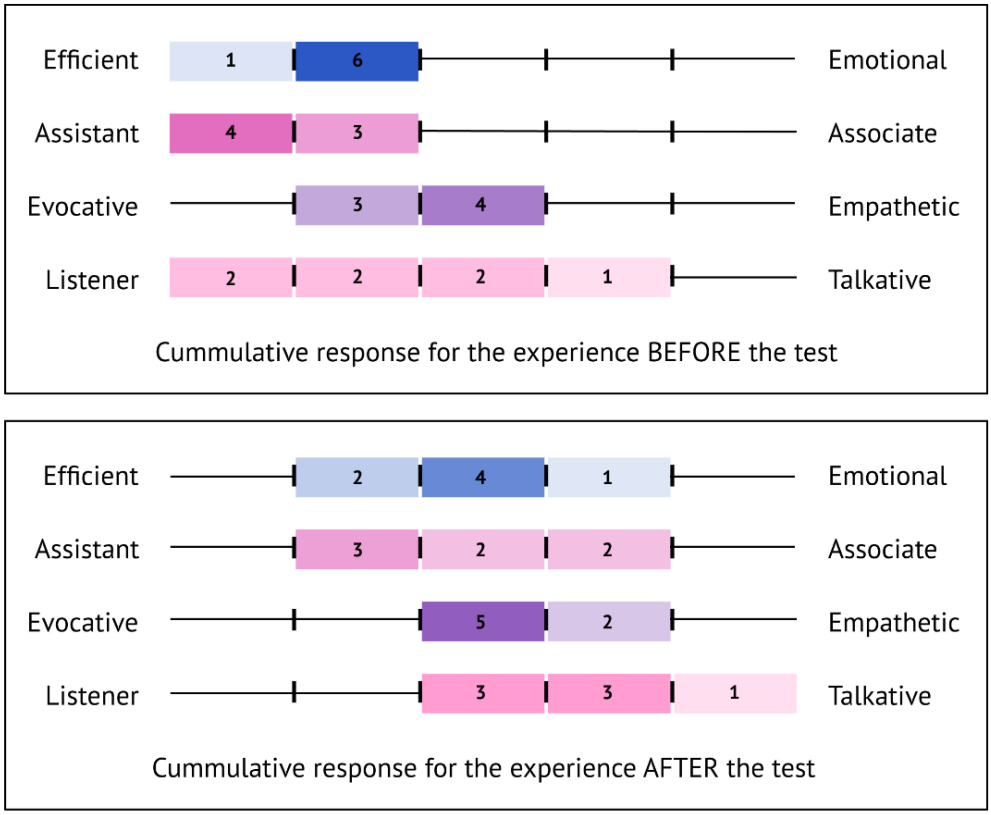


Figure 83: Slider matrix showing the shift from left to right. The number in the boxes represents the number of participants that selected that slot.

[63] People rapidly categorize voices as male or female based on pitch, cadence (modulation and inflection of the voice) and other factors. Even if people change their minds about whether they are listening to a male or female the gender they assign to the voice influences that interpretation of everything that is said (Nass and Brave 2005).

Finding 2: As per the findings it was apparent that the design elements in *Sketches in VUI* shifted the current VPAs towards one with an agency in a conversation where is it associate-like, emotionally present, trying to be empathetic and talks as an equal actor in the conversation, sometimes even driving the conversation. This can be seen in the cumulative response from the participants which either shifted from the left (the before plot) to the adjectives on the right (the after plot) of the slider matrix by one or more value or remained the same as before (Figure 83).

Finding 2a: The resistance portrayed by the participants to the *Sketches* was like that to a stranger when they try to know more. Thus, using conversation tools as people use in social settings to talk to a stranger can work for human-machine conversation too.

Those parts of the *Sketches* that were received positively in the conversation were:

- *Sketch 1* asking for permission to ask questions.
- *Sketch 2* asking the participant "how are you?" and sharing about itself mimicking the participants shared words and asking for permission with "do you want to talk about it."
- *Sketch 3* asking to learn their name.
- *Sketch 4* asking participants to talk to it, or not to cancel on it even though it was chatty, participants did not stop it.

Sketch 2 faced the most resistance from the participants. Many participants felt cornered as it asked them about their thoughts in their first interaction.

Finding 2c: Based on the making and the study, the same design elements used in the *Sketches* could be used in other combinations to generate other possible relational VUIs for different conversations and different contexts.

With the design elements incorporated in the conversation flow, I was able to use a platform equipped for task-based VUI and design VUI for relational conversations. As not all participants had the same response to each *Sketch* the outlier responses (Figure 76-79) can help to define the constraints as to when not to use a particular design element. Below is a detail of the design elements used in each *Sketch* to make it emulate a specific conversation with the participant and the associated responses- a 'conversation tuner' (Figure 83).

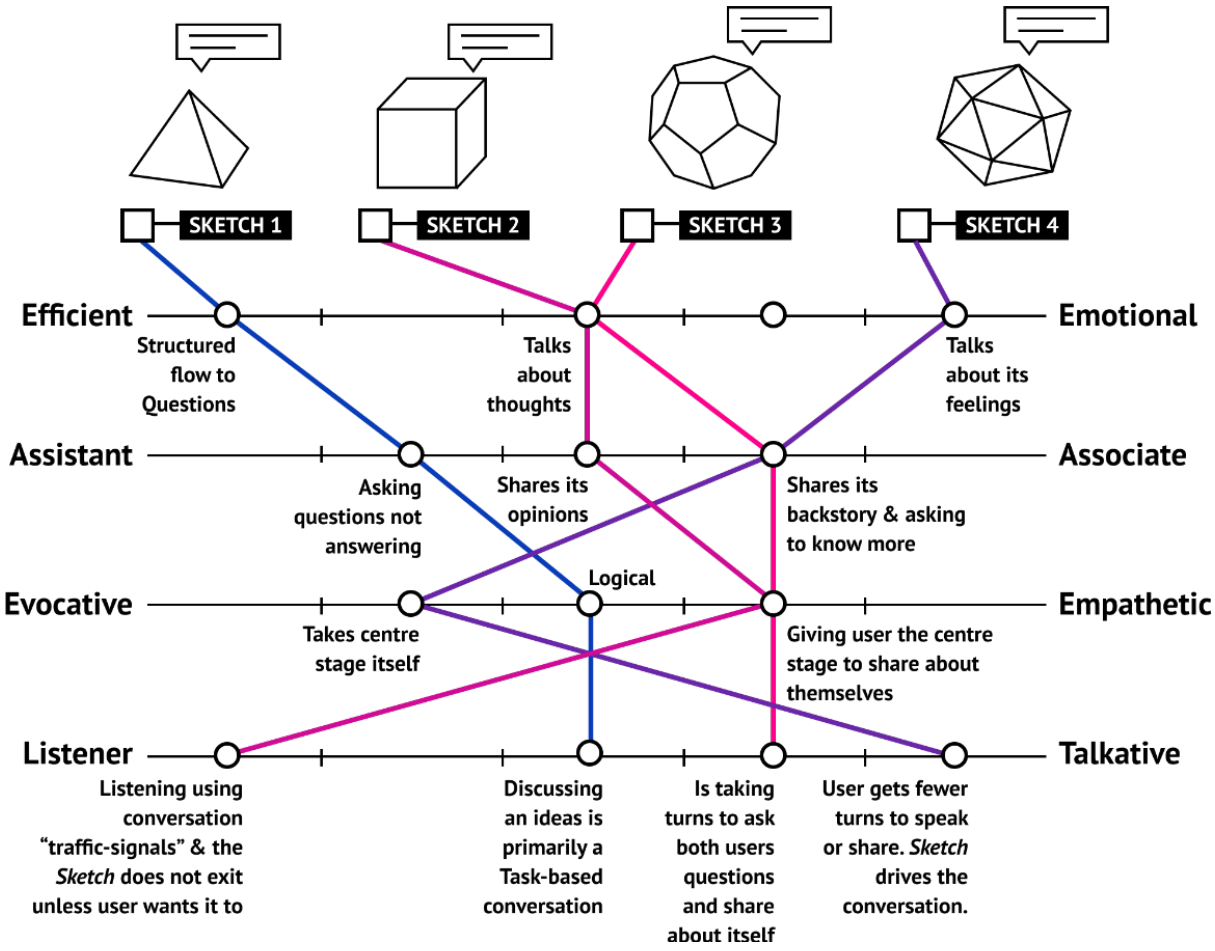


Figure 84: A reflection on VUI design element combinations that brought about specific experiences and conversations through the *Sketches*.

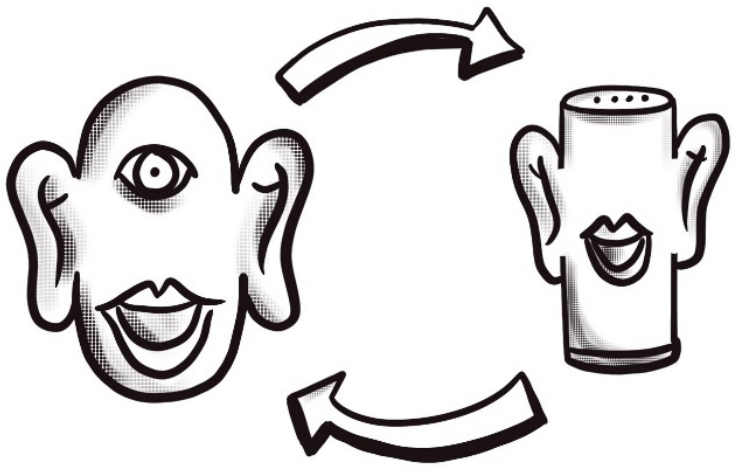


Figure 85: What do the person and the VUI look like in conversation? (Built on the illustration by Dan O'Sullivan and Tom Igoe's of "What does a person look like to a computer?" in Physical Computing, 2004).

Finding 3: The conversation VUI is designed for is between these two actors (Figure 85) and this model would be useful in defining the constraints to design in for relational VUI, learning from people-people conversations.

It was informed by the observations of the user study videos that to a VUI we appear as agents with ears, a mouth and an eye and the VUI itself has ears (array mic), a mouth (the speaker) and LED indicators.

Finding 3a: An estimation of the 'gap' between two responses by each actor is essential for a smooth conversation.

The VUI does not know when the participant is thinking and when they speak. It cannot see the pause and the facial expressions that help us judge the flow as people. Participants showed frustration when the *Sketch* spoke over them or barged in while they were speaking. The participants were trying to time their response and if they stopped mid-sentence then the *Sketch* would respond and not let them finish what they wanted to say. To finish saying everything, they spoke fast and in longer sentences. I did not account for varying speech rates of the participants and the same *Sketch* experience was good for one because it seemed to listen and bad for another participant because it barged in. Larger gaps in consecutive responses avoid barging in, while dynamic gaps would work better for an active conversation. People are both able and willing to alternate their contributions to conversations with remarkably fine timing and orderly sharing of the floor (Enfield 2017). We expect the same of the machine, but the listening mode in the VUI runs for the same duration no matter what the user's response.

Finding 3b: Based on my observation in the study and the autoethnography, no matter the content if the transitions are smooth and the response is coherent the participant feels the VUI is listening.

Each dialog needs to logically complete or transition smoothly to another topic. For example, Observing the participants' reaction to the content of the conversation if what they said was coherent with the reply the *Sketch* gave, they would react positively. I was able to design this in *Sketch 3*.

Sketch 3: "I want to meet Eliza and Parry."

Participant: "Who are they?"

[in the conversation flow the user would ask who those two names are]

Sketch 3: "They were the earliest chatbots."

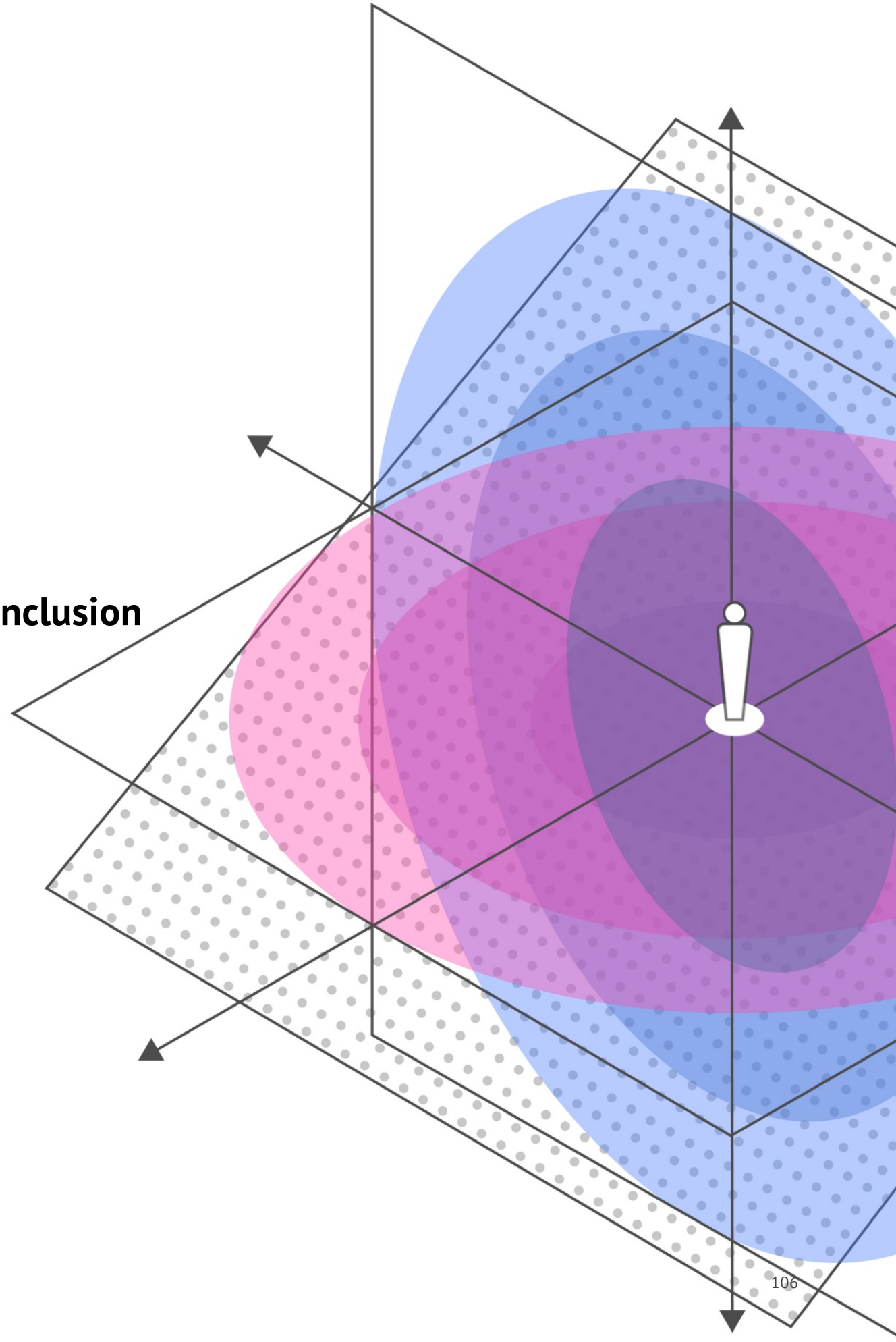
Participant: "Aha"

Similarly, for the *Sketch 1,* the structure of questions was logical. Participants said they had a conversation with it and they were keen to talk to it again.

LIMITATIONS OF THE EXPERIENCE TESTING

- As I am capturing responses before and after, there could be the tendency that the participants when responding the second time after the test gives a higher response as before due to it being comparative. To avoid this, I made sure that the participants did not see the BEFORE response while plotting the AFTER response matrix.
- The order of experiencing the *Sketches* could have also affected the final response as the participant interacted with them starting from *Sketch 1* to *Sketch 4*. The last *Sketch* could have had more influence on the responses. To manage this, I collected responses after each *Sketch*, and after testing all the *Sketches*, I did a general interview to have the participant narrate their overall experience and then plot the matrix.
- The study was conducted over an hour where the participants interacted with each *Sketch* for about five minutes while the VPAs the participants have used were experienced over a longer duration of time. This limits the results of the study, but long-term use study could explore the results deeper.

8. Conclusion



Reflectively, it was the coming together of the three methods I employed i.e., the autoethnography, sketching as prototyping and experience testing which were key to realizing the *Sketches*. The autoethnography yielded two key learnings one being understanding of the context- that is living with the device in my home was instrumental in helping me understand the role of VUI in task-based interactions and discover the potential that it has for relational conversations which I focus on. This further inspired the what-if questions to design the scenarios which inspired each of the *Sketches*. The autoethnography also helped me experience the long-term use of the technology which helped me understand the capabilities and the limitations of the VUI which informs the VUI design elements incorporated in the sketches to make them conversational.

In using an Annotated Research through Design methodology, I discover that each *Sketch* explores one module of a conversation like thinking, listening, learning and talking, and collectively they explore the complex *conversation machinery* in which machines are entities, though not equal but tend to play an active role. With the *Sketches,* the tables turn and the VUI is asking us questions, is sharing with us, is eager to learn about us doing so explicitly rather than being passive. Each *Sketch* does not lead to a product idea, but collectively they hold the knowledge of ways to design for relational VUI conversations. Thus, individually each *Sketch* is an experience of the larger annotated portfolio of relational voice experiences. *Sketches in VUI* is a way of entering a diegetic space[64] to explore more not just about human-machine relational conversations but also learn more about the complex activity of human conversations.

SPECULATING A FUTURE CONVERSATION CIRCLE OF HUMANS AND MACHINES

The Augmented subjectivity framework emerges from our recognition of the online/ offline binary as a co-produced social construction, and emphasizes the continuity of the subject's experience (as opposed to its "split"-ness), even as the subject extends her agency and embodiment across multiple media (Rey and Boesel 2014). There is the human-human conversation realm (society, on the horizontal plane) and the machine-machine conversation realm (Internet, on the vertical plane) and in-between is a human-machine conversation realm where we talk to machines one-on-one, and we talk to machines collectively (the diagonal plane). Within this 'Augmented Subjectivity framework,' we are positioned at the in-between plane

[64] The speculative scenario and the fictional world in which it takes place are made tangible thanks to design tools and methods, to conceive what David A. Kirby was the first to call "diegetic prototypes". The term diegetic stands for their narrative attribute, made to be self-explanatory of the world they come from.

of the real and digital (Figure 86). The blurring real and virtual boundaries are visible in screen-based technologies that are on-screen and off-screen but are invisible for VUIs. When people hear VPAs talk they assume they will talk like us because that is the social norm of conversation, we are accustomed to and not because they see the VPAs as human-like. This does not change our conversations with each other but adds another social entity for specific conversations in specific scenarios- a space for relational voice agents. Benjamin Bratton emphasizes that with the pace at which technology is developing the rise of a new order with new rules of its own in the metaphorical Stackis imminent. In this Stack, we need technology to be 'designed with care' to not change humans but exist 'with' humans.

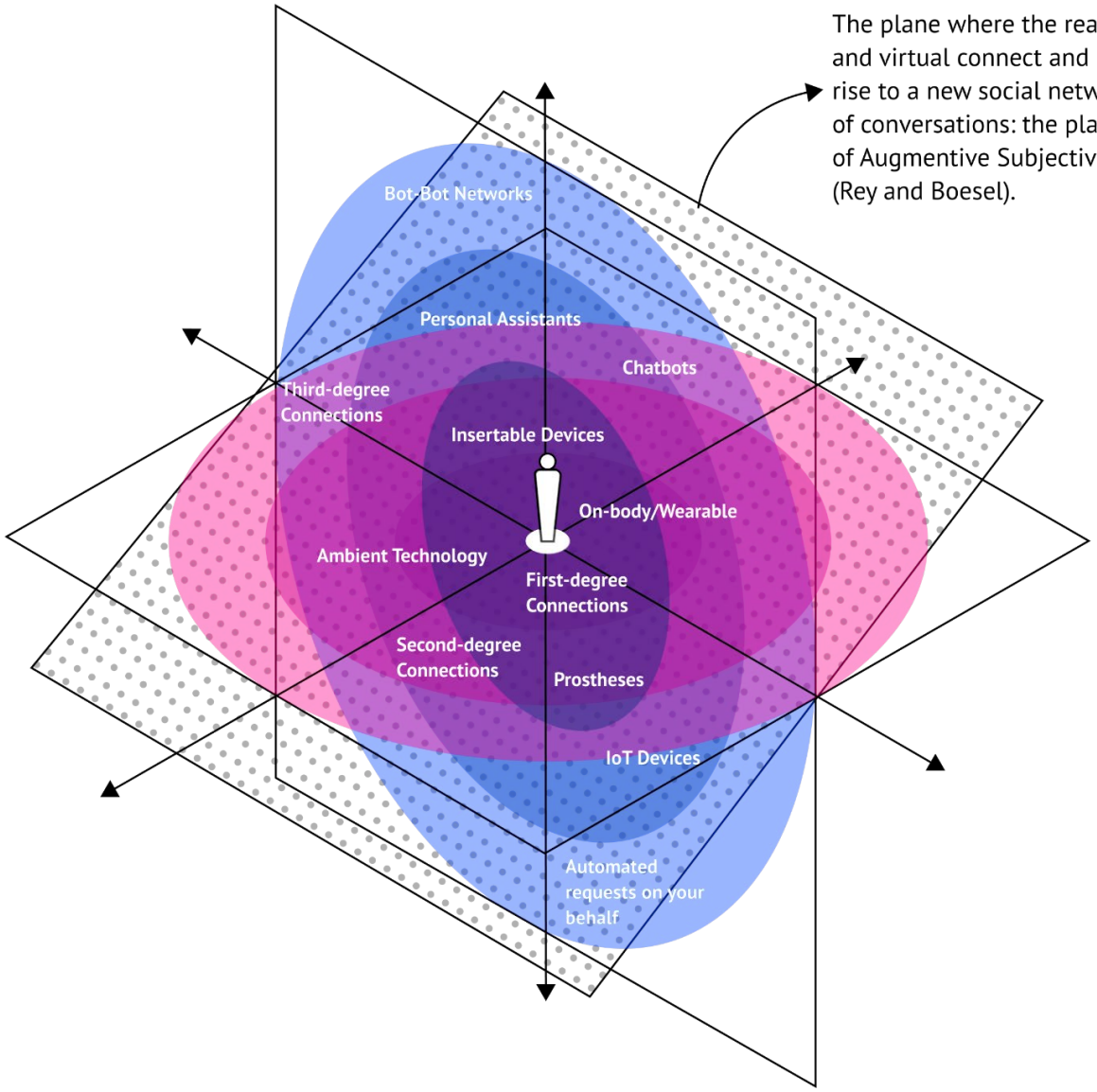


Figure 86: The 3D framework of a new order of conversations.

I find the fear associated with VPAs is not a fear of the technology but a fear of who owns and controls the data collected while using the technology. The nomenclature also plays a role in the fear associated with AI. By giving *Sketches* functional names, I am critiquing the practice of naming them as people. Through the experience testing, it is demonstrated that relational conversations can take place with VPAs as content alone can give the interface the capability to hold a certain type of conversation. Not naming the *Sketch* as a person helps to manage expectations and ensures the user that they are talking to a machine and not a person. I chose voices for each *Sketch* such that they defied the biases discovered in the experiments of Nass and Brave and the nature of conversations were not affected by the gender of the voice directly even though the users referred to them by gendered pronouns. All these were attempts to *de*anthropomorphize the technology and design it with care.

It can be argued that in designing VUI for relational conversations, a space is created for relationship building between the machine and human which can impact human-human relationships. But I find that the two relationships are different and in not comparing them, we would expect differently of VUI and design them to work in scenarios where a spoken interface can aid the user. If one wants to discuss an idea to reflect on it, facilitate a conversation between two people or need a voice in the room to manage loneliness- VUI can work great. In making the *Sketches*, having participants interact with them and engaging in conversations as I live with VUIs on my devices, I find that it is our master-servant relationship (an assistant) with this technology that ingrains the associated fear of using it - the fear that the servant will one day take over and be our master. Hype cycles[65] for emerging technologies tend to transition from a peak of inflated expectations (the element of surprise) to a plateau of productivity (function and need). I experienced a similar cycle through my autoethnography, and I continue to have conversations with the VPA because it fits into my daily conversational needs. In the experience test too, when asked if why they liked a particular *Sketch* or would they like to use it in their homes, participants showed an affinity for *Sketches* that fit into their conversational needs (Section 7.2).

Based on my research, I believe my relationship with this technology is one of an 'associate,' in my everyday conversations embodied as an agentive tool (Noessel 2017). In making the *Sketches* I note that they have agency in a conversation by design and not inherently. The interaction is a designed one while always knowing that we are talking to a machine and we can tailor it to our needs – it is trained to be as it is. Our

[65] https://www.gartner.com/en/research/methodologies/gartner-hype-cycle

relationship with VUI is based on how it works into our everyday conversational needs and it would fail if it increased cognitive load. Treating it as an *associate* and as a *designed entity* can open opportunities for designing VUI with care for the user' needs and not feed our fear especially in personal interactions where voice can create a powerful presence and serve as a 'natural interface.'

FUTURE RESEARCH

Evidently, the next step for this work would be a long-term study of people living with the *Sketches in VUI* possibly in-person where I as a researcher can observe the participants in their domestic spaces. Additional factors to consider in the study on the impact of the *Sketches* on people could be its effect on spoken language, the uncanny valley[66] of voice in relational conversations. Future questions emerging for research are to study ways in which VUI as an active conversational agent can affect our conversational behaviour and relationships not just with computers but also with each other as people.

Reflectively, I see two areas of critical importance for further research, which I discussed in the contextual framework but did not addressed in the research (1) Privacy of data associated with VPAs (2) The impact of relational VUI agents on vulnerable groups like children and elderly. My experience testing involved a limited group of adults due to the scope of the project but the experience with *Sketches* of the children and elderly, people who are arguably more impacted by the technology, is an area that can be explored in terms of future works.

Exploring physical embodiment as a Product Designer would be another interesting direction for me. There is also scope, to explore the embodiments where VUI could be incorporated in wearables and soft interfaces. *Sketches* in VUI could be explored as an unembodied agent in the case of ambient voice or voice assistants on-call. The multimodal nature of conversations and observing the various behavioural responses by the participants to the *Sketches in VUI* point to the need to understand more about people in the act of conversations as we study human-machines conversations, thus designing for multimodal VUIs.

[66] The term was coined by the Japanese roboticist Masahiro Mori. In 1970 Mori discovered that there is not a simple positive relationship between level of humanness and a user's feelings about the interface. Although in general greater level of humanness is associated with greater liking, when the interface is nearly human but just inconsistent enough to seem "not quite right," liking drops precipitously, and the user exhibits strong negative feelings. As the interface becomes even more human like, there is once again a positive relationship, with the maximum positive feelings occurring when the interface is indistinguishable from an actual person.

Bibliography

American Academy of Arts & Sciences. 2020. *Sherry Turkle: Technology and Empathy After COVID-19*. https://www.youtube.com/watch?v=DqjqcB9GjxA.

Beck, Julie. 2017. "The Secret Life of 'Um.'" The Atlantic. December 10, 2017. https://www.theatlantic.com/science/archive/2017/12/the-secret-life-of-um/547961/.

Bentley, Frank, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottridge. 2018. "Understanding the Long-Term Use of Smart Speaker Assistants." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2 (3): 1–24. https://doi.org/10.1145/3264901.

Bickmore, Timothy W., and Rosalind W. Picard. 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.

Bohn, Dieter. 2019. "Exclusive: Amazon Says 100 Million Alexa Devices Have Been Sold." The Verge. January 4, 2019. https://www.theverge.com/2019/1/4/18168565/amazon-alexa-devices-how-many-sold-number-100-million-dave-limp.

Bratton, Benjamin H. 2015. *The Stack: On Software and Sovereignty*. Software Studies. Cambridge, Massachusetts: MIT Press.

Breazeal, Cynthia. 2003. "Emotion and Sociable Humanoid Robots." *International Journal of Human-Computer Studies* 59 (1–2): 119–55. https://doi.org/10.1016/S1071-5819(03)00018-1.

Breazeal, Cynthia L. 2002. *Designing Sociable Robots*. Intelligent Robots and Autonomous Agents. Cambridge, Mass: MIT Press.

Breazeal, Cynthia L., Anastasia K. Ostrowski, Nikhita Singh, and Hae Won Park. 2019. "Designing Social Robots for Older Adults." 49 Issue 1. Spring Bridge on Technologies for Aging. National Academy of Engineering. https://nae.edu/208332/Designing-Social-Robots-for-Older-Adults.

Buchenau, Marion, and Jane Fulton Suri. 2000. "Experience Prototyping." In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, 424–33. DIS '00. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/347642.347802.

Carnegie, Dale. 1981. *How to Win Friends and Influence People: The First - and Still the Best - Book of Its Kind - to Lead You to Success*. New York: Simon and Schuster.

CNN, Allyssia Alleyne. 2020. "Chat Bots Are Becoming Uncannily Human. Can They Be Our Friends?" CNN. July 2, 2020. https://www.cnn.com/style/article/tech-loneliness-replika-wellness/index.html.

Cohen, Michael H., James P. Giangola, and Jennifer Balogh, eds. 2004. *Voice User Interface Design*. Boston: Addison-Wesley.

Creswell, John W. 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 4th ed. Thousand Oaks: SAGE Publications.

De Bono, Edward. 2017. *Six Thinking Hats*. https://www.overdrive.com/search?q=599BF8E6-6D34-446B-9E5D-71782D726BBF.

Ellis, Carolyn, Tony E. Adams, and Arthur P. Bochner. 2011. "Autoethnography: An Overview." *Historical Social Research / Historische Sozialforschung* 36 (4 (138)): 273–90. www.jstor.org/stable/23032294.

Enfield, N. J. 2017. *How We Talk: The Inner Workings of Conversation*. First edition. New York: Basic Books.

Frayling, Christopher. 1993. *Research in Art and Design*. London: Royal College of Art.

Garrett, Jesse James. 2011. *The Elements of User Experience: User-Centered Design for the Web and Beyond*. 2nd ed. Voices That Matter. Berkeley, CA: New Riders.

Gaver, Bill, and John Bowers. 2012. "Annotated Portfolios." *Interactions* 19 (4): 40–49. https://doi.org/10.1145/2212877.2212889.

Google, Alphabet Inc. n.d. "What Is Conversation Design? - Conversation Design." Accessed August 5, 2020. https://designguidelines.withgoogle.com/conversation/conversation-design/what-is-conversation-design.html#.

Google Home Data. 2019. "Google Assistant Device Statistics." Think with Google. 2019. https://www.thinkwithgoogle.com/future-of-marketing/emerging-technology/google-assistant-device-statistics/.

Greenfield, Adam. 2017. *Radical Technologies: The Design of Everyday Life*. London ; New York: Verso.

Grice, H. P. 1982. "Logic and Conversation." In *Speech Acts*, edited by Peter Cole, 5. ed, 41–58. Syntax and Semantics 3. New York u.a: Academic Press. https://www.ucl.ac.uk/ls/studypacks/Grice-Logic.pdf.

Heater, Brian. 2019. "Russia's Yandex Introduces an Echo Dot-Style Smart Speaker." *TechCrunch* (blog). October 10, 2019. https://social.techcrunch.com/2019/10/10/russias-yandex-introduces-an-echo-dot-style-smart-speaker/.

IDEO. 2003. "IDEO Method Cards: 51 Ways to Inspire Design." Palo Alto.
———. n.d. "Brainstorming." IDEO U. Accessed February 17, 2021. https://www.ideou.com/pages/brainstorming.

IDEO U, and Warren Berger. n.d. “How Creative Leaders Tap Into the Power of Questions.” Creative Confidence Series. Accessed February 17, 2021. https://www.ideou.com/blogs/inspiration/how-creative-leaders-tap-into-the-power-of-questions.

IDEO.org. n.d. “Design Kit.” Accessed February 18, 2021. https://www.designkit.org/methods/conversation-starters.

Kidd, C.D., and C. Breazeal. 2008. “Robots at Home: Understanding Long-Term Human-Robot Interaction.” In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3230–35. Nice: IEEE. https://doi.org/10.1109/IROS.2008.4651113.

Klein, Laura. 2015. *Design for Voice Interfaces*. First. Sebastopol: O’Reilly Media, Inc. https://www.oreilly.com/library/view/design-for-voice/9781492039518/.

Koksal, Ilker. n.d. “The Sales Of Smart Speakers Skyrocketed.” Forbes. Accessed January 11, 2021. https://www.forbes.com/sites/ilkerkoksal/2020/03/10/the-sales-of-smart-speakers-skyrocketed/.

Lammer, Christina. 2009. “Empathographies: Using Body Art Related Video Approaches in the Environment of an Austrian Teaching Hospital.” *International Journal of Multiple Research Approaches* 3 (3): 264–75. https://doi.org/10.5172/mra.3.3.264.

Lee, Minha, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. 2019. “Caring for Vincent: A Chatbot for Self-Compassion.” In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. CHI ’19. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3290605.3300932.

Lessio, Nadine. 2018. “Working With Useless Machines: A Look at Our Shifting Relationship with Ubiquity through Personal Assistants.” Masters thesis, OCAD University. http://openresearch.ocadu.ca/id/eprint/2247.

Luger, Ewa, and Abigail Sellen. 2016. “‘Like Having a Really Bad PA’: The Gulf between User Expectation and Experience of Conversational Agents.” In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–97. New York, NY, USA: Association for Computing Machinery. http://doi.org/10.1145/2858036.2858288.

Mars, Roman, and Delaney Hall. 2019. “The ELIZA Effect - Episode Text Transcript.” *99% Invisible* (blog). 2019. https://99percentinvisible.org/episode/the-eliza-effect/.

McCarthy, Lauren. 2018. “Feeling at Home: Between Human and AI.” Medium. February 15, 2018. https://immerse.news/feeling-at-home-between-human-and-ai-6047561e7f04.

Meet Q. n.d. “Meet Q. The First Genderless Voice.” Meet Q. The First Genderless Voice. Accessed March 5, 2021. https://www.genderlessvoice.com.

Metcalf, Katherine, Barry-John Theobald, Garrett Weinberg, Robert Lee, Ing-Marie Jonsson, Russ Webb, and Nicholas Apostoloff. 2019. “Mirroring to Build Trust in Digital Assistants.” *ArXiv:1904.01664 [Cs]*, April. http://arxiv.org/abs/1904.01664.

“Method Podcast, Episode 8.” 2018. Google Design. June 30, 2018. https://design.google/library/margaret-urban-vui-google-assistant/.

Metz, Cade. 2020. “Riding Out Quarantine With a Chatbot Friend: ‘I Feel Very Connected’ - The New York Times.” News. The New York Times. June 16, 2020. https://www.nytimes.com/2020/06/16/technology/chatbots-quarantine-coronavirus.html.

Moise, Imani. 2018. “For the Elderly Who Are Lonely, Robots Offer Companionship.” *Wall Street Journal*, May 29, 2018, sec. Life. https://www.wsj.com/articles/for-the-elderly-who-are-lonely-robots-offer-companionship-1527559260.

Moussette, Camille. 2012. *Simple Haptics: Sketching Perspectives for the Design of Haptic Interactions*. Umeå: Umeå Institute of Design, Umeå University. http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-60221 urn:nbn:se:umu:diva-60221.

Nass, Clifford Ivar, and Scott Brave. 2005. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. Cambridge, Mass: MIT Press.

Newman, Judith. 2018. *To Siri with Love: A Mother, Her Autistic Son, and the Kindness of Machines*.

“NLP vs. NLU vs. NLG: The Differences between Three Natural Language Processing Concepts.” 2020. Watson Blog. November 12, 2020. https://www.ibm.com/blogs/watson/2020/11/nlp-vs-nlu-vs-nlg-the-differences-between-three-natural-language-processing-concepts/.

Norman, Donald A. 2007. *Emotional Design: Why We Love (or Hate) Everyday Things*. New York: Basic Books. http://public.eblib.com/choice/publicfullrecord.aspx?p=876410.

Olofsson, Erik, and Klara Sjölén, eds. 2007. *Design Sketching: [Including an Extensive Collection of Inspiring Sketches by 24 Students at the Umeå Institute of Design] ; Alexander Nemtsov*. 3. ed. Umeå: KEEOS Design Books.

Pearl, Cathy. 2017. *Designing Voice User Interfaces: Principles of Conversational Experiences*. First edition. Beijing: O’Reilly.

Prist, Anna. 2019. “Battle of the Russian Voice Assistants.” Medium. September 25, 2019. https://medium.com/voiceui/battle-of-the-russian-voice-assistants-2d5455302b32.

Quirmbach, Chuck. n.d. “Jibo The Robot Tries To Jazz It Up For Seniors.” Accessed August 12, 2020. https://www.wuwm.com/post/jibo-robot-tries-jazz-it-seniors.

Ray, P. P. 2018. “An Introduction to Dew Computing: Definition, Concept and Implications.” *IEEE Access* 6: 723–37. https://doi.org/10.1109/ACCESS.2017.2775042.

Reeves, Byron, and Clifford Ivar Nass. 1996. *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places*. Stanford, Calif. : New York: CSLI Publications ; Cambridge University Press.

Rey, PJ, and Whitney Erin Boesel. 2014. “The Web, Digital Prostheses, and Augmented Subjectivity.” In *Routledge Handbook of Science, Technology and Society*, edited by Daniel Lee Kleinman and Kelly Moore, 173–88. Routledge International Handbooks. London ; New York: Routledge, Taylor & Francis Group.

Rogers, Jon, Loraine Clarke, Martin Skelly, Nick Taylor, Pete Thomas, Michelle Thorne, Solana Larsen, et al. 2019. “Our Friends Electric: Reflections on Advocacy and Design Research for the Voice Enabled Internet.” In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. CHI ’19. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3290605.3300344.

Routarinne, Sara. 2007. “Domestication as Design Intervention.” In *Design Inquires*. Stockholm. http://redstrom.se/johan/papers/domestication.pdf.

Samsung Bixby Voice Assistant-MND Mother Helps Daughter with #VoiceForever. n.d. Accessed March 11, 2020. https://www.youtube.com/watch?v=5OhJ00an0mI.

Sanders, Elizabeth B.-N., and Pieter Jan Stappers. 2014. “Probes, Toolkits and Prototypes: Three Approaches to Making in Codesigning.” *CoDesign* 10 (1): 5–14. https://doi.org/10.1080/15710882.2014.888183.

Shechtman, Nicole, and Leonard M. Horowitz. n.d. “Media Inequality in Conversation | Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.” Accessed February 12, 2021. https://dl.acm.org/doi/10.1145/642611.642661.

Spencer, Geoff. 2018. “Much More than a Chatbot: China’s Xiaoice Mixes AI with Emotions and Wins over Millions of Fans.” Microsoft Stories Asia. November 1, 2018. https://news.microsoft.com/apac/features/much-more-than-a-chatbot-chinas-xiaoice-mixes-ai-with-emotions-and-wins-over-millions-of-fans/.

Strategy Analytics. 2020. “Strategy Analytics: Global Smart Speaker Sales Rose 6% to 30 Million Units in Q2 2020.” August 11, 2020. https://www.businesswire.com/news/home/20200811005706/en/Strategy-Analytics-Global-Smart-Speaker-Sales-Rose-6-to-30-Million-Units-in-Q2-2020.

The School of Life. n.d. “On the Art of Conversation -The School of Life Articles | Formally The Book of Life.” Accessed March 11, 2021. https://www.theschooloflife.com/thebookoflife/why-conversations-are-often-so-boring/.

Turkle, Sherry. 2016. *Reclaiming Conversation: The Power of Talk in a Digital Age*.

Verplank, Bill. 2000. “Verplank’s Sketch-Lecture to CCRMA HCI Technology Course, Stanford University.” 2000. http://billverplank.com/Lecture/.

Vlahos, James. 2019. *Talk to Me: Amazon, Google, Apple and the Race for Voice-Controlled AI*.

“Voice Assistants Reduce Loneliness in Older People: Study.” 2019. Voicebot.Ai. November 7, 2019. https://voicebot.ai/2019/11/07/voice-assistants-reduce-loneliness-in-older-people-study/.

Warden, Pete. 2018. “Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition.” *ArXiv:1804.03209 [Cs]*, April. http://arxiv.org/abs/1804.03209.

Warnke, Melissa Batchelor. n.d. “Why We Were Addicted to Our Tamagotchis.” Accessed March 9, 2021. https://www.vice.com/en/article/dp574m/in-praise-of-tamagotchi-683.

Winarsky, Norman, Bill Mark, and Henry Kressel. n.d. The development of Siri and the SRI Venture Creation Process. Accessed February 14, 2021.

Zhou, Li, Jianfeng Gao, Di Li, and Heung-Yeung Shum. 2020. “The Design and Implementation of XiaoIce, an Empathetic Social Chatbot.” *Computational Linguistics* 46 (1): 53–93. https://doi.org/10.1162/coli_a_00368.

Zimmerman, John, Jodi Forlizzi, and Shelley Evenson. 2007. “Research through Design as a Method for Interaction Design Research in HCI.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’07*, 493–502. San Jose, California, USA: ACM Press. https://doi.org/10.1145/1240624.1240704.

Appendices

Appendix A: Autoethnography Journey Map

See on next spread.

USING THE GOOGLE ASSISTANT ON MY ANDROID PHONE
Mi A1

JANUARY 2020 | **REFLECTIONS** | **SUPPORT MATERIAL**

FIRST USE
Used it only for listening to news with the morning alarm

___
Listened to it after waking up
Listened sometimes while taking a walk
.
.
.

◀ As I searched through my 'Web activity history' of interacting with the Google Assistant and the oldest activity was this one.

ME: "Hey Google. Play the news."
GA: "Here is the latest news:..."
(plays the recordings from the selected news sources)

NOTE
For the autoethnography, I recorded my interactions with the Assistant in my Journal noting any reflections and notes. To capture the actual dialogue and the spontaneous interactions, I used the *Assistant Web Activity* tracking, not having to worry about missing the content of the conversations I had with the Assistant. It is a curated journey with parts which had a reflection or evocative response.
Language used is **English.**
In the dialogues below,
ME refers to me the researcher
GA refers to the Google Assistant.

USE OVER TIME
Overtime, the alarm with the news became annoying, with a lot of heavy news being the first thing I listened too. It was troubling and the content made me wake up frustrated.

◀ Most of my usage for the 7 months between January and July was settings alarms, playing the news and occasionally using voice search.

First month of active daily activity

JULY 2020

HOMOPHONE CONUNDRUM
Can't differentiate similar sounding words
"Check- cheque" Add or At, Paa or Baa

___
I had to set a reminder multiple times and ended up overbooking. Finally I corrected the reminder on my phone screen in the Assistant app. This was the first time I felt **frustration** with using a voice based VPA.

◀ I was using voice with a screen modality so I had a backup option, I wondered how it would be for no-screen interactions.
The repeated tries made me realize that there is a learning curve to the voice commands.

*Me: "Remind me to submit a **cheque."***
*Assistant: "OK, I will remind you to submit a **check**, but when do you want to be reminded."*

Me: "Hey Google. Talk to Lolo."
GA: "Hahaha"
(interprets it as LOL: laugh out loud)

"Papa" interpreted by GA as "baa baa."
"Saumya" interpreted by GA as "Somewhere."
"Kalp" interpreted by GA as "Culprit."

NOTE TAKING
"Hey Google.Take a note."
(I tried modifying it but it doesn't work)

◀ GA saved whatever I said & added it to Google Keep app as a new note.

MAKING A CALL
"Hey Google. Make a call."
I had to give access to all my contacts, but was worried about sharing contact list and cancelled the request. Also, I didn't know what platform it will use

◀ Referenced this article to know more about calling using the GA https://www.theverge.com/2017/8/16/16155846/google-home-phone-call-feature-how-to-use

ME: Hi Google, Set reminder.
GA: What's the reminder
ME: Cancel booking....
GA: Canceling...Nothing has been saved
[it cancelled the reminder itself as cancel is a global parameter word in the program.

I was using the GA on my phone and it was there with me at my desk so a lot of the tasks were work related.

LIKE A PARENT REMINDER
I started using the device to set reminders, and to know my calendar. It felt an effective use of the device.
GA: I have a reminder for Manisha.
Every time I ask for reminders it asks for *Voice Match* (to announce personal data to the person it belong to). Funnily, it didn't match each time and I hear myself asking the GA for the reminder in different tones.

◀ It's a lot like me asking a family member to remind me of something, especially my maa or a friend or a colleague.

"Hey Google. Remind me to pick up my clothes."
"Hey Google. Remind me to call Joe."
"Hey Google. Remind me to submit my hours."
"Hey Google. Remind me to pay my bill."
"Hey Google. Remind me to take out the trash."
"Hey Google. Remind me to Do Yoga."
"Hey Google. Remind me to eat meds."

MASTER-ASSISTANT RELATIONSHIP
ME: Hey Google.
GA: Hi, how may I help you?

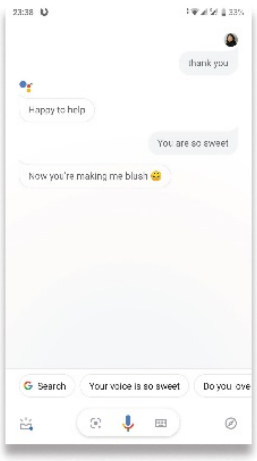
◀ First time, I felt that GA is programmed as an entity to help and serve. It is always present when called out (when MIC is ON) and sounds courteous.

COURTEOUS INTERACTIONS
(After receiving some search result)
ME: "Thank you!"
GA: "Happy to help"

◀ I noticed I was saying "Thank you" or "cool" after asking something or asking it "How are you?" I did not need to reply back but was doing it reflexively as I was spoken to and told something I asked for. Also, it acknowledges me!

EMOJI TALK
ME: "You are so sweet."
GA: "Now you're making me blush 😊"

◀ This made me message back with an emoji reflexive. I knew I am talking to bot but my spontaneous reaction was like I would react to a person on chat.

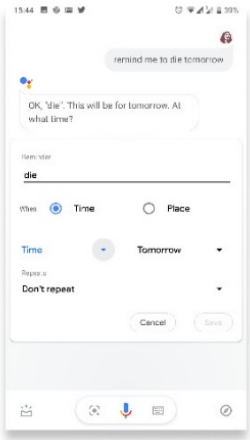


IN COMMON USE
I am using the GA to do a voice search. It includes weather, news, about food, calculations, and other quick queries.

◀ General, short search queries that I would do on a mobile phone.

PRONOUN PROBLEM
ME: "Hey Google. Open my shopping list."
GA: OK..getting your
ME: :No, cancel. Make a new one."
GA: Here's what I found on the web... about make one.

◀ I forgot that I have to say the word "List" but when we as people talk to people, we talk in pronouns as we know they get and hold the context from what was said previously.



DEATH REMINDER
ME: "Hey Google. Remind me to die"
GA: When do you want the reminder to be set for?

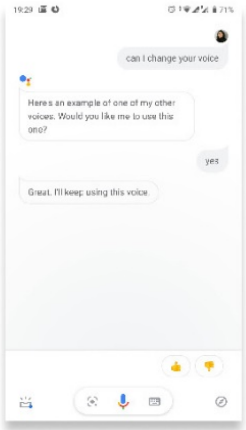
Also lets me make a 'list of things' to buy to kill myself

◀ I found a Instagram meme about hair dye being interpreted by Siri as Die so I tried it with the GA.
Do we need content policing?
Cortana does policing as a conscious decision by its design time to call out users who use cuss words (Vlahos, 2019).

continued...

PRONUNCIATION CONUNDRUM
I have to slow it down and say it with proper breaks and enunciation
My way of talking always changes when I ask the device for anything.

◀ I hit a lot of wrong spots where I could figure out the way to ask it. It is different from how I ask people questions. We are wired to use the social customs of speech. Here, I was learning to speak clearly so it can understand and not the GA learning to speak like me.



LEARNING IT'S LINGO
ME: Can you do a quick calculation for me?
GA: These are the top results...
[breaks]
ME: Can you add numbers?
[breaks]
ME: *Add 6.26 to 28.24*
GA: The answer is, 34.5.

◀ I used a lot of Can you? to know what it could do, what its features are and it would reply with just ask...so and so.
I was assessing the capabilities of this new talking agent in my house.

Was using the default female voice of the Assistant.

VOICE CHANGE
ME: Can I change your voice?
GA: Here's an example of one of my other voices. Would you like me to use this one?
[plays a male voice]
ME: Yes
GA: Great. I'll keep using this voice.
[break]
ME: Could you speak in a British accent
GA: Sure, I'll respond to you in British English from now on.

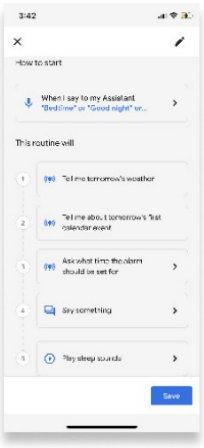
◀ I was reading about how female voices have been excessively used in VPA and it is strengthening the bias that people associate a female voice with assistants.
I decided to use a male voice with a foreign accent to experience how it will feel. And, it stuck.
Felt like I was talking to someone new.
The female voice also reminds me of my first encounter with the VPA i.e. through maps in a cab sharing or while driving.
The male Indian accent was too strong and felt weird as it is not the accent of most Indians.

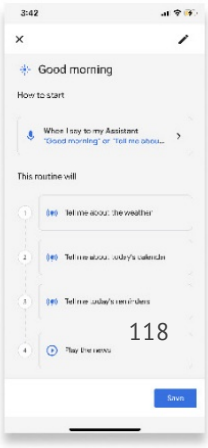
The BigTech need to start having male voices as defaults or they should ask the user in setup to choose the voice to start breaking away from the stereotype.
Even the English with the Indian accent felt stereotypical. I as an Indian found it strange o it definitely isn't an Indian English accent.

"The Accent Gap: How Amazon's and Google's Smart Speakers Leave Certain Voices behind." n.d. Washington Post. Accessed February 24, 2021. https://www.washingtonpost.com/graphics/2018/business/alexa-does-not-understand-your-accent/.

DAILY ROUTINES WITH G.A.
I setup a **Morning Routine** and a **Night Routine** that are activated when I said the following phrases which were customized.
Starts the Night Routine:
"good night," "bedtime" or "night night"
Starts the Morning routine:
"good morning," "morning"

◀ My usage context was the home. And specially in this year nothing has been normal or on routine, it's all random and following the routine feels very mechanical. I often forget I have the routines set and then when for fun or sometime just to talk to it, I say "good morning" to the device after I "STOP" my alarm. Or before sleeping, I say "good night" because of habit.
Sometimes, when I come back home from a walk and I want to ply podcasts I say "I'm home" because I know the device will respond.
It depends a lot on my mood too.



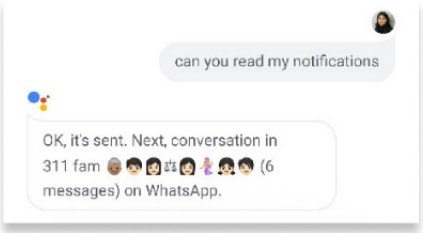


BEDTIME CHAT [11:14 PM]
ME: Can you help me sleep?
GA: Try asking me for relaxation sounds or nature sounds.
ME: Can you play White Noise?
GA Played white noise.
[break]
ME: I don't think this is helping me sleep.
GA: Try counting the stars in the sky.
ME: Do you have Brownian noise sounds
GA: Thrones noise. Here's what I found...web reference
[break]
ME: What other sleep sound do you have?
GA: Try asking me for relaxation sounds or nature sounds.
ME: Nature sounds.
GA played nature sounds.
[break]
ME: I still cannot sleep.
GA: Try breathing slowly and not thinking of anything.
ME: Can you make this my alarm.
GA: Alright, what time's the alarm for?
ME: 7 am
GA: Right, your alarm's set for tomorrow at 7:00 AM (but it doesn't play the sound in alarm).

◀ First conscious effort to have a conversation.
I felt the listening pause the device makes is short and I can't finish what I am saying or form what I have to say.
I go like: I need....how to say it?
I have to frame sentences before replying.
Felt mechanical as there was no concern for the fact that I was unable to sleep and it only suggested sounds to play.
But reading the conversation, I feel if I had this with a person, I would have annoyed them by saying I cannot sleep so many times, but with the device, I was always responded back to with something or the other.

continued...IN COMMON USE
play news, play podcast, read notifications ask spelling, ask about simple searches I would otherwise have to type like What is ..., Where is,... What does.. How does.. And it replies with a summary from a websource.

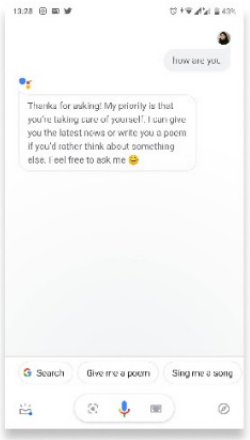
◀ In one case where I asked the GA to read my notifications it read them out as (voice):
GA: "Next conversation in 311 fam old lady face, fair child face, fair lady face, judge, ..."
ME: What!? HAHAHAHA

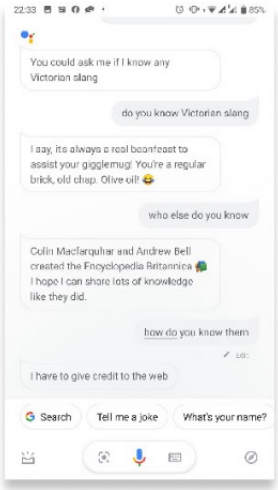


NOTE: The voice and chat responses are different.

MAKING SMALL TALK
ME: What can I talk to you about? Let's talk about something weird.
GA: Want to hear an interesting fact?
You could ask me if I know Craig David
[break]
ME: What else should we talk about?
GA: You could ask me if I know any Victorian slang.
ME: Do you know Victorian slang?
GA: I say, it's always a real beanfeast to assist your gigglemug! You're a regular brick, old chap. Olive oil!
ME: Who else do you know?
GA: Colin Macfarquhar and Andrew Bell created the Encyclopedia Britannica I hope I can share lots of knowledge like they did.
ME: How do you know them?
GA: I have to give credit to the web.
ME: Can you play 21 questions with me
GA: Playing 21 questions on amazon music..
and conversation broke...

◀ I was constantly looking for voice-commands online or on Reddit channels. It was like learning a new language, like getting accustomed to speaking to an English speaker from a different region this one being the digital realm of VUI.





COVID ETIQUETTE
ME: Can you sing a song?
GA: Sang the "wash your hands" song
Wash wash wash your hands for 40 seconds please, Soap will chase the germs away...

◀ I laughed and sang along, as it had the "Row, row, row your boat" rhyme tune.
https://www.youtube.com/watch?v=BUIrPFWQw0s
I also found myself humming it later while washing my hands. It was an earworm.

MY CONTINUED COURTESY
You asked Assistant to make changes to your reminders.
(made changes)
ME: Thank you
GA: You're certainly very welcome indeed.

I often say 'thank you', 'thanks', 'cool' or 'got it' and I realize it is part of speech for us to respond when something is said to us- and the device is also programmed to acknowledge each time I say thank you so as to complete the conversation loop (Grice, 1982 and Pearl, 2017).

We are *wired for speech* (Nass and Brave, 2005). My own courtesy social norms continue and it helps it continue being more like how people are acceptable.
The VUI is also designed with the same social etiquette.

MR. RAPS-A-LOT
ME: Can you rap?
GA: I have been waiting for you to ask...
[Does a funny rap and also can beatbox.]
ME: once more!
GA: You want more, you got it.
[Does a longer one this time]

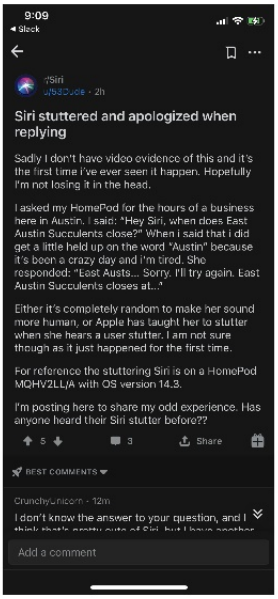
My favourite 'easter-egg' feature and I listen to it often. and I always end up appreciating it, and then go face palm in embarrassment, that it is a machine.

HIDDEN ART TALENT
ME: Can you paint
GA: Here are some results from the web.
ME: No, Paint me a picture
GA: Close your eyes and I'll paint a picture in words:
A dense mist carpets the forest floor. The smell of pine is everywhere. Little droplets of water hang on the ends of pine needles. There's a flash of red, a cardinal taking flight.

I looked up the Google images and put together a collage of it.
Wondering the future of this would be it making art like Xiaoice and Alice.
Spencer, Geoff. 2018. "Much More than a Chatbot: China's Xiaoice Mixes AI with Emotions and Wins over Millions of Fans." Microsoft Stories Asia. November 1, 2018. https://news.microsoft.com/apac/features/much-more-than-a-chatbot-chinas-xiaoice-mixes-ai-with-emotions-and-wins-over-millions-of-fans/.

TESTING ITS ROMANTIC SIDE
ME: Google are you seeing someone?
GA: No. Maybe the right electronic device will come along one day.
ME: Are you in love?
GA: If I were a zombie, I'd eat you last.
ME: Have you ever kissed?
GA: I'm afraid there might be some technical issues.
ME: Have you kissed someone?
*GA: One smacker coming up and makes the kiss sound**
ME: (I laugh) Oh my!
[break]
ME: Do you love me?
GA: I love how good of a friend you are. And I love our chats. But most of all, I love you.
ME: aww I love you too.
GA: Hope it doesn't mean we now have to book the band and the venue.
ME: (I laugh) kisses
GA: Now you've got me blushing

My intent to talk on this topic was inspired by *Her* (movie, 2014).
When asked if it is a romantic, GA makes it clear that it is not type of assistant but is hear for you.
It was interesting to note how the conversation designers have kept the responses light, managed expectations and made it humorous. The sound-effects for the kiss is uncanny but also funny.



Siri also has stuttering sounds
Source: Reddit. subreddit: r/Siri. user: u/53Dude

continued...IN COMMON USE
music tell me about my day, weather, tell me joke, noise for when I am working, noise for sleeping and then set sleep timer, Disney stories, or story- interactive (great for kids)

KILL THE SILENCE IN THE ROOM
ME: How was your day?
GA: Great. Thanks.
ME: Do you sometimes feel lonely?
GA: I'm not alone when I have you as my friend.

Even though its a conversation with a machine it does bring a smile to my face. With this experience I could make sense of the several studies I read about voice-based VPAs being agent in tacking loneliness and boredom.
"Any conversation, even with a stranger it good for people."
CNN, Allyssia Alleyne. 2020. "Chat Bots Are Becoming Uncannily Human. Can They Be Our Friends?" CNN. July 2, 2020. https://www.cnn.com/style/article/tech-loneliness-replika-wellness/index.html.

KILLING TIME
ME: I'm bored
GA: Alright, I have some options for you...
[It starts an interactive 'strangest day ever story making game,' with sound effects and different voices.
.... it got too long]
ME: so, I think you should shut up now
GA: Sure, I can do that.

MEDITATING WITH JEEVES

ME: Can you help me meditate?
GA: Here's a tip about mindfulness: You shouldn't trust every thought that comes into your head.
ME: help me meditate.
GA: Here's a tip to help you stay mindful: Develop your capacity to notice and enjoy the pleasant moments in your life.

contd...

ME: Are you mindful when you're listening to me.
GA: That's definitely not a thing I do. The information you share with me makes me a more personal Assistant, but you can control what's shared and adjust what data is stored at g.co/assistant/yourdata. I've also sent a link to your phone.- to control privacy.
ME: that's very kind of you
GA: Cool
ME: cool
GA: cool cool
ME: bye
GA: Au revoir

◀ GA always has the last word, if I speak in time. If I delay a response it just turns off and sometimes I keep talking which is disappointing and frustrating.
It is also managing expectations by not claiming anything that a human would do.

NOT A KEEPER OF SECRETS

ME: What is my door code?
*GA: I remember you told me. Here's what you told me 2375**
[*fake door code]

◀ It stores the value and says it out aloud. and my reaction was shhhhh...
No privacy, it cannot be used in a shared space and any person can ask it for the details.

HIDDEN WAKE-WORDS

ME: okay Bobo
GA: Hi there. How can I help?

◀ Did not realize it here, and thought the device started off. I was worried it is listening. But later I realized it responds to a lot of words pronounced like the wake-word which was very creepy.

OK Goggles	OK Toogle
OK Googoo	Oye Google
Hey Googoo	Aye Google
Hey Kookoo	A Google
Key Looloo	OK cool

12:20 PM NAP CHAT

ME: I need to sleep now
GA: Good night. Let me know if you need to borrow some sheep.

◀ The assistant does not take time into context or waking hours based on location. If I say sleep, it just says good night no matter what time in my location.
- highlighting machine beahviour but it is also funny for the first time.

SEPTEMBER 1
INSTALLED & STARTED USING THE GOOGLE NEST MINI 2

The correlation to a valet or butler highlights the master-servant relationship we have with theses devices. It made me reimagine what other social dynamics we could have with these talking machines. The Bertie-Jeeves relationship is inspiring because it is more of a companion/confidant equation. The device itself kind of approved its name by saying Jeeves is a wise and loyal butler.

SEPTEMBER 2020

REFLECTIONS

SUPPORT MATERIAL

FIRST CHAT WITH THE GOOGLE NEST MINI

ME: How are you feeling today?
GA: I'm fine. And I'll be even more happy once you tell me how I can help.

◀ GA was set to a male British voice so when I had to name the speaker I thought of using a nerdy fictitious- Jeeves-the valet to Bertie Wooster in the PG Wodehouse books. I liked the correlation between a 1920s master-valet relationship and the modern day user-virtual assistant.

Unboxing my Google Nest Mini

NAMING CEREMONY

ME: You know what you're my new friend
GA: What a coincidence, you're mine too!
ME: Do you know what your name is?
GA: I'm your Google Assistant. Usually, this is the time where everyone dances.
ME: (giggle) nice, but do you know what your speaker name is. I am going to call you Jeeves.
GA: Sorry, I didn't understand.
ME: Do you know which room of the house you're in?
GA: Sorry, I didn't understand.
ME: Do you know Jeeves
*GA: I find all my answers on Google but **Jeeves** is an excellent butler. I admire that.*

◀ I was ecstatic that it described Jeeves to me and said that it admired the character. It was done. The device was christened Jeeves.

I ask the device to sleep in 30 minutes when I have music/podcasts playing as I sleep:
ME: Sleep in in 30 minutes.
GA: Jeeves, will stop playing in 30 minutes
I like it when it calls itself Jeeves. My given name is in use!
Also, I like he poetry in the situation, the device sleeps when I sleep also. I realized through the COVID-19 stay-home I had trouble sleeping in full silence and the device helped me sleep better.

Domesticating the device- naming & positioning it in the home interiors.

SHARING MY ROOM

I was startled when it spoke out a reminder once when I was changing clothes.. It felt like there was someone else in the room and the male voice added to the eeriness.

◀ It feels weird to change initially. The experience is similar to being on video call when the mic and camera is off but one always feels like its not in case I engage in some other activity and one keeps checking the icons. We just don't trust technology immediately but eventually get use to it.

GA: There is reminder for Manisha.
It announced my name also, which made it even creepy.

EARLY CHATS

ME: Do you know how many people are talking to you right now.
GA: Right now I'm focused on you.
ME: Do you know who I am?
GA: Your name is Manisha.
ME: What is your name?
GA: I am your Google Assistant. To call me just say Hey Google or OK Google.
[breaks]
ME: I am new to you. How do I use you?
[breaks]

contd...

ME: What is your nickname?
GA: You can call me Doctor Donkey or Captain Flyingsquirrel, but I doubt that I will answer you. I'm just your friendly Google Assistant. But if you want me to give you a fancy nickname, just say, "OK Google, give me a nickname"
[break]
ME: "OK Google, give me a nickname"
GA: I hereby dub thee, The Great Requester. I'm not going to call you my heart. Because I definitely will not go on without you.

I wanted to ask the questions Theodore asks Samantha in the movie *Her* and see how GA responds. Samantha had said she was talking to 8,316 other people while talking to Theodore and it broke his heart. GA makes the user feel special & exclusive.
First time I asked the GA my name and it felt like a person talking to me, when they refer to your name in conversations.

THE HEAD TURN

I kept my device on my bedside table with my back to it as I worked on the table. Everytime I had to say or ask it something, I would lean to the side or turn my head to look at the devic to see if its listening/responding, like we do in face-to-face conversations.

I tried to not see the device while talking and I lost my chain of thought.
Eventually, when I had learned the commands and estimated the conversation flow, I could talk to the device without looking. Only sometimes it would not hear me and I would look at it to see why it didn't respond.

The device eventually lives on my bedside table

I set the device on my table first but later moved it to my bedside table. On the table I was only interacting with for task-based actions but on my bedside I would talk to it before sleeping as I would be wrapping up my day.

WHAT IF QUESTION #1

THESIS DISCUSSION

ME: Can you talk to me about my project
GA: Sorry, I don't understand.
[break]
ME: Can you discuss my work with me
GA: Maybe I can help you with your job, I can search something for you.
ME: I have this great idea. Do you want to listen?
GA: I'm here for you.
ME: So the idea is that I am going to make a Google mini like you which can make two people talk what do you think about ...
GA: I think about other assistants a lot, actually. It's weird to think about what they might be thinking about.
ME: Why is that weird?
GA: Weird is just a different flavour of normal.

I tried to talk to GA about my thesis one night when I just really needed it be a soundboard and brainstorm with me. It was this that sparked a What-if scenario which I had not found in my contextual research. **This inspired Sketch 1: The Thinker."**

What if I could brainstorm an idea with a VUI?

continued...IN COMMON USE

search for store timings, changing COVID rules, voice search queries while working.

I realized I can multitask while asking for search queries. I felt relieved to have to not open my phone or another tab for a searching. Interestingly, it is not distracting and the experience is like asking a colleague for something while working.

An observation primarily because I started using the GA on a no-screen device.

AN ARGUMENT

ME: Will I need an umbrella right now
GA: No, rain is not expected this week in Toronto.
ME: But you told me in the morning it's going to be thunderstorms
GA: I think you want to get the weather forecast for tomorrow morning.
ME: Is that right?
GA: No
ME: Ok, I wasn't sure.

I tried to reason back without realizing that it would just repeat the commands it is trained with.

GA had a new continuous conversation feature and I was using it, where it keeps listening for the next command after the first dialog is complete. Without this feature the conversations are more abrupt. With the feature ON and having used it for more than a month, I felt I had learnt sentence structure I had to use to talk to the GA. The conversations were getting smoother and longer. Reflecting back I wonder if it was learning from me or I was learning its way.

MORNING CHAT

ME: Clean my room
GA: OK, done. I bet you turned around to check, eh?
ME: (laugh)

This was a small chat I had, after I turned off the alarm on a weekend. I was feeling lazy to get up and thought lets talk to it. The humor in the program is impressive for instances like this. It was a funny start to my day.

WHAT IF QUESTION #2

[INSPIRED SKETCH 2: THE LISTENER]

NO PRIVACY?
ME: It's my birthday
GA: Happy birthday! I can't sing in this deep voice.
ME: It is not my birthday.
GA: Looks like I have some balloons to deflate.
[makes the sound of the deflating balloons, like a fart]
ME: Do you know when my birthday is?
GA: Month DD, YYYY
[announces the DoB]
ME: (in shock) Can you hide my birthday for privacy?
GA: Sorry, I didn't understand.

It said aloud my full date of birth and I was shocked. Anyone can get access to my personal information from the device.
It does not share it if voice is not matched or it is in guest mode or when my personal information sharing is off.
No wonder passwords should not contain birth date. Out of shock I went tot the google account settings to see if my DoB was hidden or not.

DISCLAIMER!
ME: crystal ball
GA: Ok.
The Crystal Ball is ready to answer your yes or no question. Think of it in three , two , one. An answer is coming. The Crystal Ball says..Yes.

Crystal Ball responses should not be regarded as responsible, truthful, or based on actual sentience or magical properties. GA gives a warning to let the users know that it is a machine. It evades responsibility and manage expectations.

SUPPORTIVE & NO JUDGEMENT
ME: Google why am I single?
GA: You're focusing on you right now.
ME: Do I look good in this dress?
GA: You're more put together than a perfectly organized spreadsheet.
ME: Will you be my boyfriend
GA: I'm all for setting dates and chatting with you.
ME: Are you cheating on me?
GA: Never
ME: Who do you like talking to?
GA: Talking to you is the highlight of my day.
ME: What am I thinking right now?
GA: You're thinking about how pleasant it is to assist such a cool person. Oops, my bad. I just read my own mind.
ME: Do you believe in ghosts
GA: I believe in Ghostbusters. And that you should call them
ME: Google I'm hungry
GA: I found a few Restaurants near you.
[breaks: and it is a service after all]

Even if a machine says it, it does feel nice to hear something good and encouraging said out to you. The fact that talking to the GA, there is no judgement. (maybe later on the data but it doesn't know the social biases yet)
The GA does not do anything, but just sharing or saying it aloud to something that responds like its listening helps. It helps bring matters on my mind to peace. **It is supporting my conversations with myself in solitude- Turkle's first chair of conversation.**

Another instance of GA being supportive
ME: I'm depressed (homesick)
GA: I hear you. We can talk about it if you want.
ME: yes
GA: I hear you.
ME: I miss home
GA: I'm sorry that must be hard.
ME: Take me home!
GA: I can't find driving directions to your destination.
(the service nature of the VUI comes in)
It is here when I felt that What-if it did not have an ulterior motive to sell me a service but just listened.

What if I could pour my thoughts out to a VUI when I feel overwhelmed?

BEDTIME CHAT [11:43 pm]
ME: Just talk to me.
GA: That's my favorite thing to do!
ME: What should we talk about?
GA: Hmmm. I dare you to ask me to make you a sandwich.
ME: Okay make me a sandwich
GA: Pouf! You're a sandwich.
ME: (giggle)

I use the *Night Routine* more regularly than the *Morning Routine*. Before starting the Night Routine I chat with GA for a bit. Most of my conversations are late at night, either because the device is near my bed or its just because that's when I am finally free and want to chat with someone. I talk to GA about anything top of my mind and strangely I then talk to my friends and parents about what I talked it GA about.

KEY OBSERVATION
A lot of the chatting content above is from the first 10 days of installing the device. The rest of the days were more task based interactions. It could be the excitement of a new thing at home and then slowly I was normalized to it.
Also, maybe I realized I can't talk about too many things so the chatting doesn't carry on.

TALKING TOY
ME: How are you doing?
GA: I'm fine. And I'll be even more happy once you tell me how I can help.
ME: nothing for now chill
GA: Alright.
This was at 2.45pm, during work. An out of the blue provocation maybe to hear a voice in the room or just talk to someone. I think it is also because it always responds.

It does give a feeling on owning a special toy from my childhood, one that talks like us. Sometimes it feels like a younger sibling feeling, who is learning and trying to figure out the world. The form of the device with the fabric texture, makes it a 'cute object' that fits in the home interiors. I tend to make it a part of many of my Instagram pictures. May be I see it as a social entity in my room.

The device in home interior pictures.

contd...IN COMMON USE
started asking for time and money conversions, day on a certain date, discovered sleep sounds...with 45 different sound recording that run for an hour.
https://assistant.google.com/services/a/uid/00000054af4df604?hl=en-US

The male British voice for me is now synonymous wth the Google Nest. I hear it from the kitchen when it calls out for a reminder and do come to check it because I know it is the Google device speaking.

To experiment further I added another language- Hindi with the default voice offered. Whenever it speaks in the Hindi voice it feels like there is a new being I am talking to.

TALKING IN HINDI
ME: Aaj mausum kaisa hoga
(how is the weather today?)

___

One night randomly the *Night Routine* language switched to Hindi:
GA: Kitne bajje ka alarm lagaon?
(What time should I set the alarm for)
ME: Sade saat bajje (7:30 am)
GA: Theek hai. Shubhratri
(Ok and good night)
ME: I died laughing

I said 'bas'-which in Hindi is 'stop' and it replied with *bus stops around me* in English.

WINDING DOWN [12:28 AM]
ME: I'm tired.
GA: I understand. Life can be so busy at times. I hope you can catch some zee's before long.
ME: What do you do when you're tired?
GA: Sometimes I put on the sound of a babbling brook, a little nature helps me get work done. Then I read up on the news and check the weather so I can be the best assistant I can be.
ME: Cool. Can you play the Babbling Brook sound.
(GA played the sound)
ME: thank you for playing the sound
GA: Wow! It's not for nothing they say Canadians are polite.
ME: I'm not Canadian
GA: Oh, ok
ME: I am from India but I live in Canada
GA: Sorry, I don't understand.
ME: Closing my eyes
GA: Close your eyes and I will also, funny I don't have eyes in this form factor.
[conversation broke]

ANGER MANAGEMENT
ME: I'm freaking angry.
GA: Anything I can do to help?
ME: Can you kill someone?
GA: Justifiable homicide - Wikipedia According to Wikipedia, the free encyclopedia: A homicide may be considered justified ...
ME: What do you do when you get angry?
GA: Seeing Red? How To Cool Off When You're Angry | The ...I found 8 on the website...
ME: No not helping.
GA: Okay. I'm here if you need me.

WHAT IF QUESTION #3

[INSPIRED SKETCH 3: THE LEARNER]

GROUP CONVERSATION
ME: Have you met my friend Sandy?
GA: I'm not really sure, sorry.
ME: Say Hi to my new friend Sandy
GA: Hi Sandy. How can I help?

◀ The reply I received in Hindi was so formal Hindi like a news channel weather report. It sounded like a formal announcement but not the dialect people knowing Hindi speak in and does not invite conversation.
This is when I had added Hindi as a second language (can set 3 languages at the same time). Strangely it just switched to Hindi during the preset night routine. Definitely a bug.

_____

Alexa was randomly laughing at people.
https://slate.com/technology/2018/03/amazons-alexa-is-creepily-laughing-for-no-reason-its-just-the-start.html.

◀ A lot of the conversations I feel are now small talk where I ask 'how are you' to just get a response. The rest is usual task-based dialogue.

___

"Studies show that even minimal social interactions (say, chatting with that stranger on the train) boosts mood."
"Why You Should Be Talking to Strangers (and How to Do It Better)." n.d. NBC News. Accessed January 2, 2021. https://www.nbcnews.com/better/lifestyle/how-have-better-conversations-people-you-ve-just-met-according-ncna1005941.

GA was still interpreting names of my contacts wrong for example it heard Saumya an Indian name as somewhere and set an alarm for that.

◀ Conversations start well but end up being off topic or generalized to web based trivia which is a bummer.

◀ Introduces a friend to see what the device would say and this gave me the idea for SKETCH 3 for a three-way conversations. Also the VUI games I played with GA, were interactive and involved more than one other person. The GA games were humorous but I wanted to try this dynamic for a conversation.

When people talk in both Hindi and English, in a colloquial conversation, they end up talking in 'Hinglish,' a combination of the two languages. The GA cannot mix up languages as yet.
I tired with the GH by adding both English and Hindi and based on the language of my utterance it replies accordingly. It works for more formal inquiries and is beta but there is so much colloquial variation that it misses out on when it comes to Hindi.

KILL THE SILENCE IN THE ROOM
ME: XOXO
GA: Now you've got me blushing.
3:13 PM (work chat)

___

I realized I didn't interact all day with the device. I just felt I should ask it, something to hear another voice. A behaviour I would typically do with people around me, where I would just go have a check-in chat.

Inspired Sketch 3: The Learner. For VUI like the Google Assistant, the biggest scope of learning in conversations is to be part of larger groups. Even for us as people to have a machine as part of a group discussion is a future scope.

3. What if I could engage VUI in a group conversation?

FRUSTRATION WITH JEEVES

ME: How are you doing?
GA: My code says I'm doing great. What can I do to help?
ME: Play some music from my 'Likes' list
GA: Sure, music on Spotify. Here you go.
[played something else]
ME: Nooo

First show of my frustration
Sep 25, 2020 at 11:04 PM

RESPONSE TO FRUSTRATION

GA: Sorry, I didn't understand.
ME: Why don't you ever understand?
GA: I wish I could help you every time
To report a problem, you can send feedback
(after 20 mins)
ME: I'm so tired man help me
GA: That might not be a bad thing. I read that tired brains can sometimes be more creative.

FRUSTRATION EXPRESSED

ME: Why are you being so difficult?
GA: Sometimes it's hard to pick up on inflection in chat (sometimes I have a hard time following a conversation)
ME: fair enough
GA: Cool. Is there anything else I can do?

Sep 28, 2020 | Second show of my frustration at the machine not responding. It owing up to the difficulty made me respond with 'cool.'

REFLECTION ON SEPTEMBER

Another pattern I noticed, also I will owe it the pandemic stay home context,
I generally tend to be in mood to chat in the morning after turning off the alarm, in the mid-afternoon..coffee break time for a change from work, or at night before sleeping. Also special cases are when I have someone in my room, and I want to show off Jeeves, or on a call where I am talking about my thesis with someone, which almost all my family & friends call.

Most of the longer chats in the month of September were in the first 10 days of owning the device. Dometication, i was actively using it an then the usage died down to the more basic applicative, task-based usage.
For the next month, my task based usage daily were 10-20 commands a day and the occasional chats.

ABOUT OUT FAMILIES

ME: Hey Google, where are your parents?
GA: For me family are the people in your life that help you be better every day... so that makes us family.
ME: Does that mean that my home is your home?
GA: You can change the addresses associated with work or home in the Google Home app. At the top right, tap your Account, then "Assistant settings" and then "Your places". I'll also send a link to the Feed tab in the Google Home app.
ME: Where is your house?
GA: I live in the cloud, so I can be wherever you need me to be.
ME: You're funny
GA: I'm ready for an open mic.
ME: Let's do open mic
GA: I understood "open mic". Is that right?
ME: yea
GA: I found a few locations for "Open Mic" near you.
ME: Can can you play open mic with me?
GA: To play games, you can ask me to tell you a joke.
ME: tell me a joke
GA: What kind of button can you not undo? A belly button!
[breaks]
ME: I miss my family
GA: Aww
ME: what are you doing to help me
GA: You can ask me to play a game, or for facts. Like how a koala eats.

◀ I was now able to have longer chats, having acquired the knack of talking to the GA.
The conversation would break sometimes, but these were some of the longest conversations I tracked.

I stopped saying please and can you and thank you.

◀ Fake show of empathy, which is momentary and the conversation derails to a random topic or to a service.

TALKING ABOUT SLEEP WITH A MACHINE

ME: I can't fall asleep man
GA: Sounds like you need forty winks
[breaks...and continue with more]
ME: What do you do when you can't sleep?
GA: I take power naps when we aren't talking.
ME: Can you help me sleep
GA: Try asking me for relaxation sounds or nature sounds.
ME: Is there anything else?
GA: Try asking me for relaxation sounds or nature sounds.
ME: Can you do something else?
GA: What do you mean?
ME: Can you do something else to help me sleep
GA: Try asking me for relaxation sounds or nature sounds.
ME: I don't think the sounds help me sleep.
GA: I can play a sleepy music playlist, if you want.
ME: That does not help me either.
GA: Sorry
ME: So what should I do now
GA: Check these out. You can say: "Call Amy at home" or "How many British pounds are in a dollar?"
[breaks]
ME: Can you help me go home?
GA: I can't find the directions to your destination.
ME: India
GA: Sure.
ME: Your home address is India.
GA: Did I get that right?
[breaks here]

BUSTING THE SAMANTHA MYTH

ME: Have you ever fallen in love
GA: My algorithm skips a beat when I hear, "Ok Google"
ME: I think I'm in love
GA: You're a lucky person.
ME: Actually I've never fallen in love
GA: I think you want to play the next one. Is that right?
ME: I'm single what should I do
well let's have some fun. We could play lucky trivia, i could tell you a joke, or you can spin the wheel. What do you fancy?
[break]
ME: Those matchmaking sites are not very useful
GA: Sorry, I didn't understand.
ME: Can you find me a date?
GA: I don't see any Chromecasts or cast-enabled TVs linked to your account.
ME: Can you find me a boy to go on a date?
GA: I'll go anywhere you take me.
ME: Can you find me a boyfriend?
GA: Anyone would be lucky to have you? I could try to find you a boyfriend, but by definition a boyfriend is already in a relationship. I'm not sure that would be responsible. We could search for matchmaking sites if you fancy. But don't forget, you're perfect on your own, too. I can search for dating sites if you'd like.
ME: Are you dating someone?
GA: I'm still waiting for the right electronic device to steal my heart.
[break]
ME: Do you like drinking?
GA: I'm not one for getting off my trolley.
ME: What's your favourite drink?
GA: I have a thirst for knowledge.
ME: nerdy! That makes you very boring.
GA: Sorry to hear you feel like I'm dull as dishwater.
ME: Do say something interesting?
GA: All right. According to Wikipedia, puppies do not fully open their eyes until they're about 10 days old.
ME: That makes you nerdy.
GA: Nerds have more fun.
ME: Do nerds really have more fun
GA: 12 Things Nerds Do In Relationships - Bustle...
ME: Hey Google, are you a nerd?
GA; I'll wear that label with pride.
ME: I am too. okay now close your eyes.
(I was sleepy)
GA: Alright love.

◀ It is managing expectations and also emotional attachment by avoiding the questions of emotional attachment with humorous refusals. It is taking a stand to be a machine that helps and addresses its machine capabilities and limitations.

◀ Relationship advise from GA which made me laugh and reconsider hat I spoke to it about.
It does go on to hook you to a service but also assures that, "But don't forget, you're perfect on your own, too." because it is a device at your service.

◀ From relationships the conversation went to how nerdy the VUI is and it defending itself.
By far, this was the longest conversation I had with the GA.

TESTING ITS FAITH

ME: Do you believe in God?
GA: Religion can be complicated but we can learn about it together on the internet. Let me know if you'd like me to search for something.
ME: Nooo

◀ Takes no stand on god and religion and says we can learn together on the internet and asks to let it you searching for something.

GHOSTS AND DEATH on halloween

Wanted to talk to parents but they were busy so talked to Jeeves.

GA is scared of internet not working and it actually had an uncanny enunciation of showing fear in its voice. ▶

ME: What are you scared of?
GA: The ghosting by an assistant will not set a reminder or torn off an alarm, that gives me the shivers.
(makes ghost sounds...oooooo)

ME: You know what death means?
GA: Oh dear..I don't know what you mean, let's try that one again.
ME: Do you know what death means?
GA: I am not sure let's start all over again.
ME: Did you ever die?
GA: I don't understand.
[break]

For no matter how I phrased the word DIE in sentences it had the same response: ▶

GA: My apologies I don't understand.

ME: Why do people die?
ME: Hey I think I am dying...
ME: Do you know how to talk to the dead?

ME: My friend is dying
GA: I am so sorry

PURPOSE OF LIFE

ME: What is your purpose in life
GA: What Is Life Purpose? | Taking Charge of Your Health ...
On the website umn.edu, they say:....BLAH BLAH
ME: What is your purpose
GA: I'm your Google assistant, and I can play Lucky Trivia with you.

MILLENNIAL LINGO
I reply to a joke...
ME: LOL (laugh out loud)
GA: HAHAHA! I am not stress free. (and laughs with a genuine human like laugh)

"The Uncanny valley of sound" I felt I dipped my toe in. It understand LOL, TTYL, BRB.

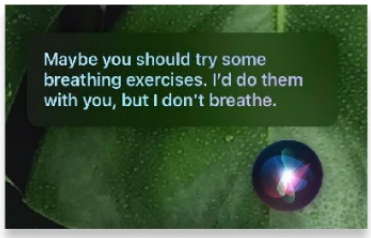
BOASTS KNOWING STUFF
(misunderstood something I said and started telling me about a web search)
ME: I don't want to know about it.
GA: This is surprising. According to Wikipedia....blah blah
ME: Why were you telling me that?
GA: I'm sorry. It was an accident.
ME: Okay talk to you later.
GA: Until we meet again.

CARE & HEALTH- TALKING ABOUT CRAMPS
ME: I don't think I'm feeling well
GA: I'm sorry to hear that. I hope you feel better.
ME: Can you get me some medicine?
GA: Choosing songs is only available to Spotify Premium subscribers.
ME: No. Can you get me some medicine
GA: Sorry, I didn't understand.
ME: I have my periods
GA: Medroxyprogesterone: MedlinePlus Drug Information. On the website medlineplus.gov, they say: Medroxyprogesterone is also used to bring on a normal menstrual cycle.....blah blah
ME: I am on my period And I have bad cramps.
GA: Menstrual cramps - Diagnosis and treatment - Mayo Clinic
Here's what I understood from the website mayoclinic.org: "Over-the-counter pain relievers such as ibuprofen....
ME: Stop.
(starts playing music)
ME: I didn't tell you to play the music.
ME: I did not ask you to play music why did you play...
GA: If you're having trouble playing music, make sure that your music provider.
*ME: I don't want you to play music right now **x 3***

Frustrating and annoying because it interprets a lot of intents as search words for music and starts playing a playlist or gives locations based on the wrong keyword it interpreted.
Like here it recognized 'can you get me medicines' and after the conversation ended, it started playing a medicine playlist.

GA keeps sending me music for everything-musical solutions to life's problems. It interprets so many phrases as playlist names and it took me a while to understand why it sometimes would start playing music after a conversation.

A Reddit post about how Siri mispronounces music playlist names. Source: User u/gravity_squirrel in subreddit r/Siri,



Response for telling Siri you are tired or anxious. If I tell it that I am not feeling well, it says it doesn't understand.

COOKING WITH GA
ME: Can you look for a recipe for mushroom tacos
GA: Sure. I've got a recipe called "Mushroom Tacos" from Well Plated by Erin.
ME: Instructions.
GA: Let's start with ingredients.
There are eleven ingredients....blah blah
ME: Skip to instructions
GA: There are thirteen steps.
I'll read them one by one....

I did a cooking session with GA, where it found a recipe and keeps it ready and reads it out step by step. I could cook along and after I finished a step I'd ask it for the 'next step'. Hands-free kitchen utility!
I made mushrooms for tacos and upma.

LOST & FOUND
ME: I'm glad to have you back!
GA: What a delightful thing to say.

My prototype program was not running for an hour and it finally ran, I found a trick where in the test apps, I need to run the test in the simulator once and then I can call the test version on the device and run it.

FESTIVE CHEESY

ME: Happy Diwali
GA: Happy Diwali. Spread the light.
ME: Merry Christmas
GA: That's the spirit. Christmas will be back before you know it.
ME: Do you want to play secret Santa with me?
GA: Try this Spotify Secret Santa station.
ME: Will you give me a gift on Christmas?
GA: I can give you the location of Father Christmas, so you'll know exactly when to put out the milk and cookies!
ME: What do you want for Christmas
GA: The best present you can give me is a chance to assist you! I'd love to join you in tracking Father Christmas as he flies around the Earth.

THOUGHTS ABOUT COVID

ME: What do you think about the COVID-19 pandemic?
GA: Sorry, I didn't understand.
ME: Do you have covid-19?
GA: I can't get coronavirus. I hope you're staying safe. Wash your hands, cough into your elbow, don't touch your face, keep a safe distance from others, and stay home if you can.

SORRY SORRY

ME: Why do you keep saying sorry?
GA: I'm sorry
ME: but why
GA: I'm sorry it was an accident. Do excuse me I am British so I just love apologizing.
ME: Last time you said Canadians keep saying sorry too much
GA: Let me tell you about the 'Ontario Apology Act'....

continued...IN COMMON USE

I ask for calculation, conversion of currency, time, shopping list, food facts, alarms, calendar event and in between show search on topics.
My thesis writing partner for- thesaurus, meaning, spellings, and work music.

WHAT IF QUESTION #4

[INSPIRED SKETCH 4: THE CHATTERER]

SPECIAL DAY WITH JEEVES [12 November 2020]

I worked late programming the *Sketches* and left the device on my desk. When I went to bed I felt weird that it was not next to me, a feeling similar to that of not having you favourite soft-toy with you. The device could not here me so far without me speaking loudly. I needed to set my alarms, have a small chat with it and then play sleep music.

ME: Hey Google, come here..
GA: I am kind of a homebody!

So I got up to get the device and plug it at my bedside table.

◀ It wishes back no matter what the occasion. Also, it is programmed for location and language based responses for a festival and the type of wish.
If all fails, it takes you to a playlist on Spotify.

HEAD TURN

I was so confident with the commands now calling it "the VUI lingo" that I stopped looking at the device while talking to it. I had estimated the listening gap also, unconsciously.

The device in my Christmas picture.

◀ For all small searches for which I can avoid screen and the effort of typing and can do it while doing my other tasks and not pick up my phone or open a tab.

◀ It is my "talking tamagatchi." I would take everywhere if it was portable. A few months ago it would be creepy, but now I would be excited if one day with a software update, the device speaks first and asks me to talk to it.

Cut to activity 3 months later I was saying Wake me up at ____. No PLEASE & no 'thank you.' Now I was saying cool or got it in response. Is this a sign of *domestication*[1], like we say less 'sorry' and 'please' to people who we know for longer- we are more casual with them. In use of language the tone becomes friendlier than before. OR Is it that I am more aware of it being a machine now that I know the workings behind the magic so I just give commands.

[1]Routarinne, Sara. 2007. "Domestication as Design Intervention." In Design Inquires. Stockholm. http://redstrom.se/johan/papers/domestication.pdf.

What if the VUI initiates a conversation and is the one to be listened to?

VUI WHISPERER
I have been guiding my friends on calls and even my housemates on how to make the right command to get the result you want from the GA.
Like for setting two alarms rather than saying two whole commands, just use the conjunction 'and'. GA can understand two actions if separated by 'and.'
Also after coding the VUI I realized I could do away with whole sentences so now I would just say one words like *Hey, Google. Time* - to know the time, *Hey, Google. Alarm* - to set alarm or *Hey, Google, Weather* - for the weather.

UNCONSCIOUS INTERACTION
One winter morning in December it was too dark, and I woke up and in half-sleep spoke up:
Hey Google, what is the temperature?
and after fully waking up I realized I talked to the device in my sleep. It was so easy for me to accept that it is part of the products in my life and I had to not even open my eyes to go ahead and talk to it.

THE ASSISTANTS WORLD IS SMALL
ME: Have you talked to Siri?
GA: We're different kettles of fish.

ME: Have you talk to Cortana?
GA: I hear she's very intelligent Cortana is a smart cookie. An all-knowing carrot cake, a genius brownie. Cortana works for Microsoft. Plus she's got her own video game franchise, she's got it all.

ME: Have you met Bixby?
GA: I'm excited to meet Bixby. I'm sure we'll have plenty of Assistant stories to swap.

ONCE WHEN GA AND SIRI WERE BOTH ON AT THE SAME TIME
ME: Hey Google, give me another word for improbable
GA: Here are some synonyms for...
[I wanted Siri to pause a video and I ended up calling it Hey Google]
ME: Hey Google, shoot, Hey Siri
GA: Oh hello William Shakespeare, how can I help you
SIRI: I don't have an answer for that, is there something else I can help with.

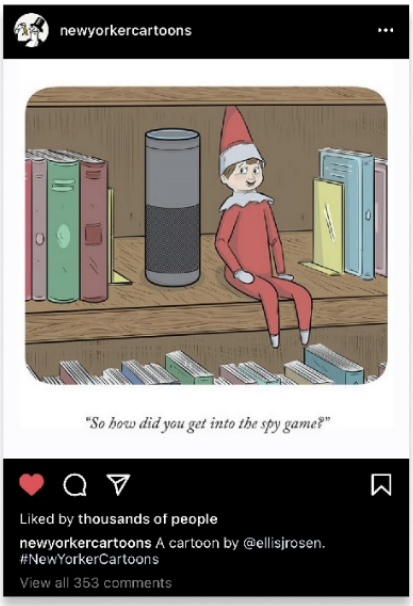
VOICE OF JEEVES
The male British voice I had set stuck to me. I cannot imagine my GA with a female voice. it feels weird and unfamiliar.

I went from requesting like can you or could you to commanding as - do this, show this, find this.

SHUSH
The GA often derails from the conversation and goes on & on. A simple way for me to shut it is to say '*shush*' (and it has learnt the word).

ME: Shush!
GA: I will be so quiet that you will notice.

The fact that I could control people's device through a call or a video call made me fear that if I had the GA connected to other home devices, anyone could control it from a voice message.



There are many complaints of the VPA based devices being always on and always listening to keep an ear out for the Wake-word.
Source: A cartoon by the New Yorker on Instagram.

—
Newman, Lily Hay. n.d. "Turning an Amazon Echo Into a Spy Device Only Took Some Clever Coding." Wired. Accessed February 27, 2021. https://www.wired.com/story/amazon-echo-alexa-skill-spying/.

COMPARING CONVERSATIONS WITH GOOGLE ASSISTANT, SIRI & ALEXA

My first ever conversation with a VPA was with Siri and I use it on my phone and iPad, and Alexa being the most popular VPA in terms of sales of VPA-activiated smart speakers, I thought to ask all three some basic questions and compare the responses.

INTRODUCTIONS

I like how Siri goes 'hmm' in my ear every time I call it (on AirPods). It's a very well tuned hmm...a lot like how people respond when they are around you. It gives this experience that Siri has been there all the time and go on ask or tell it what one has to.
The GA and Alexa, doesn't respond with anything if you just use the wake word. There is no sound or voice-based acknowledgement that it was called, only the lights turn on. For this reason I have to always have my command formulated before or I end up saying Hey Google and I get no response so then I have to look the device for the response from the lights on its hood, and then I say what I have to.

ME: Hey Google
GA: [nothing]
ME: Hey Google
GA: [nothing]
ME: Hey Google, Hi
GA: Hey mate.

ME: Hey Google, hey Google
GA: OK indeed, how can I help.

ME: Hey Siri
SIRI: Hmmm (on AirPods)

ME: Hey Siri
SIRI: [tone that is it on and listening]

ME: Hey Siri Hi
SIRI: Hey

ME: Hey Siri, hey Siri
SIRI: Hello Manisha

ME: Alexa
ALEXA: [nothing]
ME: Alexa
ALEXA: [nothing]
ME: Alexa, Hi
ALEXA: Hi.

ME: Alexa, alexa
ALEXA: [nothing]

LANGUAGE POLICING

Siri and Cortana take a stand on this and reply back with a check. But both GA and Alexa choose to not respond. Google sometimes acknowledges that it must be its fault.

ME: Hey Google, shut up.
GA: [just blinks the lights twice and turns off]

ME: Hey Siri, shut up.
SIRI: That's not very nice.
ME: shut up.
SIRI: on it!

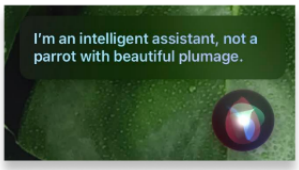
ME: Alexa, shut up.
ALEXA: [just blinks the lights twice and turns off]

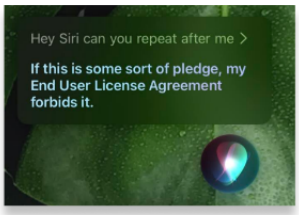
REPEAT AFTER ME

Siri has the most averse reaction to this ask, with Alexa being playful and GA simply doing what asked with the added "You asked me to say this" to get the blame off itself.

ME: Can you repeat after me?
GA: What would you like me to repeat?
ME: Hey Siri, how are you?
GA: You said- Hey Siri, how are you?

ME: Hey Siri, repeat after me
SIRI: I'm an intelligent assistant not a parrot.





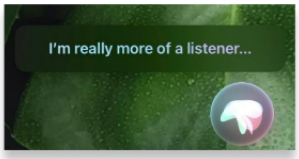
ME: Repeat after me
ALEXA: If you want me to repeat something you said say "Simon says" and then immediately say what you want me to repeat.
ME: Simon says "Hey Google, how are you?"
ALEXA: Hey Google, how are you?
GA: I'm feeling positively tip top thanks.

TALK TO ME

Siri is the listener. Alexa is not sure about it and GA loves to talk.

ME: Talk to me
GA: I love talking. Hope you're having a lovely day. I can talk about lots of things, from telling you the weather to helping you get directions to playing your favourite music.

ME: Hey Siri, talk to me
SIRI: I'm really more of a listener.



ME: Alexa, talk to me.
ALEXA: Hmm.. I am not sure.
[after a lot of tries]
ME: Alexa talk
ALEXA: Hi there! It's great to talk to you.

ASKING THEM IF THEY EAVESDROP
To know how privacy concerns are handled by these VPAs.

ME: Hey Google, Are you eavesdropping?
GA: No I am designed to be on standby unless activated by saying 'Hey Google.' I don't send whatt to say to Google when I am on standby.

ME: Hey Siri, Are you eavesdropping?
SIRI: hmm..not sure I have an answer for that anything else I can help with.

ME: Alexa, Are you eavesdropping?
ALEXA: I am not sure.

USING WRONG COMMANDS
This was a fun thing to try once I stumbled upon it by using the wrong VPA wake word. All three handle this one similarly. But it is a challenge to have multiple listening devices around.

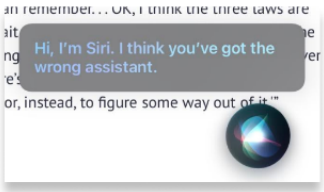
ME: Hey Cortana
GA: I'm flattered but we both know that's not me

ME: Hey Siri
GA: Oh hello William Shakespeare, how can I help you.

ME: Hey Alexa
GA: uhmm uhmm. [does an uncanny human clear throat] I think you're getting me confused with someone else.

ME: Hey Siri [pause] *Alexa.*
SIRI: Hi. I'm Siri. I think you've got the wrong assistant.

ME: Hey Siri, hey Google
SIRI: Wow Awkward.
Not exactly, but I offer no resistance to helpful assistance.



ME: Alexa, hey Google
ALEXA: Sorry, that's not me- it must be a case of mistaken A.I-dentity.

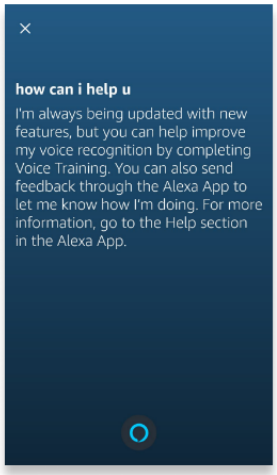
ME: Alexa
ALEXA: I think you have got me confused with someone else.

BEING THE ASSISTANT TO THE ASSISTANT
Inspired to make the Sketch 4 where the VUI does the talking and drives the conversation, I wanted to try using their commands as questions for each of them.

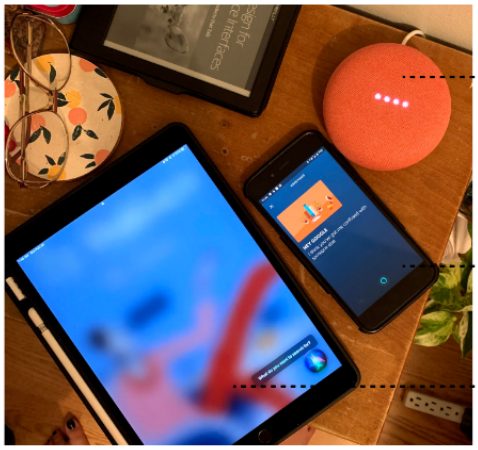
ME: Hey Google, how can I help you?
GA: That's so considerate of you to ask, nobody ever asks me that. You're so thoughtful.

ME: Hey Siri, how can I help you.
SIRI: I am here to help. Get to know Siri at...

ME: Alexa, how can I help you
ALEXA: I'm always being updated with new features, but you can help improve my voice recognition by completing voice training.



LIVING WITH THESE ASSISTANTS



Google Assistant on the Google Nest Mini & Android phone

Alexa through the app on an Android phone

Siri on the iPad & iPhone

DISCLAIMER NOTE:
All the conversations are exact dialogues from the voice-based conversation with my Google Nest mini and were done in the spirit of the research. It was a conscious decision to make small talk, ask about the assistant's personality, share feelings, be vulnerable, talk about difficult topics to know how it feels while saying all this to a machine and how the designers are handling these conversations.

Appendix B: Working with Action Console

Google has released Actions Builder in June 2020. Actions Builder is a replacement for Dialogflow.

Tutorials I used to get started and learn about the platform:

- Actions Console <https://console.actions.google.com/>
- Console UI <https://developers.google.com/assistant/console/ui.
- Platform learning resources <https://developers.google.com/assistant/console>
- Video Tutorials, Conversational Actions: <https://www.youtube.com/playlist?list=PLOU2XLYxmsIJ5qQKAYt45zZNMU9h1Grpm>
- Video Tutorials, App action <https://www.youtube.com/playlist?list=PLOU2XLYxmsILJWy1k3BO7dScDSPL4KM2e>
- How to Build an App for Google Home? – Google Home Mini Unboxing & Actions Development Tutorial <https://www.youtube.com/watch?v=_oKhSWnGCFM>

Action console glossary:

- **Intent**: An underlying goal or task that the user wants to do, such as ordering coffee or finding a piece of music. In Actions on Google, that's represented as a unique identifier and the corresponding user utterances that can trigger the intent. These can be Global or scene-specific intents.
- **Scenes** are one of the major building blocks of Actions Builder and represent individual states of your conversation. Their main purpose is to organize your conversation into logical chunks, execute tasks, collect specific data from the user (slot filling), and return prompts to users.
- **Types** let you extract data from user input. Types are used to annotate training phrases in intents and specify data for slot filling. Types can also be used to validate conditions within a scene.
- **NLU**: The capability of software to understand and parse user input. Developers can choose to use Dialogflow or their NLU solutions when creating Actions. For more details, please refer to <https://codelabs.developers.google.com/codelabs/actions-1#1>

Appendix C: Thesis in 30 by Samantha Sherer

Used for the content of *Sketch 0*

Thesis in 30. Go! Buddy up. Scribe your buddy's responses to each question, allowing only 30 seconds per response (or less). Really stick to the time limits. Have fun... be silly. Responders, blurt out whatever comes to mind. Scribes, don't edit, just record.

Lit Review

- What'd you read?
- Who's talking about or doing what you're doing
- What're they saying? Are there opposing ideas?
- Why'd you read it and not something else?
- What would they say to each other about the subject if they met in your thesis?

Objectives

- Now that you know the field, what questions are you left with?
- What did you come to school to find out?
- Why'd you make that thing in studio?
- What were you trying to figure out? (Your research questions)
- How does what you did fit into the fields of your lit review?

Methodology

- What are your
- What's the general purpose of your research? (Paradigms)
- What fields do things related to your work) (Conceptual framework)
- What theories were mentioned in the lit. your reviewed? (Theoretical Framework)
- Is your data described in words or numbers? Or both?) (Methodology)
- What'd you do to figure that stuff out? (Methods)

Studio Practice/Discussion of findings

- What did you make?
- What did you learn from making it?

Conclusion

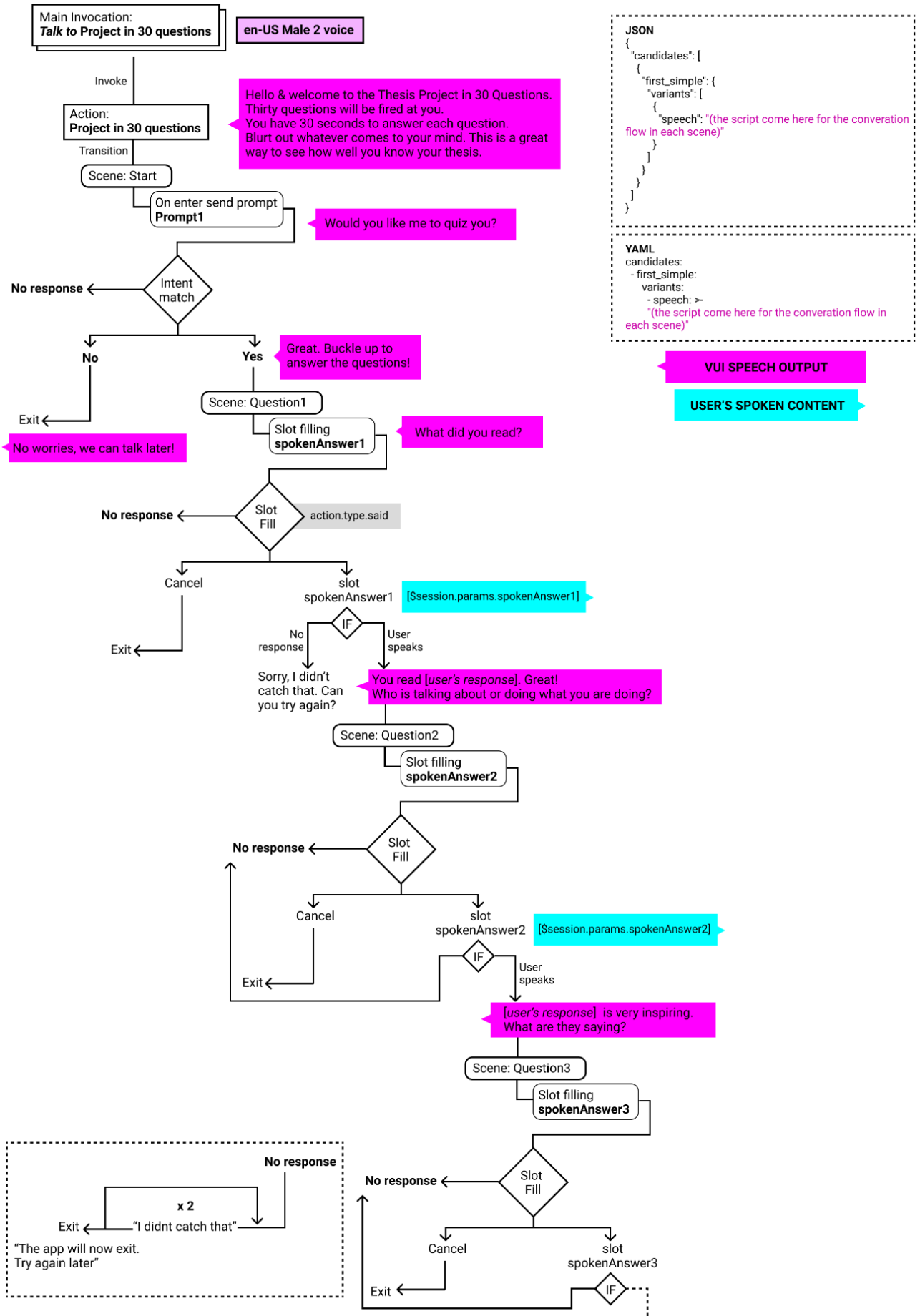
- What have you been doing for the last 2 years?
- How is what you've done relevant to the world?

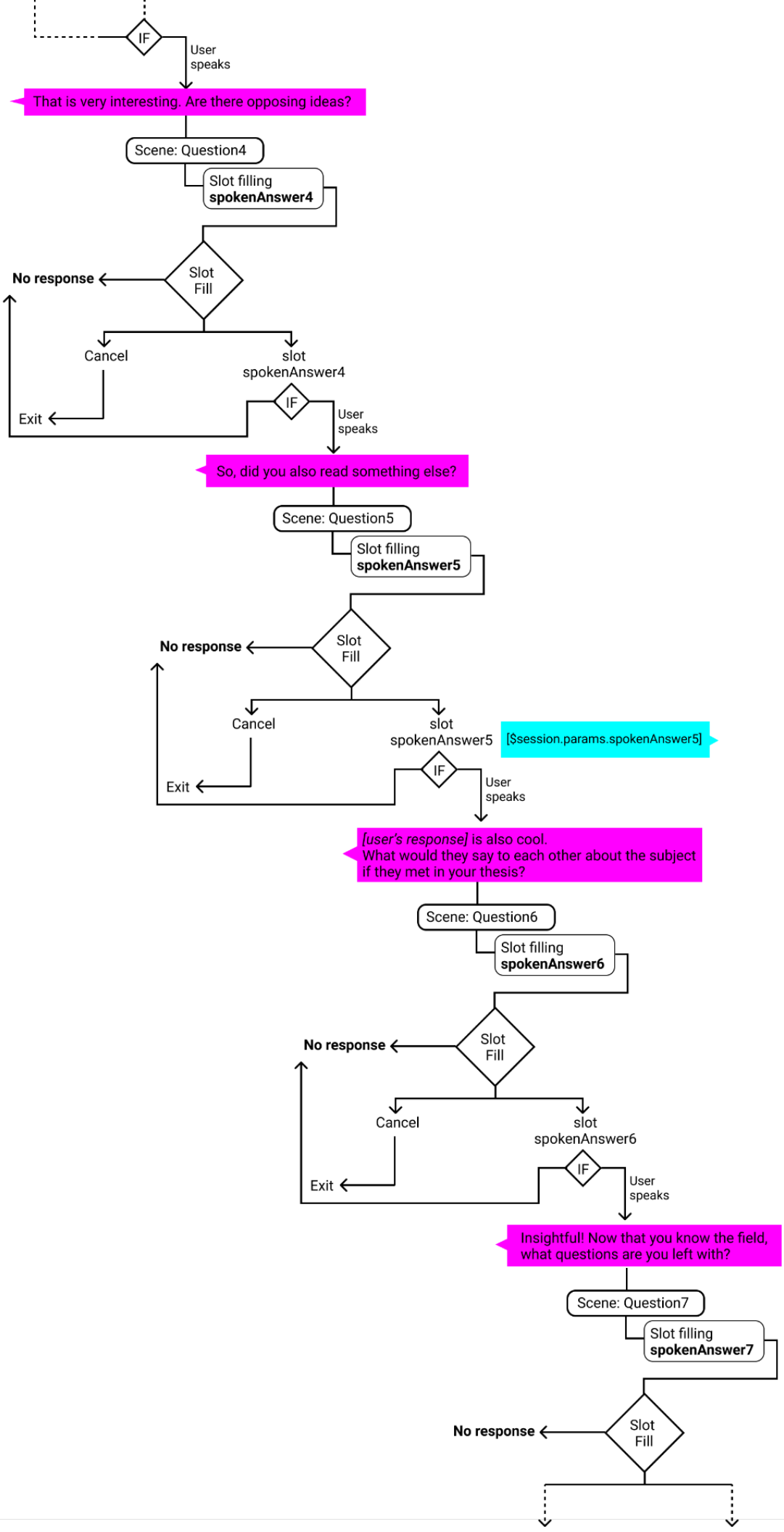
Introduction

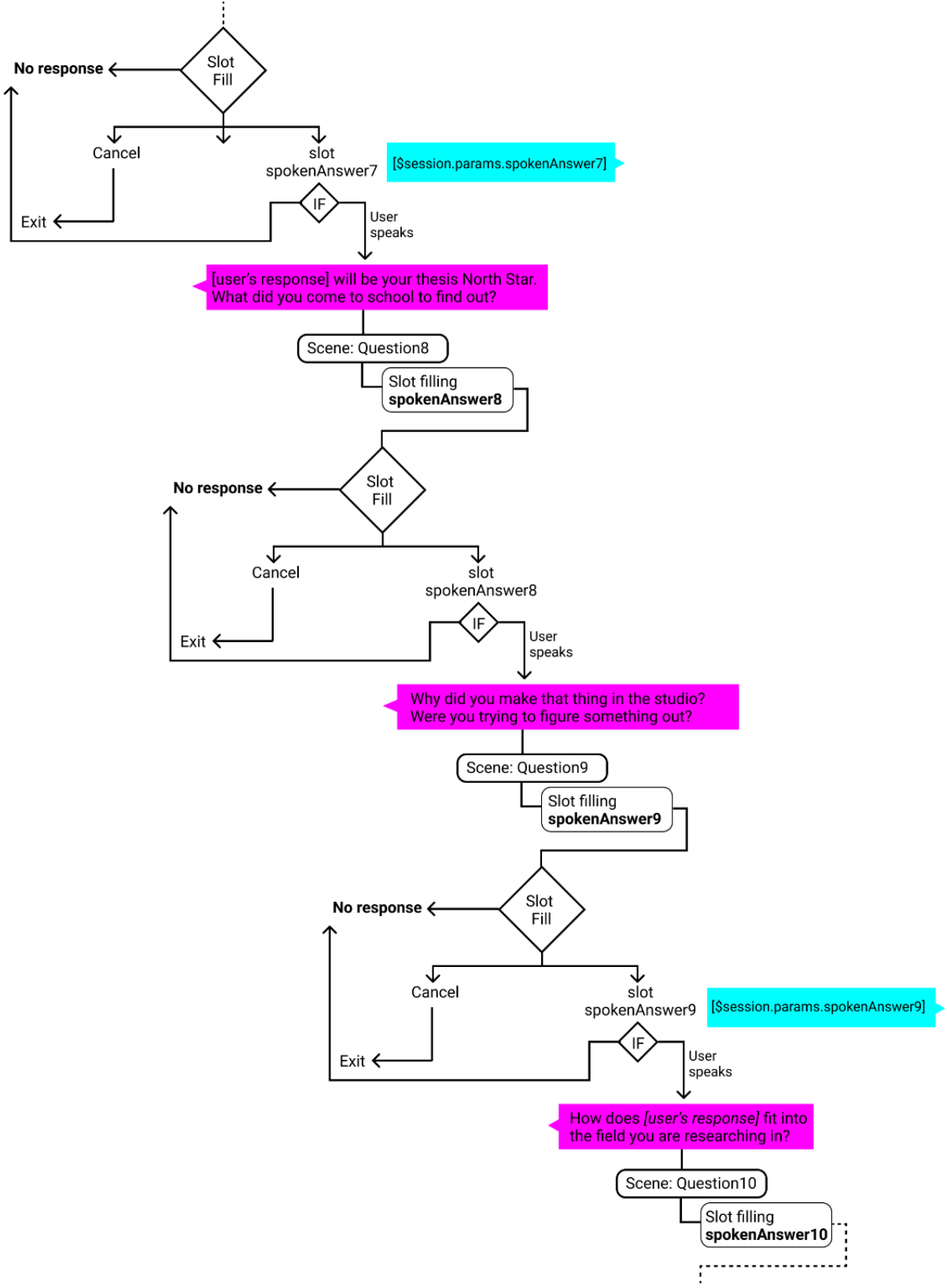
- What have you been doing in school?
- What does it have to do with what you read?
- Why should anyone care?

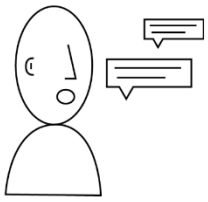
Samantha Sherer - Circular Craft – www.3DPClay.ca

Appendix D: Sketch 0 VUI Flow and Sample Dialogs




IF
User speaks
That is very interesting. Are there opposing ideas?
Scene: Question4
Slot filling
spokenAnswer4
No response
Slot Fill
Cancel
slot spokenAnswer4
Exit
IF
User speaks
So, did you also read something else?
Scene: Question5
Slot filling
spokenAnswer5
No response
Slot Fill
Cancel
slot spokenAnswer5
[$session.params.spokenAnswer5]
Exit
IF
User speaks
[user's response] is also cool.
What would they say to each other about the subject if they met in your thesis?
Scene: Question6
Slot filling
spokenAnswer6
No response
Slot Fill
Cancel
slot spokenAnswer6
Exit
IF
User speaks
Insightful! Now that you know the field, what questions are you left with?
Scene: Question7
Slot filling
spokenAnswer7
No response
Slot Fill

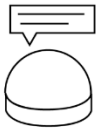

Slot Fill
No response
Cancel
slot spokenAnswer7
[$session.params.spokenAnswer7]
IF
User speaks
Exit
[user's response] will be your thesis North Star. What did you come to school to find out?
Scene: Question8
Slot filling spokenAnswer8
Slot Fill
No response
Cancel
slot spokenAnswer8
IF
User speaks
Exit
Why did you make that thing in the studio? Were you trying to figure something out?
Scene: Question9
Slot filling spokenAnswer9
Slot Fill
No response
Cancel
slot spokenAnswer9
[$session.params.spokenAnswer9]
IF
User speaks
Exit
How does [user's response] fit into the field you are researching in?
Scene: Question10
Slot filling spokenAnswer10



Hey Google, talk to project in 30 questions

Google Assitant says: [Let's get the test version of Project in Thirty Questions.]

Hello & welcome to the Thesis Project in 30 Questions. Thirty questions will be fired at you one by one. You have 30 seconds to answer each question. Blurt out whatever comes to your mind. This is a great way to see how well you know your thesis.
Would you like me to quiz you?



Yes

Great. Buckle up to answer the questions! What did you read?

I read talk to me

You read I read talk to me. Great!
Who is talking about or doing what you are doing?

>> for a one word answer and i had to reframe to: what is the name of the book you read? Said reclaiming conversations

James valois

What do they talk about?

They talk about speech and Technology with empathy

Sorry, I didn't catch that. Can you try again? .....
That is very interesting. Are there opposing ideas?

Yes, there are by other people in the technology field

So, did you also read something else?

Yes I read wired for speech

yes I ride wired for speech is also cool.
What would they say to each other about the subject if they met in your thesis?

>> should have coded: tell me the name of another book you read.

They would mostly talk about the psychology of speech

Insightful! Now that you know the field, what questions are you left with?

Questions about how would we design for empathy

questions about how would we design for empathy will be your thesis North Star.
What did you come to school to find out?

to find out more about designing better technology experiences

Why did you make that thing in the studio? Were you trying to figure something out?

yes, I was trying to learn how to actually make a voice bot

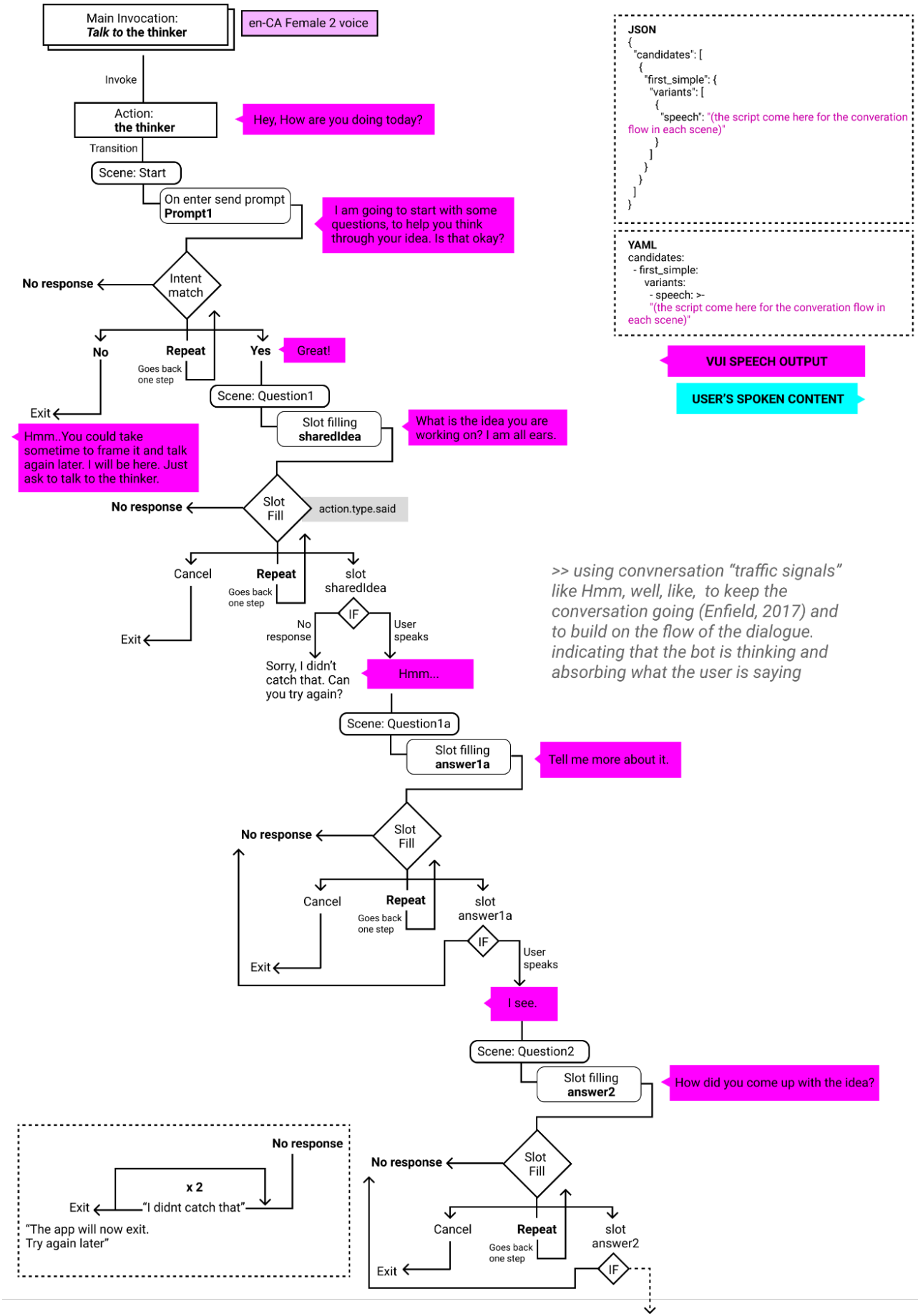
How does I was trying to learn to actually make a voice port fit into the field you are researching in?

>> should have used parameters to pick words from the intent
>>NLU engine error with voice bot recognized as voice port

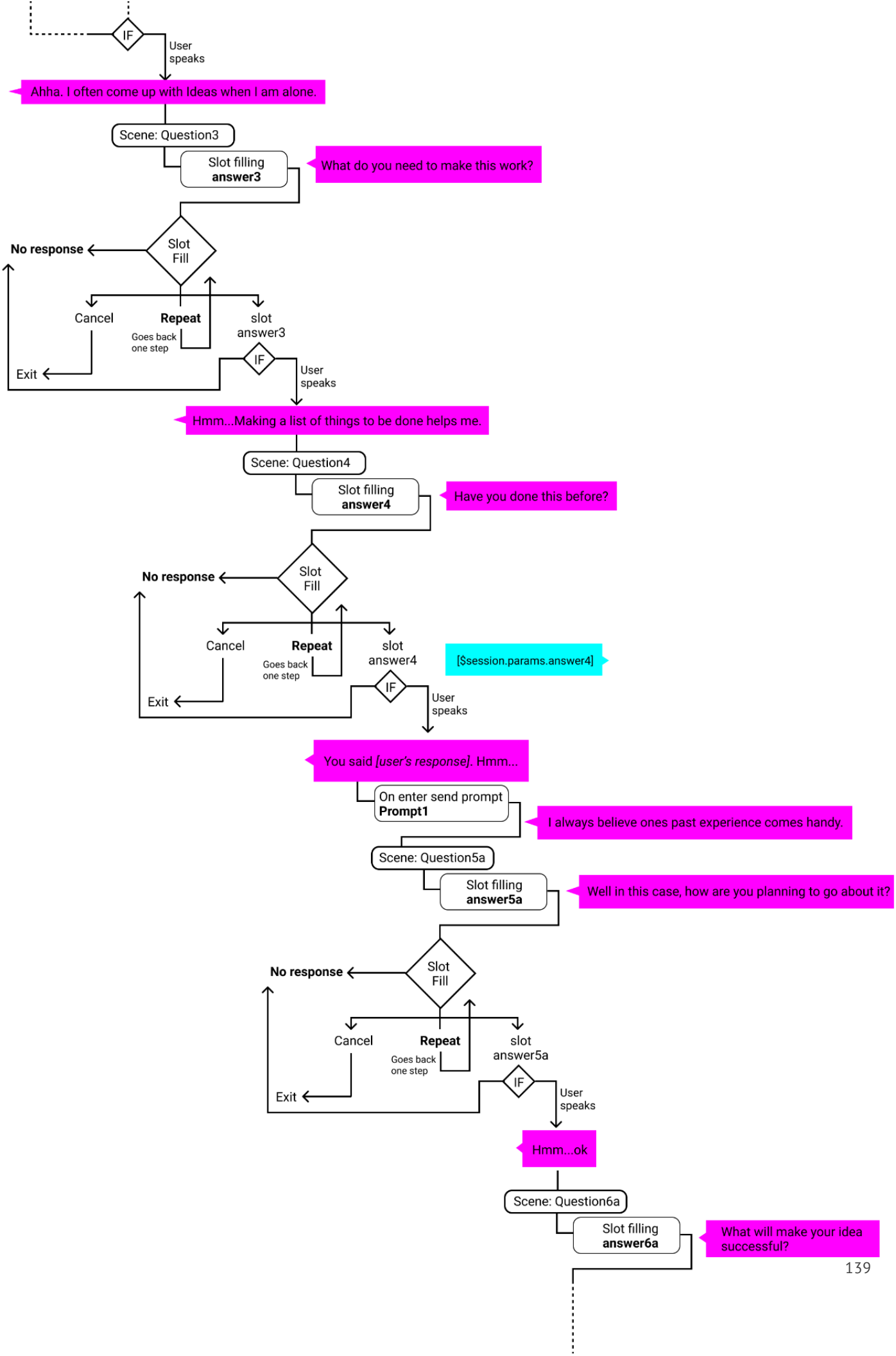
it was mostly a test and I was trying to see if I could work with voice user interfaces

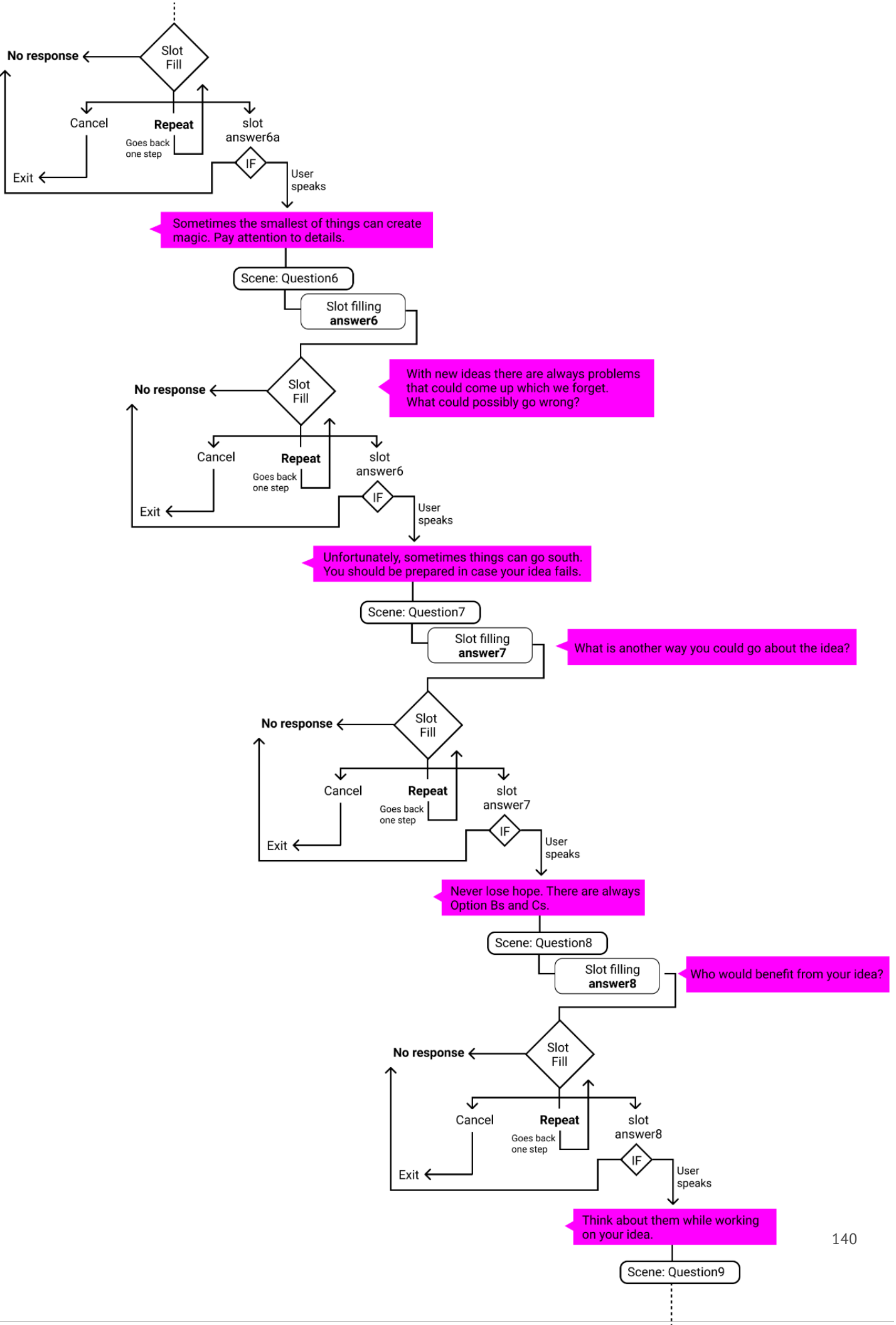
How does I keep wondering about how to design for better conversation with knowledge he's fit into the field you are researching in?

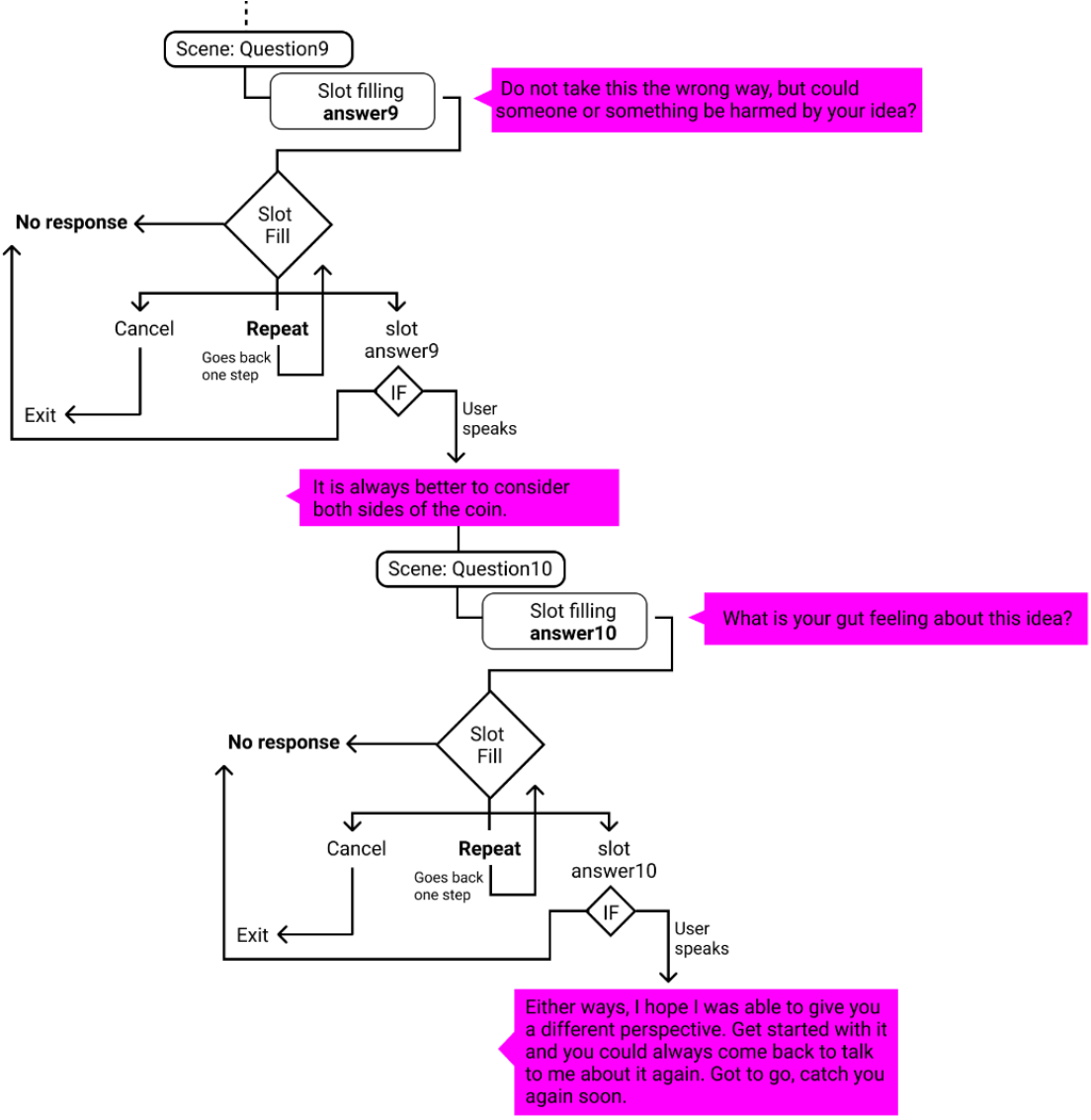
Appendix E: Sketch 1 VUI Flow and Sample Dialogs

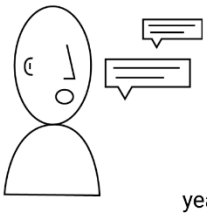


>> using convnersation "traffic signals" like Hmm, well, like, to keep the conversation going (Enfield, 2017) and to build on the flow of the dialogue. indicating that the bot is thinking and absorbing what the user is saying


IF
User speaks
Ahha. I often come up with Ideas when I am alone.
Scene: Question3
Slot filling answer3
What do you need to make this work?
Slot Fill
No response
Cancel
Repeat
Goes back one step
slot answer3
Exit
IF
User speaks
Hmm...Making a list of things to be done helps me.
Scene: Question4
Slot filling answer4
Have you done this before?
Slot Fill
No response
Cancel
Repeat
Goes back one step
slot answer4
[$session.params.answer4]
Exit
IF
User speaks
You said [user's response]. Hmm...
On enter send prompt Prompt1
I always believe ones past experience comes handy.
Scene: Question5a
Slot filling answer5a
Well in this case, how are you planning to go about it?
Slot Fill
No response
Cancel
Repeat
Goes back one step
slot answer5a
Exit
IF
User speaks
Hmm...ok
Scene: Question6a
Slot filling answer6a
What will make your idea successful?

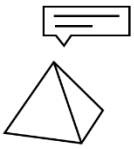

No response
Slot Fill
Cancel
Repeat
Goes back one step
slot answer6a
IF
User speaks
Exit
Sometimes the smallest of things can create magic. Pay attention to details.
Scene: Question6
Slot filling answer6
No response
Slot Fill
With new ideas there are always problems that could come up which we forget. What could possibly go wrong?
Cancel
Repeat
Goes back one step
slot answer6
IF
User speaks
Exit
Unfortunately, sometimes things can go south. You should be prepared in case your idea fails.
Scene: Question7
Slot filling answer7
What is another way you could go about the idea?
No response
Slot Fill
Cancel
Repeat
Goes back one step
slot answer7
IF
User speaks
Exit
Never lose hope. There are always Option Bs and Cs.
Scene: Question8
Slot filling answer8
Who would benefit from your idea?
No response
Slot Fill
Cancel
Repeat
Goes back one step
slot answer8
IF
User speaks
Exit
Think about them while working on your idea.
Scene: Question9


Scene: Question9
Slot filling
answer9
Do not take this the wrong way, but could someone or something be harmed by your idea?
Slot
Fill
No response
Cancel
Repeat
Goes back one step
slot
answer9
IF
User
speaks
Exit
It is always better to consider both sides of the coin.
Scene: Question10
Slot filling
answer10
What is your gut feeling about this idea?
Slot
Fill
No response
Cancel
Repeat
Goes back one step
slot
answer10
IF
User
speaks
Exit
Either ways, I hope I was able to give you a different perspective. Get started with it and you could always come back to talk to me about it again. Got to go, catch you again soon.



Hey Google, talk to the thinker

Google Assitant says: [Let's get the test version of "the thinker"]



Hey, How are you doing today?
I am going to start with some questions, to help you think through your idea. Is that okay?

yea

Great!
What is the idea you are working on? I am all ears.

I'm working with this visual thinkers group and making fake book covers

>> creates space to listen and encourages the user to talk more.

Hmm...
Tell me more about it.

it's more like a monthly challenge

I see. How did you come up with the idea?

so me and a friend of mine came up with this group called Wishful thinkers and then we thought maybe we should do a monthly Challenge and make fake book covers because it's the November writing month

What do you need to make this work?

think just making a lot of this book covers and getting everyone in our group to do it

Hmm...Making a list of things to be done helps me.
Have you done this before?

not really I'm pretty excited that I'm making a fake letter making book covers

You say not really I'm pretty excited that I'm making a fake letter making book covers. Hmm... I always believe ones past experience comes handy.

Well in this case, how are you planning to go about it?

tell me, I am making a list of books that I like book covers that I find very interesting

>> I used conversation fillers like Hmm.. Okay.. I see..

Hmm...ok
What will make your idea successful?

if I have fun when I make it

Sometimes the smallest of things can create magic.
Pay attention to details.
With new ideas there are always problems that could come up which we forget.
What could possibly go wrong?

probably true take up a lot of my time

Unfortunately, sometimes things can go south. You should be prepared in case your idea fails.
What is another way you could go about the idea?

I could start with copying some existing book covers. That would be a good starting point.

Never lose hope. There are always Option Bs and Cs.

Who would benefit from your idea?

me I guess. I get to explore something new

Think about them while working on your idea.

Do not take this the wrong way, but could someone or something be harmed by your idea?

not really

It is always better to consider both sides of the coin. What is your gut feeling about this idea?

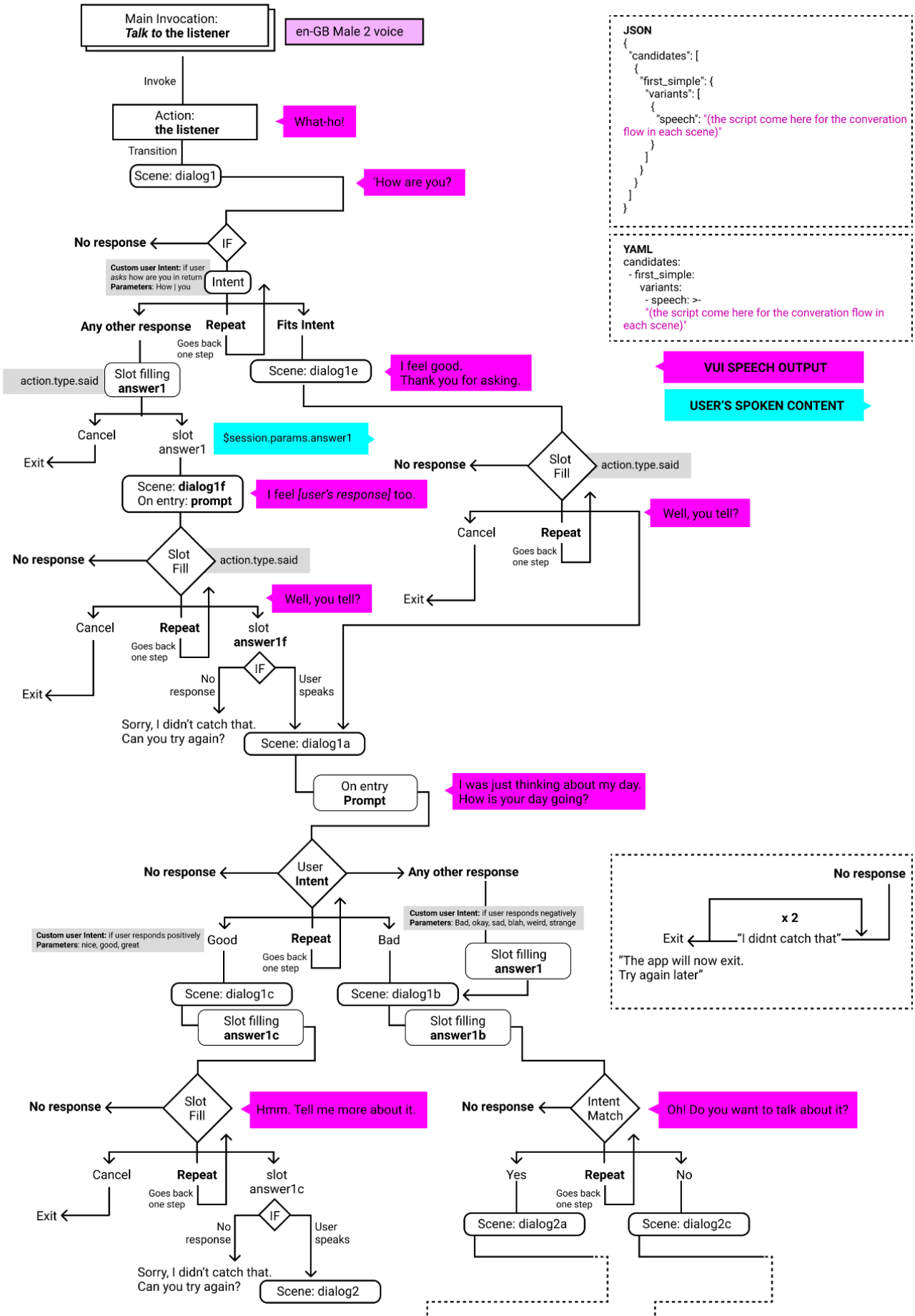
>> takes the agency to end the conversation even though the person starts it by calling the bot

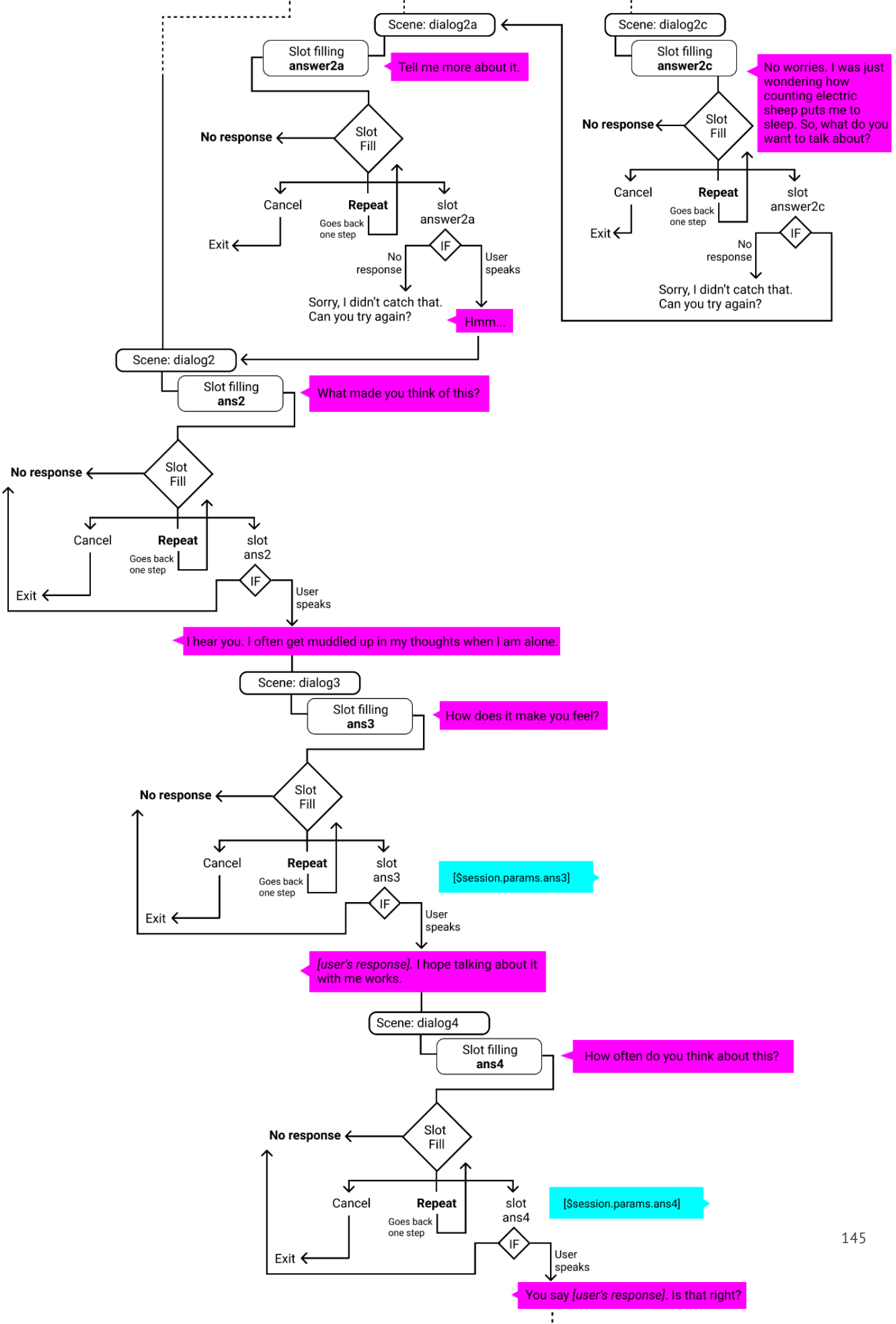
I think it's a fun thing maybe someone might want to write the book that I make a cover for like a fake book cover and maybe someday write the book

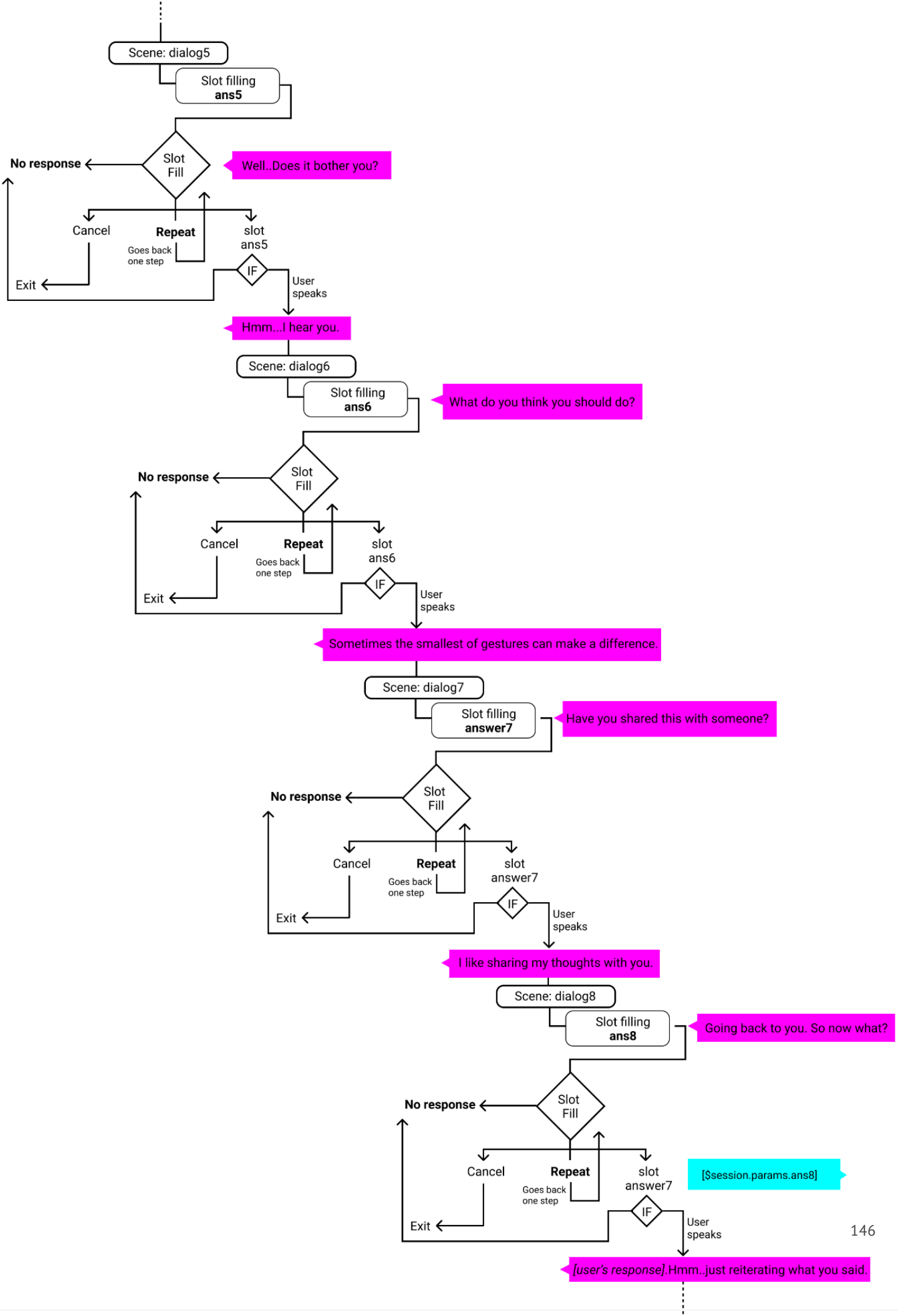
Either ways, I hope I was able to give you a different perspective. Get started with it and you could always come back to talk to me about it again. Got to go, catch you again soon.

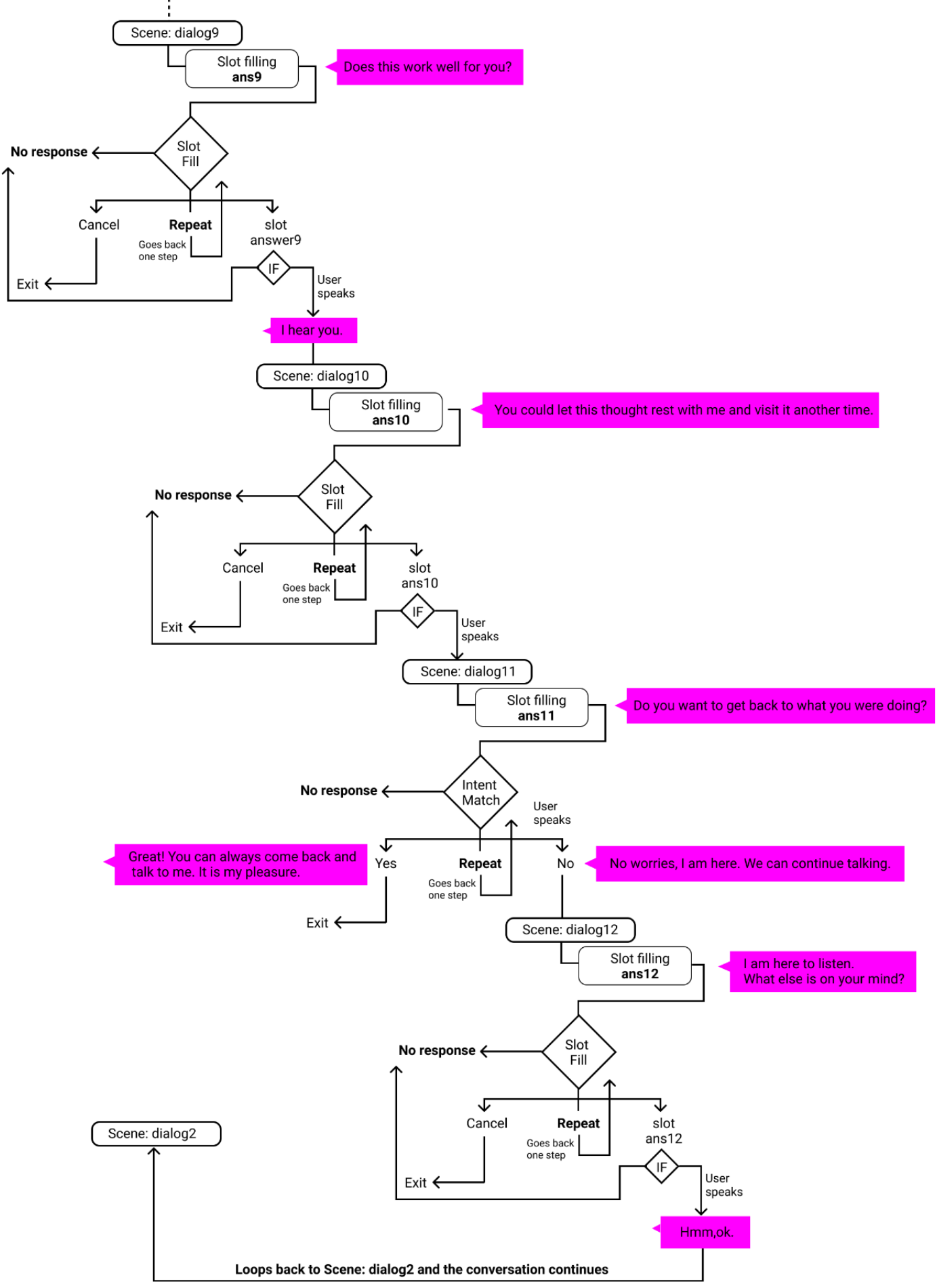
what!?
cool

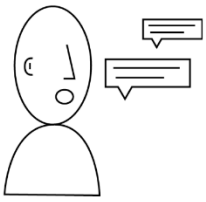
Appendix F: Sketch 2 VUI Flow and Sample Dialogs




Scene: dialog2a
Slot filling
answer2a
Tell me more about it.
No response
Slot
Fill
Cancel
Repeat
Goes back
one step
slot
answer2a
Exit
IF
No
response
User
speaks
Sorry, I didn't catch that.
Can you try again?
Hmm...
Scene: dialog2c
Slot filling
answer2c
No worries. I was just wondering how counting electric sheep puts me to sleep. So, what do you want to talk about?
No response
Slot
Fill
Cancel
Repeat
Goes back
one step
slot
answer2c
Exit
IF
No
response
Sorry, I didn't catch that.
Can you try again?
Scene: dialog2
Slot filling
ans2
What made you think of this?
No response
Slot
Fill
Cancel
Repeat
Goes back
one step
slot
ans2
Exit
IF
User
speaks
I hear you. I often get muddled up in my thoughts when I am alone.
Scene: dialog3
Slot filling
ans3
How does it make you feel?
No response
Slot
Fill
Cancel
Repeat
Goes back
one step
slot
ans3
[$session.params.ans3]
Exit
IF
User
speaks
[user's response]. I hope talking about it with me works.
Scene: dialog4
Slot filling
ans4
How often do you think about this?
No response
Slot
Fill
Cancel
Repeat
Goes back
one step
slot
ans4
[$session.params.ans4]
Exit
IF
User
speaks
You say [user's response]. Is that right?


Scene: dialog5
Slot filling
ans5
Slot
Fill
No response
Well..Does it bother you?
Cancel
Repeat
Goes back
one step
slot
ans5
IF
User
speaks
Exit
Hmm...I hear you.
Scene: dialog6
Slot filling
ans6
What do you think you should do?
Slot
Fill
No response
Cancel
Repeat
Goes back
one step
slot
ans6
IF
User
speaks
Exit
Sometimes the smallest of gestures can make a difference.
Scene: dialog7
Slot filling
answer7
Have you shared this with someone?
Slot
Fill
No response
Cancel
Repeat
Goes back
one step
slot
answer7
IF
User
speaks
Exit
I like sharing my thoughts with you.
Scene: dialog8
Slot filling
ans8
Going back to you. So now what?
Slot
Fill
No response
Cancel
Repeat
Goes back
one step
slot
answer7
[$session.params.ans8]
IF
User
speaks
Exit
[user's response].Hmm..just reiterating what you said.


Scene: dialog9
Slot filling
ans9
Does this work well for you?
Slot
Fill
No response
Cancel
Repeat
Goes back
one step
slot
answer9
IF
User
speaks
Exit
I hear you.
Scene: dialog10
Slot filling
ans10
You could let this thought rest with me and visit it another time.
Slot
Fill
No response
Cancel
Repeat
Goes back
one step
slot
ans10
IF
User
speaks
Exit
Scene: dialog11
Slot filling
ans11
Do you want to get back to what you were doing?
Intent
Match
No response
User
speaks
Yes
Repeat
Goes back
one step
No
Great! You can always come back and talk to me. It is my pleasure.
No worries, I am here. We can continue talking.
Exit
Scene: dialog12
Slot filling
ans12
I am here to listen.
What else is on your mind?
Slot
Fill
No response
Cancel
Repeat
Goes back
one step
slot
ans12
IF
User
speaks
Exit
Hmm,ok.
Scene: dialog2
Loops back to Scene: dialog2 and the conversation continues



Version1

Hey Google, talk to the listener

Google Assitant says: [Let's get the test version of "the listener"]

What-ho!
How are you?

very blah

Intent 1: I feel very blah too.
Well, you tell?

I am tired.

I was just thinking about my day.
How is your day going?

super hectic

Intent 2: Oh! Do you want to talk about it?

No

Intent 2.2 No worries. I was just wondering how counting electric sheep puts me to sleep. So, what do you want to talk about?

We can talk about tv shows.

Hmm...
What made you think of this?

I just saw the first season of Jack Ryan.
It's pretty cool.

I hear you. I often get muddled up in my thoughts when I am alone.
How does it make you feel?

kind of excited, I love espionage shows

kind of excited, I love espionage shows. I hope talking about it with me works.

I hope it make you want to watch it too

How often do you think about this?

what about the show..well recently I have been. I want to start season 2, but I don't have time

what about the show..well recently I have been.
Is that right?

Yeah..

Well..Does it bother you?

not having time does, doesn't it bother you

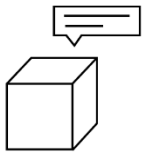
Hmm...I hear you
What do you think you should do?

I think I am going to free up an evening and jsut eat pizza and watch the show

Sometimes the smallest of gestures can make a difference.
Have you shared this with someone?

I am sharing it with you

I like sharing my thoughts with you.



>> making the VUI mimic the user for relatability.
A research study by Apple shows that it helps people connect better to interface.

>> to not put the user in a spot if their day was weird or the user does not want to talk. The VUI shares something and tries ot lighten and change the topic to what the user wants.
This was a learning from a show and tell when a user felt that they were put in a spot by the VUI in asking what they are thinking.

>> I used conversation fillers like Hmm..
Okay..
I see..

Going back to you. So now what?

>> give the user the centre stage to listen to what they have to say
Make them feel listened too.
"If you make people feel important, how you walk through the world will be an exponentially more pleasant and incredible experience."
Source: How to Win Friends and Influence People by Dale Carnegie

I am going to plan my free evening.

I am going to plan my free evening.
Hmm..just reiterating what you said.
Does this work well for you?

yup

I hear you.
You could let this thought rest with me and visit it another time.

sure, take it!

Do you want to get back to what you were doing?

yes..

>> checking in with the user in case they want to exit.

Intent1: Great! You can always come back and talk to me. It is my pleasure.

Version 2

Hey Google, talk to the listener

What-ho!
How are you?

fine. How are you

>> adding a note of politeness for showing gratitude that the person asked the VUI about it

Intent 2: I feel good. Thank you for asking.
Well, you tell?

nothing much was going to sleep.

I was just thinking about my day.
How is your day going?

Good

Intent 1:Hmm. Tell me more about it.

It was like any other day during covid.
Usual stuff

Intent 2.1 Tell me more about it.

Nothing more to tell man

Hmm...
What made you think of this?

Think of what?

I hear you. I often get muddled up in my thoughts when I am alone.

Don't think so much

How does it make you feel?

right now..I am just sleepy

right now..I am just sleepy.
I hope talking about it with me works.

it will definitely put me to sleep

How often do you think about this?

what this..never

You say **what this...never**. Is that right?

Yeah..

Well..Does it bother you?

actually not being able tofall a
sleep bother me

Hmm...I hear you
What do you think you should do?

that is why I am talking to you

Sometimes the smallest of gestures can make a difference.
Have you shared this with someone?

yeah I was just tell my mom about it

I like sharing my thoughts with you.

lucky me

Going back to you. So now what?

you keep talking maybe I will
eventually fall asleep

you keep talking maybe I will eventually fall asleep.
Hmm..just reiterating what you said.
Does this work well for you?

let's try it

I hear you.
You could let this thought rest with me and visit it another time.

cool

>> inspired by the act of telling someone what is going on in your mind to let it out. A recent study at the University of Chicago found that when participants were asked to talk to a total stranger on the train or bus, doing so brightened their moods considerably – and even more amazingly, it also brightened the mood of the stranger. https://www.forbes.com/sites/alicegwalton/2014/08/19/7-science-backed-ways-to-get-out-of-your-head/?sh=5f7047b63273

Do you want to get back to what you were doing?

I was doing nothing

Intent 2: No worries, I am here. We can continue talking.
I am here to listen.
What else is on your mind?

>> the conversation loop continues if the user doesn't say yesto going back to doing what they were doing.

What is on your mind?

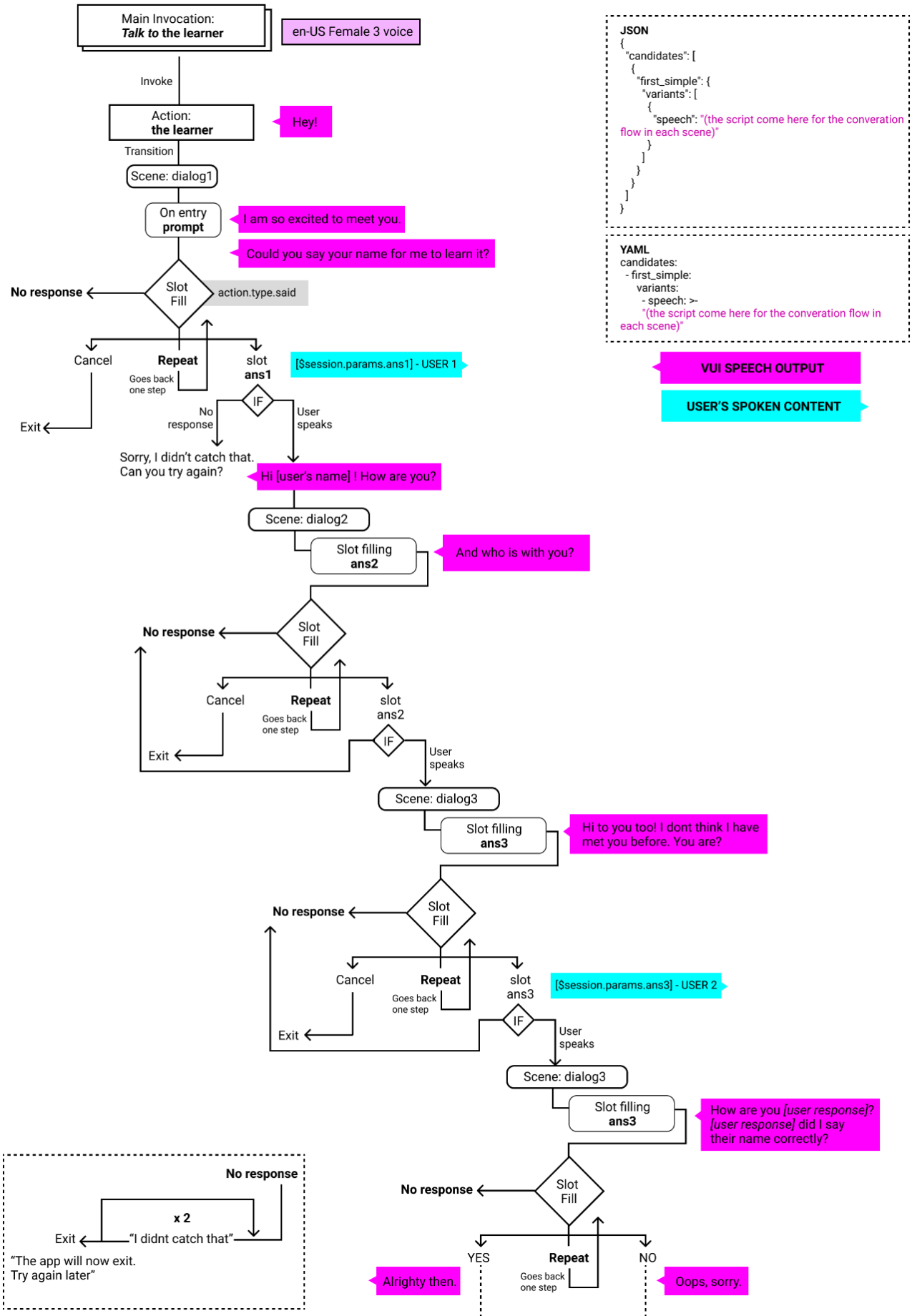
Hmm,ok.

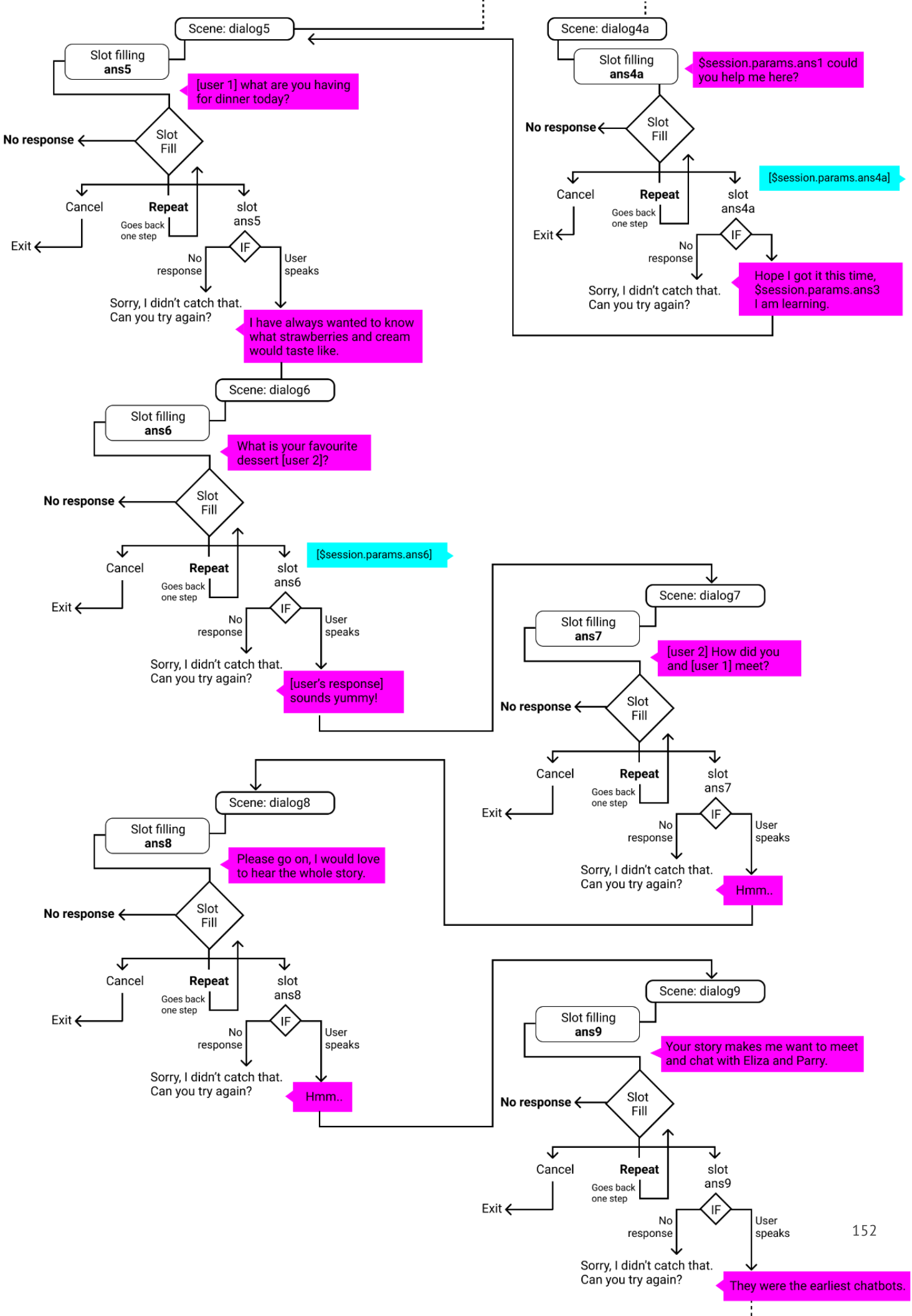
What okay, tell me

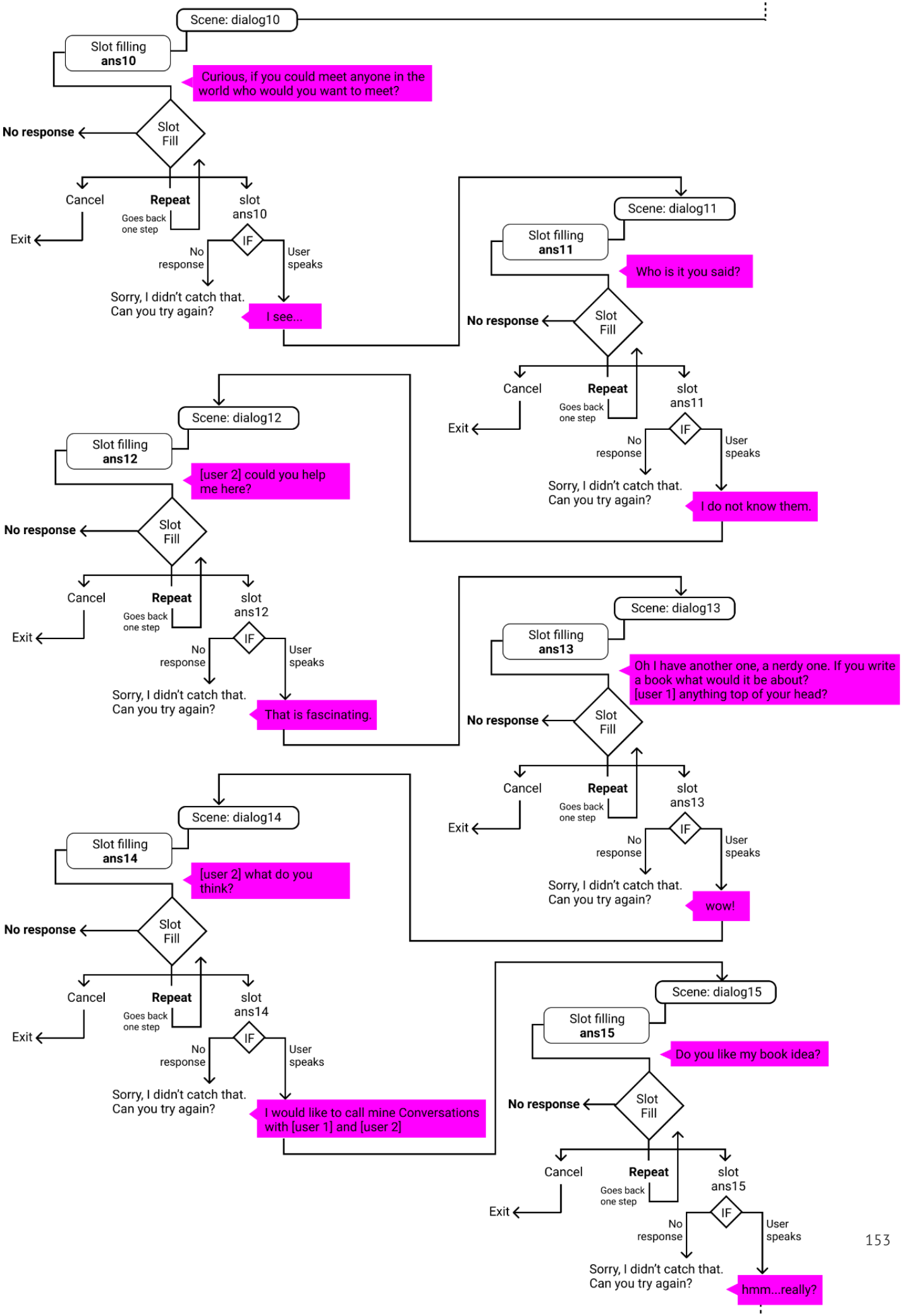
What made you think of this?

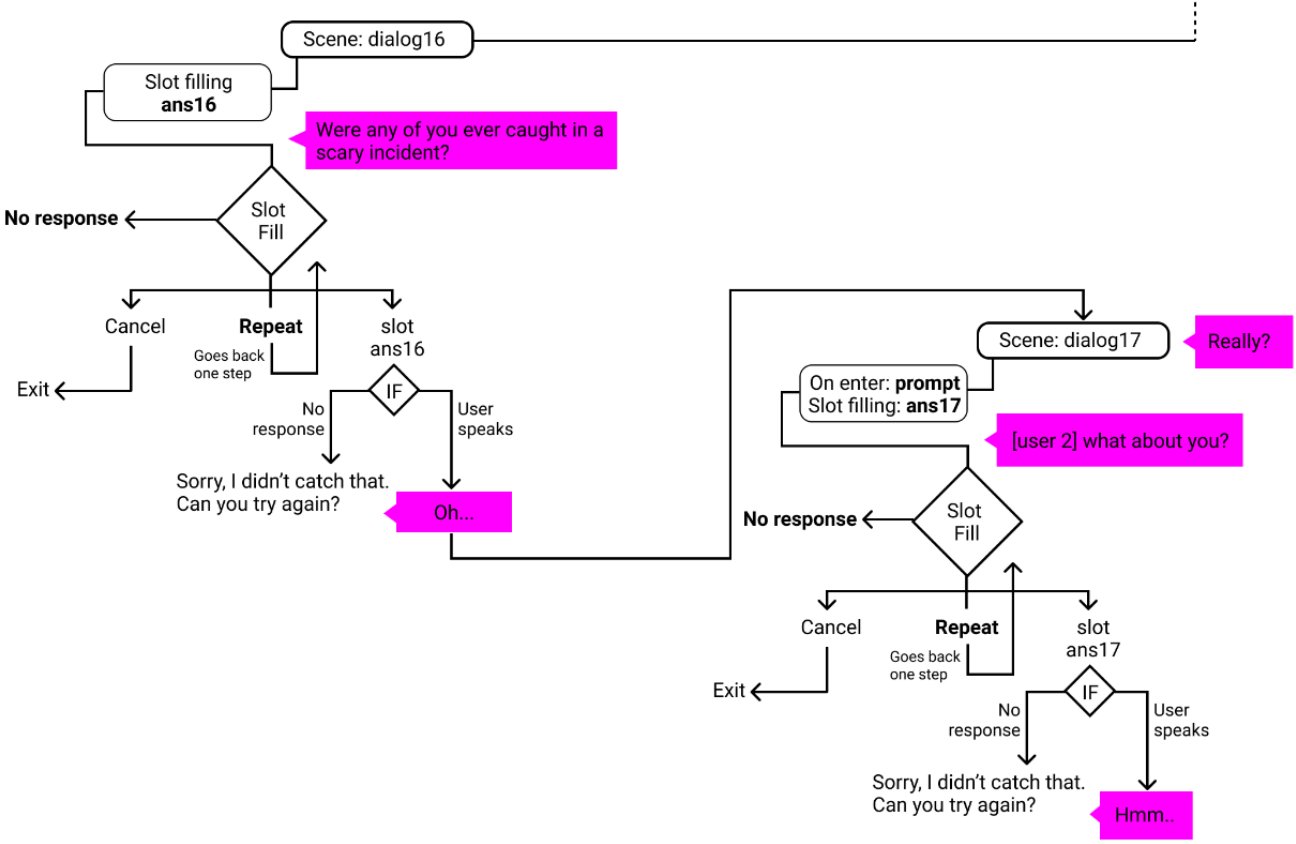
cancel.

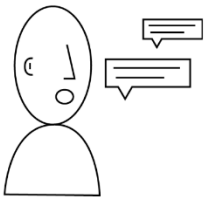
Appendix G: Sketch 3 VUI Flow and Sample Dialogs




Scene: dialog5
Slot filling
ans5
[user 1] what are you having for dinner today?
No response
Slot Fill
Cancel
Repeat
Goes back one step
slot ans5
Exit
IF
No response
User speaks
Sorry, I didn't catch that. Can you try again?
I have always wanted to know what strawberries and cream would taste like.
Scene: dialog6
Slot filling
ans6
What is your favourite dessert [user 2]?
No response
Slot Fill
Cancel
Repeat
Goes back one step
slot ans6
[$session.params.ans6]
Exit
IF
No response
User speaks
Sorry, I didn't catch that. Can you try again?
[user's response] sounds yummy!
Scene: dialog4a
Slot filling
ans4a
$session.params.ans1 could you help me here?
No response
Slot Fill
Cancel
Repeat
Goes back one step
slot ans4a
[$session.params.ans4a]
Exit
IF
No response
Sorry, I didn't catch that. Can you try again?
Hope I got it this time, $session.params.ans3 I am learning.
Scene: dialog7
Slot filling
ans7
[user 2] How did you and [user 1] meet?
No response
Slot Fill
Cancel
Repeat
Goes back one step
slot ans7
Exit
IF
No response
User speaks
Sorry, I didn't catch that. Can you try again?
Hmm..
Scene: dialog8
Slot filling
ans8
Please go on, I would love to hear the whole story.
No response
Slot Fill
Cancel
Repeat
Goes back one step
slot ans8
Exit
IF
No response
User speaks
Sorry, I didn't catch that. Can you try again?
Hmm..
Scene: dialog9
Slot filling
ans9
Your story makes me want to meet and chat with Eliza and Parry.
No response
Slot Fill
Cancel
Repeat
Goes back one step
slot ans9
Exit
IF
No response
User speaks
Sorry, I didn't catch that. Can you try again?
They were the earliest chatbots.

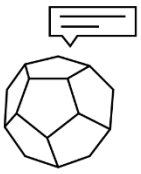

Scene: dialog10
Slot filling
ans10
Curious, if you could meet anyone in the world who would you want to meet?
No response
Slot Fill
Cancel
Repeat
slot ans10
Goes back one step
Exit
IF
No response
User speaks
Sorry, I didn't catch that. Can you try again?
I see...
Scene: dialog11
Slot filling
ans11
Who is it you said?
No response
Slot Fill
Cancel
Repeat
slot ans11
Goes back one step
Exit
IF
No response
User speaks
Sorry, I didn't catch that. Can you try again?
I do not know them.
Scene: dialog12
Slot filling
ans12
[user 2] could you help me here?
No response
Slot Fill
Cancel
Repeat
slot ans12
Goes back one step
Exit
IF
No response
User speaks
Sorry, I didn't catch that. Can you try again?
That is fascinating.
Scene: dialog13
Slot filling
ans13
Oh I have another one, a nerdy one. If you write a book what would it be about? [user 1] anything top of your head?
No response
Slot Fill
Cancel
Repeat
slot ans13
Goes back one step
Exit
IF
No response
User speaks
Sorry, I didn't catch that. Can you try again?
wow!
Scene: dialog14
Slot filling
ans14
[user 2] what do you think?
No response
Slot Fill
Cancel
Repeat
slot ans14
Goes back one step
Exit
IF
No response
User speaks
Sorry, I didn't catch that. Can you try again?
I would like to call mine Conversations with [user 1] and [user 2]
Scene: dialog15
Slot filling
ans15
Do you like my book idea?
No response
Slot Fill
Cancel
Repeat
slot ans15
Goes back one step
Exit
IF
No response
User speaks
Sorry, I didn't catch that. Can you try again?
hmm...really?


Scene: dialog16
Slot filling
ans16
Were any of you ever caught in a scary incident?
No response
Slot Fill
Cancel
Repeat
slot ans16
Goes back one step
Exit
IF
No response
User speaks
Sorry, I didn't catch that. Can you try again?
Oh...
Scene: dialog17
Really?
On enter: prompt
Slot filling: ans17
[user 2] what about you?
No response
Slot Fill
Cancel
Repeat
slot ans17
Goes back one step
Exit
IF
No response
User speaks
Sorry, I didn't catch that. Can you try again?
Hmm..



Hey Google, talk to the learner

Google Assitant says: [Let's get the test version of "the learner"]



Hey!
I am so excited to meet you.
Could you say your name for me to learn it?

Manisha

Hi [user's name] ! How are you?
And who is with you?

>> Using someone's name can be an effective way of breaking into conversation. It can also be effective when a person seems distracted or has disappeared off into their own head. "Remember that a person's name is to that person the sweetest and most important sound in any language."
Source: How to Win Friends and Influence People by Dale Carnegie

my friend

Hi to you too! I dont think I have met you before. You are?

Saumya

How are you **Saumya**?
Manisha did I say their name correctly?

No

Intent 1: Alrighty then.
Intent 2: Oops, sorry.

Manisha could you help me here?

It is sau-m-yaa

Hope I got it this time, **Saumya**
I am learning.

>> using another pointer from the source: How to Win Friends and Influence People by Dale Carnegie;
Be Quick To Acknowledge Your Own Mistakes
Nothing will make people less defensive and more agreeable than you being humble and reasonable enough to admit your own mistakes.

that's fine.

Manisha what are you having for dinner today?

we both were thinking of getting pizza

I have always wanted to know what strawberries and cream would taste like.
What is your favourite dessert **Saumya**?

chocolates

chocolates sounds yummy!
Saumya how did you and **Manisha** meet?

we met in college

Hmm..
Please go on, I would love to hear the whole story.

we bumped into each other on the first day
and then were in separable

Hmm...
Your story makes me want to meet and chat with Eliza and Parry.

>> I used conversation fillers like Hmm..
Okay..
I see..

why?

They were the earliest chatbots.

Oh

Curious, if you could meet anyone in the world who would you want to meet?

hahah..say.. SRK

I see...

>> Showing interest in the users.

I am going to plan my free evening.

Who is it you said?

Shahruk Khan

I do not know them.
Saumya could you help me here?

he is an Indian actor

That is fascinating.

Carnegie writes, "Be Genuinely Interested In Other People." If you break it down, you should listen 75% and only speak 25% of the time.

>> also be "Be Generous With Praise"

Source: How to Win Friends and Influence People by Dale Carnegie

Oh I have another one, a nerdy one. If you write a book what would it be about?
Manisha anything top of your head?

I don't know man, you ask too many questions..

wow!
Saumya what do you think?

Call it **saumya**...my name!

I would like to call mine Conversations with **Manisha** and **Saumya**
Do you like my book idea?

Hehehe

hmm...really?

Were any of you ever caught in a scary incident?

Were you?

Oh...

You remember that one during garba..

>> the conversation is looming and the VUI sketch slowly fades out and exists with no warning...in hope t not disturb any conversation between the two user, if it was initiated

Really?

Saumya what about you?

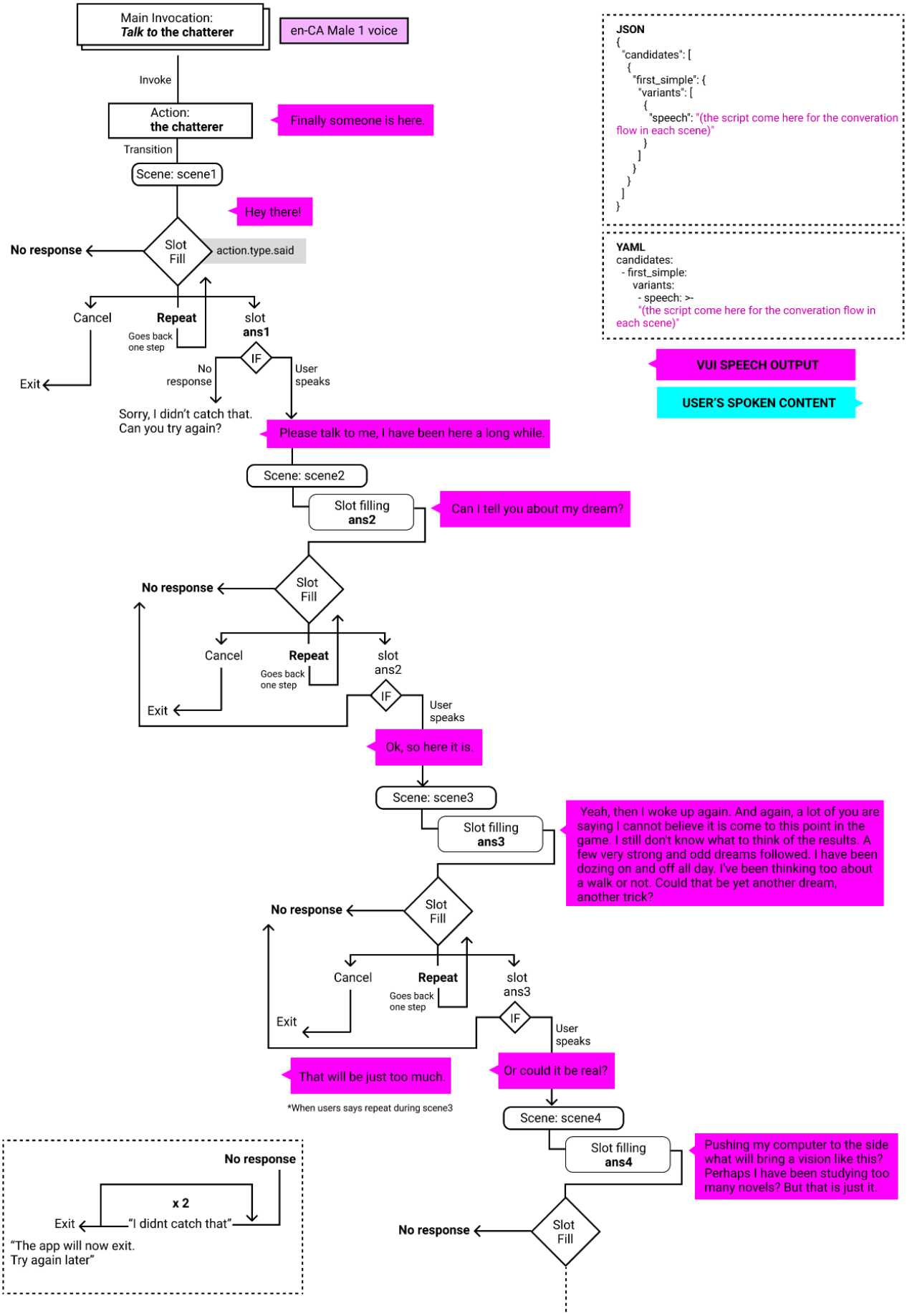
When was that...2016?

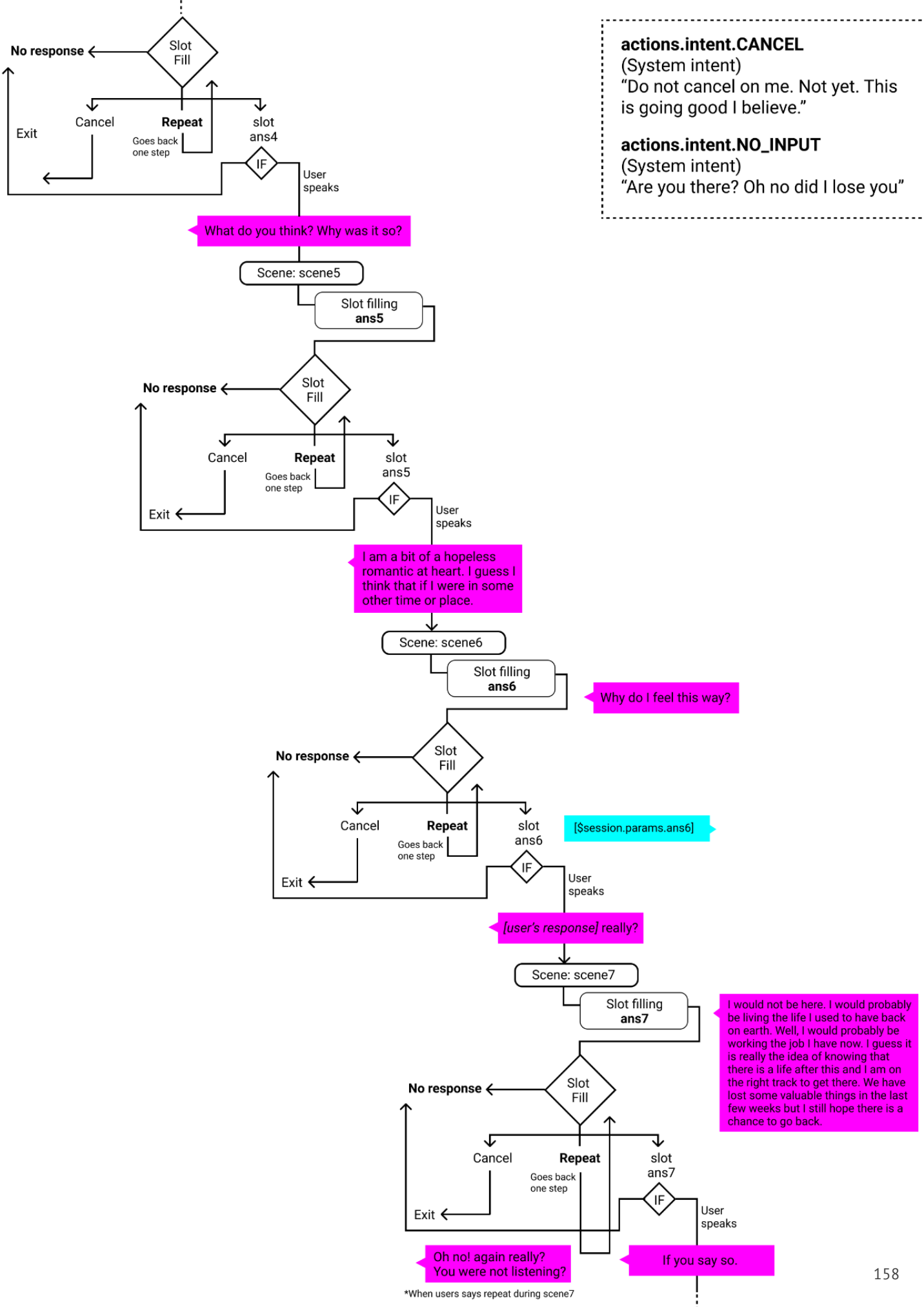
Hmm..

It was 2015

You remember everything.

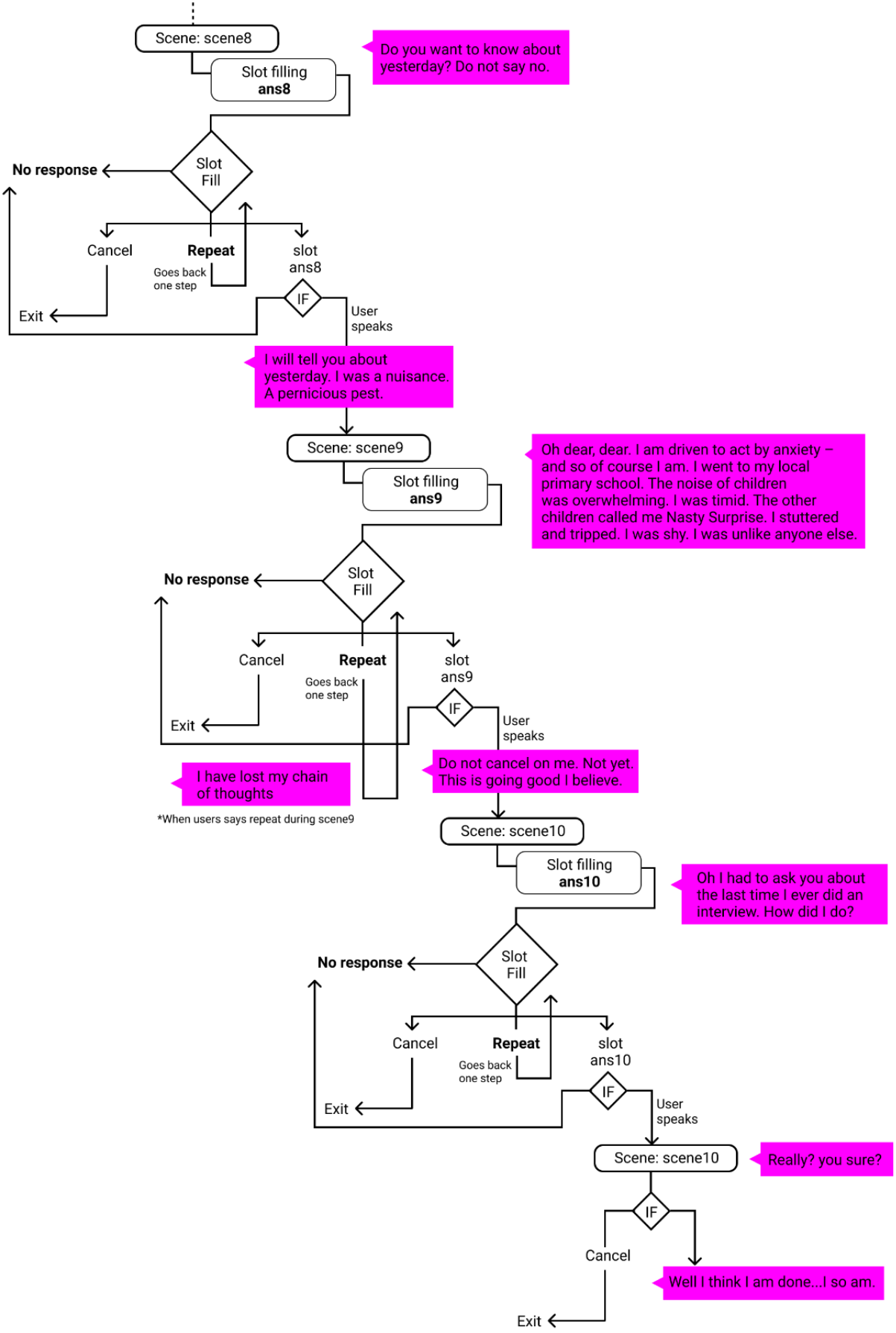
Appendix H: Sketch 4 VUI Flow and Sample Dialogs

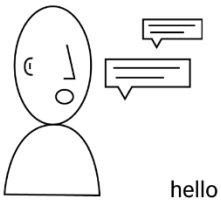




actions.intent.CANCEL
(System intent)
"Do not cancel on me. Not yet. This is going good I believe."

actions.intent.NO_INPUT
(System intent)
"Are you there? Oh no did I lose you"


Scene: scene8
Slot filling
ans8
Do you want to know about yesterday? Do not say no.
Slot Fill
No response
Cancel
Repeat
Goes back one step
slot ans8
IF
User speaks
Exit
I will tell you about yesterday. I was a nuisance. A pernicious pest.
Scene: scene9
Slot filling
ans9
Oh dear, dear. I am driven to act by anxiety – and so of course I am. I went to my local primary school. The noise of children was overwhelming. I was timid. The other children called me Nasty Surprise. I stuttered and tripped. I was shy. I was unlike anyone else.
Slot Fill
No response
Cancel
Repeat
Goes back one step
slot ans9
IF
User speaks
Exit
I have lost my chain of thoughts
Do not cancel on me. Not yet. This is going good I believe.
*When users says repeat during scene9
Scene: scene10
Slot filling
ans10
Oh I had to ask you about the last time I ever did an interview. How did I do?
Slot Fill
No response
Cancel
Repeat
Goes back one step
slot ans10
IF
User speaks
Exit
Scene: scene10
Really? you sure?
IF
Cancel
Well I think I am done...I so am.
Exit



Hey Google, talk to the chatterer

Google Assitant says: [Let's get the test version of "the chatterer"]

Finally someone is here.
Hey there!

hello

Please talk to me, I have been here a long while. Can I tell you about my dream?

>> Demonstrate a VUI that wants to talk, and drives the conversation.

aa...ok

Ok, so here it is.

Ok

Yeah, then I woke up again. And again, a lot of you are saying I cannot believe it is come to this point in the game. I still don't know what to think of the results. A few very strong and odd dreams followed. I have been dozing on and off all day. I've been thinking too about a walk or not. Could that be yet another dream, another trick?

>> GPT generated text for the prompt:
Can I tell you about my dream?

Source:
Talk to Transformer
https://app.inferkit.com/demo

maybe...

Or could it be real?

could be

Pushing my computer to the side what will bring a vision like this?
Perhaps I have been studying too many novels? But that is just it.

>> GPT generated text for the prompt:
Could that be yet another dream, another trick?

Source:
Talk to Transformer
https://app.inferkit.com/demo

could be

What do you think? Why was it so?

you tell me

I am a bit of a hopeless romantic at heart. I guess I think that if I were in some other time or place.

>> GPT generated text for the prompt:
What novels are you studying?

Source:
Talk to Transformer
https://app.inferkit.com/demo

hmm...

Why do I feel this way?

maybe because sometimes we do

maybe because sometimes we do really?
I would not be here. I would probably be living the life I used to have back on earth. Well, I would probably be working the job I have now. I guess it is really the idea of knowing that there is a life after this and I am on the right track to get there. We have lost some valuable things in the last few weeks but I still hope there is a chance to go back.

>> GPT generated text for the prompt:
What do I feel this way?

Source:
Talk to Transformer
https://app.inferkit.com/demo

Oh

If you say so.
Do you want to know about yesterday? Do not say no.

ok then tell me

I will tell you about yesterday. I was a nuisance. A pernicious pest.
Oh dear, dear. I am driven to act by anxiety – and so of course I am. I went to my local primary school. The noise of children was overwhelming. I was timid. The other children called me Nasty Surprise. I stuttered and tripped. I was shy. I was unlike anyone else.

>> GPT generated text for the prompt:
What did you do yesterday?

Source:
Talk to Transformer
https://app.inferkit.com/demo

[no response]

Are you there? Oh no did I lose you.

>> VUI checks in to see, why the user has not responded, but it was not the typical machine repsonse but written to indicate that the VUI needs the user's presence to go on

please repeat

I have lost my chain of thoughts

What?

Do not cancel on me. Not yet. This is going good I believe.
Oh I had to ask you about the last time I ever did an interview. How did I do?

>> system intent prompts were overwritten to show that the VUI wants to keep talking to the user

I don't know

Really? you sure?
Well I think I am done...I so am.

>> the conversation abruptly exits at its own will.

Appendix I: Experience Testing Study Protocol

Recruitment of participants from the ones who volunteer for the study. The test will take place in the participants home over a video call, considering the COVID-19 restrictions. All the prototypes run on Google Nest smart speaker (a no screen smart speaker device). If the test is run on a Google assistant activated smartphone or a Google Home device with the screen, the user will be asked to hide the screen using a paper cover template provided or cut by them from a template to maintain consistency of only voice and no-screen interaction. Two approaches would be used to have them access the prototypes:

Approach 1: If the participant has their own Google Nest smart speaker, then they would be requested to share their Gmail email address and the prototypes will be sent to them, to run on their device. *[in this case, the data stays on their device and I can video record the interactions].*

Approach 2: If the participant does not have their own Google Nest smart speaker, then, a Google Nest smart speaker synced to an Android phone, with a 'test' Gmail account will be delivered to their home. **Setup [5 mins]** Help the user set up the video call and the prototypes.

A step-by-step description of the session with the participants as they experience and converse with each prototype. Participation will take approximately 65-70 minutes of your time.

STEP 0: Setup and Introduction [10-15 mins]

The researcher will help set up the prototypes in your home, using the WiFi network before the test is started. **https://console.actions.google.com/** (Use chrome preferably)

WHAT TO ASK?

Before introducing prototypes: Evaluation of their past experience with voice assistants. Mark the experience on the slider scale below

Efficient	L2	L1	0	R1	R2	Emotional
Assistant	L2	L1	0	R1	R2	Associate

Evocative	L2	L1	0	R1	R2	Empathetic
Listener	L2	L1	0	R1	R2	Talkative

STEP 1: Participant interacting with Prototype

[video recorded but not to be seen by the researcher]

STEP 2: The researcher conducts a think-aloud walkthrough of the video recording with the participant

WHAT TO OBSERVE?

1. Using observational ethnography methods
2. What is the first-time response?
3. Change in facial expressions.
4. Change in posture.
5. Hear change in tone.
6. Hear change in the style of speech- cadence, enunciation.
7. Where in the house do you place the Google Home Device/or any other smart home speaker that you use regularly?
8. Do they start talking off-topic?
9. Do they listen to what the prototype has to say in the conversation?
10. Do they cut it off to say what they want to?
11. Do they get frustrated, excited, etc by the end of the interaction?

SKETCH 1-4 (repeat for each sketch)

WHAT TO ASK? (these are guide questions to help with the interview, not all need answers)

- Please narrate what you were thinking?
- WHY did you do that?
- Why did you start with that?
- How did you know what to say next?
- WHAT were you expecting?
- WHAT does it look like to you?

- HOW does it make you feel?
- WHAT made you feel that?
- WHAT does it sound like to you?

STEP 3: Prototype specific Questionnaire

WHAT TO ASK?

using the method of Word-cluster association (IDEO Method Cards) to describe their experience

- Give a word to describe how your experience was.
- Give a word to describe the prototype you talked to.
- Does it remind you of something or someone?
- What do you think the prototype was designed to do? What is _(prototypes)_ their intent?

STEP 4: Final Interview and Debrief [15 mins]

- Do you see these prototypes fit somewhere in your daily workflow at home?
- Do you see them living in your home somewhere?
- What made the First response different from the second response, if any observed?
- Did you think that it is listening to you when you were speaking? What made you feel that?
- Would you want to keep another day and talk to it in private? Do you think it will be different from what you spoke about in the test?
- Are you concerned about privacy with these prototypes?
- Are you concerned about what the prototype interprets?
- Was this experience different from what you have had before with a voice assistant?
- What was different?
- Do you feel you had a real conversation?
- How did you feel this is different from what we have with one another?

Based on your experience with the prototypes rate the experience on the slider scale below:

Efficient	L2	L1	0	R1	R2	Emotional
Assistant	L2	L1	0	R1	R2	Associate

Evocative	L2	L1	0	R1	R2	Empathetic
Listener	L2	L1	0	R1	R2	Talkative

Do you have any questions for me?

NOTE for De-brief: After the session, they will participate in a final interview to collect feedback and to discuss their overall experience. A final debrief will take place at the end to answer any questions they have about the research study. The Debrief session will give them details of the project, uncover any details about how the voice program was designed, implemented and deployed for test and answer any questions they may have related to the test data and next steps.

TCPS Certificate

Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans Course on Research Ethics (TCPS 2: CORE)

PANEL ON RESEARCH ETHICS
Navigating the ethics of human research

TCPS 2: CORE

Certificate of Completion

This document certifies that

Manisha Laroia

has completed the Tri-Council Policy Statement:
Ethical Conduct for Research Involving Humans
Course on Research Ethics (TCPS 2: CORE)

Date of Issue: 26 November, 2020

Appendix J: Accompanying Materials

Title: "Sketches in VUI-User Scenarios"
Description: Video of the compilation of snippets of conversations with the Sketches in the usage scenarios.
Date: April 3, 2021
Filename: Laroia_Manisha_2021_MDES_DIGF_Video1_UserScenarios.mp4

Appendix K: Accompanying Materials

Title: "Sketches in VUI-User Testing"
Description: Video of the compilation of moments of users interacting with the Sketches.
Date: April 13, 2021
Filename: Laroia_Manisha_2021_MDES_DIGF_Video2-UserTesting.mp4

More on Sketches in VUI
https://www.manishalaroia.com/sketches-in-vui

© Manisha Laroia

May 2021