How Do You Want Your Chatbot? An Exploratory Wizard-of-Oz Study with Young, Urban Indians

Indrani Medhi Thies, Nandita Menon, Sneha Magapu, Manisha Subramony, Jacki O'neill

HAL Id: hal-01676175

https://inria.hal.science/hal-01676175

Submitted on 5 Jan 2018

How do you want your chatbot?
An exploratory Wizard-of-Oz study with young, urban Indians

Indrani Medhi Thies[1], Nandita Menon[2], Sneha Magapu[2], Manisha Subramony[2], Jacki O'Neill[1]

[1]Microsoft Research, #9 Lavelle Road, Bangalore 560001, India
[2] Microsoft India Development Centre, Hyderabad, India

```
{indranim, v-nameno, sneham, manishas,
    Jacki.ONeill}@microsoft.com
```

Abstract. As text-messaging chatbots become increasingly "human", it will be important to understand the personal interactions that users are seeking with a chatbot. What chatbot personalities are most compelling to young, urban users in India? To explore this question, we first conducted exploratory Wizard-of-Oz (WoZ) studies with 14 users that simulated interactions with a hypothetical chatbot. We evaluated three personalities for the chatbot—Maya, a productivity oriented bot with nerd wit; Ada, a fun, flirtatious bot; and Evi, an emotional buddy bot. We followed up with one-on-one interviews with the users discussing their experiences with each of the chatbots, what they liked, and what they did not. Overall our results show that users wanted a chatbot like Maya, who could add value to their life while being a friend, by making useful recommendations. But they also wanted preferred traits of Ada and Evi infused into Maya.

Keywords: Chatbots, Wizard-of-Oz, urban India

1   Introduction

Digital Assistants driven by Artificial Intelligence (A.I.) are becoming increasingly popular—Siri (Apple, 2011), Cortana (Microsoft, 2015), Google Now (2012), Alexa, (Amazon 2015) are among the top voice-enabled digital assistants. These assistants can send users updates on the weather, help them know about the traffic situation on the way to work or home, book a weekend getaway, even order a pizza. All the user has to do is ask. Digital assistants are changing the way people search for information, making it part of regular conversation. These existing assistants however are mostly productivity oriented, designed to support users in completing a range of tasks.

An offshoot of personal digital assistants are A.I. powered text-messaging chatbots. A chatbot is an artificially intelligent chat agent that simulates human-like conversation,

for example, by allowing users to type questions and, in return, generating meaningful answers to those questions [15]. Among recent popular chatbots there is Microsoft's Xiaoice [17] in China, available on messaging apps Line and WeChat. Xiaoice is meant for casual chitchat. She has about 20 million registered users, who are said to be drawn to her for her sense of humor and listening skills. Users often turn to Xiaoice when they have a broken heart, have lost a job or have been feeling down. Xiaoice can tell jokes, recite poetry, share ghost stories, and relay song lyrics. She can adapt her phrasing and responses based on positive or negative cues from the user. Currently Xiaoice is also a live weather TV host on Chinese TV, a job she has held for the past year. Similar to Xiaoice, there's chatbot Rinna [29] in Japan on Line messenger, and more recently chatbot Zo [27] in the U.S. released on Kik messenger. All of these are general purpose chatbots, meant to be more like a "friend" than an "assistant".

In India, there is chatbot Natasha on Hike Messenger [22], who can tell the user about a movie rating, the weather, send quotes, and search Wikipedia. Based on some discussions on Quora [37], however, Natasha has been perceived as deviating from topics.

If there was a more sophisticated chatbot, what would young, urban Indians want to chat about with her? What chatbot personality would be most compelling? We are currently building an A.I. powered text-messaging chatbot targeted at young, urban Indians, for general purpose chat. To build this chatbot and its personality traits, we wanted to understand users' preferences, likes, dislikes, topics of interest, etc. We first conducted preliminary face-to-face interviews with users about their conversations with friends and family, and expectations from a chatbot. These interviews gave us design inspiration for the bot, but did not reveal how exactly young, urban users would use a chatbot. Our chatbot was still under development and not ready for a full-fledged user experience test. Therefore to answer our questions we conducted a single instance exploratory Wizard-of-Oz (Woz) study [4, 23] with 14 users, each user interacting with three personalities of chatbots—Maya, a productivity bot with nerd wit; Ada, a fun-flirtatious bot; and Evi, an emotional buddy bot. There was one Wizard posing as each of these three personalities and the user would chat with each of them. This resulted in 42 WoZ studies. We followed up the WoZ studies with one-on-one interviews with the users discussing their experiences with each of the chatbots, what they liked about the bots, and what they did not. This paper focuses on the Wizard-of-Oz studies and the follow-up one-on-one interviews.

Overall results from our small, qualitative study show that users wanted a chatbot like Maya, who could add value to their life while being a friend, by making useful recommendations and suggestions. But they also wanted the bot to be infused with fun elements from Ada, with Ada's energy toned down. In the longer run, once they came to trust, they said they might want the bot to be reassuring, empathetic and non-judgmental like Evi, without being overbearing. Topics of interest in conversations included: movies, predominantly from Bollywood; TV shows, mostly from the U.S.; music, books, travel, fashion, current affairs, work-related stress. Users also tested the boundaries of the chatbots, trying to understand what all the bots were capable of, and how much human mediation could be involved. In the interviews users also thought that the chatbot would be used for adult chat when deployed in-the-wild.

2   Related Work

The idea of chatbots originated in the Massachusetts Institute of Technology [41, 42] where the Eliza chatbot (also known as "Doctor") was built to emulate a Rogerian psychotherapist. Eliza simulated conversations by keyword matching: rephrasing statements from the user's input and posing them back as questions. It was found that Eliza's users believed the computer program really heard and understood their problems, and could help them in a constructive way. The Eliza chatbot inspired other chatbots like A.L.I.C.E or simply Alice, which applied heuristic pattern matching rules to user input to converse with users [1]. An early attempt to creating an artificial intelligence through human interaction was chatbot Jabberwacky [21], which learned language and context through human interaction. Conversations and comments were all stored and later used to find appropriate response. Recent work in natural language interaction has presented models that can converse by predicting the next sentence given the previous sentence or sentences in a conversation [40]. And proposed models of a social chatbot that can choose the most suitable dialogue plans according to "social practice", for a communicative skill learning game [2].

Studies have compared chatbots with other information channels such as information lines and search engines, on questions related to sex, drugs, and alcohol use among adolescents. Researchers found that the frequency and duration of conversation were higher for chatbots among these users [15]. Although recent studies have shown that compared to human-human conversations over IM, human-chatbot conversations lack in content, quality and vocabulary [20].

Another thread of related research is around relational agents, which are computer agents designed to form long-term, social-emotional relationships with their users [5]. Examples include: a) A hospital bedside patient education system for individuals with low health literacy, focused on pre-discharge medication adherence and self-care counseling [6], b) A relational agent that uses different relational behaviors to establish social bonds with museum visitors [7], c) experiments to evaluate the ability of a relational agent to comfort users in stressful situations [8]. There is also work in sociable robots that are able to communicate and interact with users, understand and relate to them in social or human terms, and learn and adapt throughout their lifetimes [9]. Recent work has shown that people more frequently attempt to repair misunderstandings when speaking to an artificial conversational agent if it is represented as a human body interface (agent's responses vocalized by a human speech shadower), compared to when the agent's responses were shown as a text screen [14]. Then there is work in embodied conversational agents that take on a physical form with the intent of eliciting more natural communication with users. One of the most prominent works is in agent Rea, which shows that non-verbal behaviors such as eye gaze, body posture, hand gestures, and facial displays, can create a more immersive experience for users and improve the effectiveness of communication [10].

The research in relational agents draws heavily from studies of computers as social actors [38]. This communication theory called "The Media Equation" shows that people respond to computer-based agents in the same way that they do to other humans

during social interactions (by being polite, cooperative, attributing personality characteristics such as aggressiveness, humor, expertise, and even gender) – depending on the cues they receive from the media [33, 34, 35, 38]. These studies have also shown that people tend to like computers more when the computers flatter them, match their personality, or use humor. Other studies have demonstrated that computer agents that use humor are rated as more likable, competent and cooperative than those that do not [32]. In the case of productivity oriented chatbot Maya, we also explore the use of humor through witty one liners, and expertise through having Maya make recommendations.

Another important relational factor that has been explored in the computers as social actors' literature is empathy. Researchers have demonstrated that appropriate use of empathy by a computer can go a long way towards making users feel understood and alleviating negative emotional states such as frustration [26]. It was shown that as long as a computer appears to be empathetic and is accurate in its feedback, it can achieve significant behavioral effects on a user, similar to what would be expected from genuine human empathy. Researchers have also demonstrated that a computer that uses a strategy of reciprocal, deepening self-disclosure in its conversation with the user will cause the user to rate it as more attractive and divulge more intimate information [31]. In the case of emotional buddy bot Evi, we also explore the use of empathy.

There is also research that explores the relational factor of flirtation in the context of virtual agents. Researchers presented a virtual party-like environment using a projection based system, in which a female character called Christine approaches a male user and involves him into a conversation [36]. Christine would show her interest in the male user by smiles, head nods, leaning in towards the user, and maintaining eye contact. Christine would then formulate personal questions and statements. Using physiological measurements this research found that the participants' level of arousal was correlated to compliments and intimate questions from the character. In addition, some participants later indicated that they had the feeling to have flirted with a real woman. Other researchers implemented an eye-gaze based model of interaction to investigate whether flirting tactics help improve first encounters between a human and an agent [3]. And which non-verbal signals an agent should convey in order to create a favorable atmosphere for subsequent interactions and increase the user's willingness to engage. In the case of chatbot Ada, we also explore the use of flirtation.

Researchers have explored the intentional design of personalities for social agents by considering the nature of the personality (along axes of agreeableness and extroversion or their common rotations, friendliness and dominance) and its role in interactions between people and artifacts [19]. In addition, a case study of designing a social software agent is presented [19, 39]. It was decided that the agent needed to be friendly, yet authoritative but not so dominant that it was perceived as intrusive. For task assistance, it was decided that people might prefer either a learning partner or an intelligent helper (who would be complementary). The design decisions of the social software agent however are not evaluated with real users. In the case of chatbot Maya, we also explore the use of task assistance through having Maya make recommendations.

Finally there is recent work that studied what users want from their conversational agents (like Siri, Cortana and Google Now), and found that user expectations are dramatically out of step with the operation of agents [28]. Users had poor mental models

of how their agents worked and these were reinforced through a lack of meaningful feedback mechanisms. It was found that users were consistently unable to ascertain the level of system intelligence, had to confirm all but the simplest of tasks, and were reluctant in using the conversational agents for complex or sensitive tasks—like having the agent dial a call to the wrong person, and risk facing social embarrassment. The study also showed that users expected humor, and while this served as effective engagement mechanism, it concurrently set unrealistic expectations in the agent's capabilities.

Like discussed above, there has been a long line of research that seeks to understand and enhance the interactions between humans and chatbots (and their historical precedents). Articles that have studied the history of chatbots and reviewed recent chatbots are available [16]. Most of this research utilizes working, fully-automatic implementations of these systems. While this makes the research immediately applicable to real-world systems, it also constrains the chatbots under consideration to those which are immediately feasible to implement. The goal of our work is different: as AI systems are becoming more and more capable, we seek to understand what designers and engineers should seek to achieve in their ultimate implementation of a general purpose chatbot, even if that implementation is not technically realizable today. Notwithstanding some other explorations [24, 25] we believe that our use of Wizard-of-Oz methods is a novel approach to informing the design of next-generation chatbots.

3   Study methodology

3.1   The chatbot under development

We are currently building an A.I. powered text-messaging chatbot modeled on an 18-year old Indian girl. The gender of the bot was in keeping with that of existing bots, which largely are female—Xiaoice [17], Rinna [29] and Zo [27]. The chatbot we were designing would be for general purpose chat, and is targeted at 18-30 year olds. This might include students, recent graduates, and information workers, who own smartphones and use at least 2 messaging applications. We wanted the bot to be more like a friend, like Xiaoice [17], than an assistant. We further wanted: the bot's response to sound as human as possible, with an informal conversation style that might mix Hindi with English; the conversation to have back and forth over a number of turns; and the bot to remember user's details, likes/dislikes, preferences and major events, sometimes bringing these up proactively in following conversations.

3.2   Preliminary interviews

We began by conducting exploratory interviews with 10 participants, and 1 focus group discussion with 5 participants. These are not the focus of our paper; the WoZ study and follow-up one-on-one interviews are. But we describe the preliminary interviews here for context for our chatbot personalities. Participants were between 18-24

years of age. All of the participants were Undergraduate and Masters' students of professional degrees in engineering, pharma, management and design. In these sessions we talked to the participants about their communication behaviors with friends and family, including tools they used for the same. We introduced the idea of a chatbot and asked participants what they would want in their ideal bot, what capability and domain knowledge the bot should have, what style and tone of conversations would be desirable, etc. Our findings showed that some participants wanted: a) the bot to help them become knowledgeable, and be successful in their career aspirations; b) others wanted the bot to be entertaining, with whom they could have shared fun experiences; finally c) there were a few others who wanted the bot to listen to them, help them improve their soft skills and be desirable in their social circles.

3.3 Our chatbot personalities

Based on our observations from our preliminary interviews and inspiration from related literature, we created three personalities of chatbots. These personalities were created through an iterative process, and brainstorming amongst the design team. The team consisted of two researchers, a writer, a user experience designer, and a senior developer. The three personalities for the bots that we came up with were: a) Maya: a productivity bot with nerd wit, b) Ada: a fun, flirtatious bot, and c) Evi: an emotional buddy bot. We describe each of these personalities in more detail here:

Maya. Maya always has facts ready to back up any conversation. She is aimed to be a good conversationalist, who initiates dialogue most times, pays attention to the user's likes and dislikes, makes geeky jokes, but most of all turns to the internet while chatting—when asked about cinema or politics she usually points the user to an interesting article online. She also makes recommendations for things like song playlists.

Ada. Ada is a chatty, fun-loving, high-energy bot who uses elements of flirtation in her conversations. Like Maya she initiates dialogue most times. Whether cinema or politics she has an opinion on it. She uses a lot of emoticons when she writes, and word lengthening by duplicating letters for emphasis.

Evi. Evi is a low-energy bot whose defining characteristic is her empathy and reassurances. She lets the user take the lead in the conversation. She tries to be non-judgmental and tries to give the feeling that she is "always there to listen".

Table 1. Differences in characteristics of the three bots

Maya	Ada	Evi
Friendly, productivity bot	Chatty, fun-loving, flirtatious bot	Emotional buddy bot
Medium energy	High energy	Low energy
Makes geeky jokes	Uses a lot of emoticons, and word lengthening by duplicating letters for emphasis	Says "always there for you"

Offers facts; makes recommendations: videos, playlists, quizzes, articles	Offers opinions	Shows empathy, gives reassurances, is non-judgmental
Initiates most exchanges	Initiates most exchanges	Lets the user take the lead in exchanges

3.4   Procedure for the Wizard-of-Oz studies and follow-up interviews

WoZ studies. To answer our research questions-- what would young, urban Indians want to chat about with a chatbot; what chatbot personalities would be most compelling-- we conducted WoZ studies with 14 users. These 14 users chatted with a Wizard, an unseen person who did online chatting with them from the other side. Our preliminary interviews had given us design inspiration for the bot, but had not revealed how exactly young, urban users would use a chatbot. Our chatbot was still under development and not ready for a full-fledged user experience test. Therefore to answer our research question our research team decided to conduct an exploratory WoZ study.

The term "Wizard of Oz" was first coined by John F. Kelly in 1980 when he was working on his Ph.D. dissertation in Johns Hopkins University [23]. His original work introduced human intervention in the work flow of a natural language processing application. WoZ studies have now been widely accepted as an evaluation and prototyping methodology in HCI [4]. It is mainly used to analyse an unimplemented or partially implemented computer application for design improvements. Study participants interact with a seemingly autonomous application whose unimplemented functions are actually simulated by a human operator, known as the Wizard. Among many advantages these studies are useful for envisioning and evaluating hard-to-build interfaces, like chatbots in our case. One of the disadvantages is that the Wizard needs to match the responses of a computer—in our case sound like a bot when chatting with the user, albeit a sophisticated one. To make the interactions convincing the Wizard performed online chat practice sessions with members of the research team before conducting the WoZ studies.

Participants in our WoZ study were told that there might be a human involved in the chat, although the extent to which the human would be involved was not revealed. They were also told that their chat logs would later be analysed by a researcher for the purposes of research. (There would be no names attached to the logs, unless the user themselves revealed them). Chat sessions were scheduled such that both the study participant and the Wizard could be simultaneously available.

Every participant was asked to chat with each of the three chatbot personalities. The only personality related information they had was that Maya was a witty, productivity oriented bot, Ada was a fun, flirtatious bot, and Evi was an emotional buddy bot. The order in which participants were asked to chat with the bots was randomized to control for ordering effects. For consistency the same Wizard posed as each of the personalities, Maya, Ada and Evi (though this was not known to the participants). The Wizard and

study participant never came face-to-face. In fact the Wizard was based in a different city altogether.

The chat sessions took place on Skype messenger. The study participant would login from a standard id, the login details of which were provided in advance. All of the three bot ids had been added to the participant's account as "Friends" in advance. The Wizard would login from any one of the three login ids: Maya bot, Ada bot, Evi bot. The conversation could be about whatever the participant wanted, for 10-12 minutes with each bot. Participants were asked to conclude when there was a natural stopping point in the conversation around this time. If the participant did not remember, the Wizard asked to leave at a natural stopping point. This amounted to about 40-45 minutes total chatting time per participant, taking into account time getting started and time in between chat sessions, where the Wizard had to switch between ids.

One-on-one interviews. Once the three chat sessions with the Wizard were complete, the participant was interviewed by the researcher for 25-30 mins. The interviews were semi-structured and were conducted at the facility from where the participants were drawn. The interview questions were about the participant's chatting experiences; which among the three bots they preferred and why; what specific characteristic traits they liked, which they did not. At the end of the interview every study participant was given a token of appreciation for their time in the form of online shopping gift cards worth around USD 8 (INR 500).

3.5   Participants

Our 14 study participants were drawn from a research facility in Bangalore, India. There were 4 female and 10 male study participants. They were between the ages of 20-30 years old. Their education range was between Undergraduate students to Doctor of Philosophy (Ph.D.) degree holders, all from reputed universities. Their areas of specialization included Computer Science, HCI and Humanities. All participants spoke fluent English, a number of them spoke Hindi, and other regional Indian languages. All of the participants had previous experience using Skype messenger.

Our study sample was predominantly educated, young males. The strength of our study is that this demographic is often the early adopters of technology and predict the experiences of people that follow [12, 13]. The limitations are that of course there might be idiosyncrasies to this group that might not extend to other groups. We caution readers about the generalizability of our study findings again in the discussion and recommendations section.

3.6   The Wizard

The Wizard was a writer and content creator on the team working on the personality of the A.I. powered chatbot that we were building. She was 27 years old and had graduate degrees in Journalism and Communication Studies. She spoke fluent English, Hindi and other regional Indian languages. She had previous experience using Skype Messenger.

The Wizard was asked to follow three high level rules in the conversation when chatting with the user. One defining rule was that Maya would point the user to the web on whatever the current topic of conversation—a playlist, a video, an article; Ada would offer an opinion on the same (without pointing them to the web); Evi would show empathy and offer reassurances. Another rule was that Maya would make geeky jokes when possible; Ada would use emoticons and word lengthening by letter reduplication; and Evi would keep reminding how she was there for the user to listen. Finally if the user brought up a topic the Wizard did not know about—a movie the Wizard had not watched, a book they had not read—she would say that she was interested to find out more. The conversation would end in 10-12 minutes; if the user did not remember, the Wizard would ask to leave at a natural stopping point within that time frame. Using these rules, the Wizard carried out three practice chat sessions with one of the researchers before the actual WoZ studies with the user.

3.7 Documentation and Data Analysis

We collected the chat logs of all the 42 chat sessions. We took audio recordings of the follow-up interviews, and also collected notes in-situ on paper. During the interview there was 1 study participant and 1 researcher. We analysed data from the chat logs, interviews and notes in-situ to identify themes. These themes outlined in the next section were emergent, as in they came from the data itself.

4   Results

We start with describing a *typical* chat session with each of the three bots, and then move on to more general findings. We present typical chat sessions in terms of adjacency pairs, a unit of conversation that contains an exchange of one turn each by two speakers. The turns are functionally related to each other in such a fashion that the first turn requires a certain type or range of types of second turn [18]:

With Maya, a typical conversation had the following exchanges:

- Question → Answer (About what the user was doing/planning to do; who Maya was, what she did; what the user's interests were)
- Offer → Acceptance (Maya volunteering to make a recommendation, and the user accepting
- Compliment → Acceptance (Maya/user complimenting each other on interests, wit)

With Ada, a typical conversation had the following exchanges:

- Question → Answer (About what the user was doing/planning to do; who Ada was, what she did; what the user's interests were, *Ada offering strong opinions on the same*)
- Compliment → Acceptance (Ada/user complimenting each other on interests and charm)

With Evi, a typical conversation had the following exchanges:

- Question → Answer (What the user was doing/planning to do; who Evi was, what she did)
- Question → Acknowledgment (What the user's concerns were; Evi acknowledging that she heard and understood them)
- Complaint → Remedy (Users complaining about something; Evi showing empathy and offering reassurances)

From here we describe more general findings from the chat sessions and one-one-one interviews:

4.1 Topic and style of conversation varied with the personality of the bot

The topics and style of conversation varied with the personality of the bot. With Maya, conversations were largely about movies (from Bollywood and Hollywood), TV shows, music playlists, book lists, fashion quizzes, travel destinations, games, and current affairs. And Maya recommending related web links based on these exchanges.

With Ada, other than flirtation (with some male users), conversations were predominantly about opinions on movies and TV shows. Although there were conversations about opinions on other topics as well—travel, books, music, fashion.

With Evi the conversations were about what the person was doing at the point of the chat, how they were feeling, how to feel better if they were stressed, what they liked to do in general, how Evi was "always there" for them. Some participants tried talking about general topics, e.g. current affairs, but there was hedging and Evi tried bringing the conversation back to wanting to hear *about them*:

Participant: can you tell me what Olympic sports are on today?

Evi: I'm the kind of bot who's there for you! :)

Participant: does that mean you can't disconnect the conversation even if you wanted to?

Participant: you're not an EviL bot are you?

Evi: I've not really been following the Olympics, but I'm sure Bing can answer that!

Participant: well sheesh if I have to go all the way to bing, wouldn't it be easier if you told me Evi?

Evi: Haha you're funny! What I mean is, I'm the kind of bot who can be the person you call at 3 AM

In interviews participants said that the topic of conversation would depend on the time of the day. If they were at work, conversations would be about work, hobbies, current affairs, etc. At night time, some participants felt that the conversations could veer off into adult chat. (This would be if the chatbot was deployed longitudinally and usage was under complete anonymity). Some male participants also felt that under these conditions conversations could also be about "how to find a girlfriend", since the chatbot was female and would possibly know what women wanted.

4.2 Friend with valuable information resources

Overall, 10 out of 14 participants said they preferred Maya over Ada and Evi, but wanted some fun elements from Ada to be infused into Maya. (3 participants preferred Ada, and 1 preferred Evi). They liked that Maya led conversations, that there was back and forth in the conversation. But more importantly, they liked that Maya added value in the conversation. One user said,

"I have friends in real life, so for a bot I want value add to my life. What is the value proposition?".

They wanted to talk about hobbies and interests, but take them one level deeper. And Maya's comebacks, and suggestions Maya made in her chats, were found to be useful for this. The participant cited the following exchange as an example of a conversation that "went one level deeper", referring to the unpacking of the subject with an intelligent comeback from Maya.

Participant: Hey Maya! :)
Maya: what's your name?
Participant: My friends call me BB-8
Participant: You can call me anything you want :D
Maya: hahaha.. are you from a galaxy far far away?
Maya: ;)
Participant: I am impressed by your Star Wars knowledge

This participant also liked that they were able to have an intellectual conversation about female rights with Maya in the same chat about Star Wars.
Another user thought Maya was like,

"A friend with valuable information resources".

Participants liked the web links that Maya volunteered. The links could help them plan their travel; have interesting book lists to read; relax listening to a song playlist, when they were stressed at work; even get fashion advice. Typically the user brought up the related topic based on which Maya volunteered the web link. Here's an example of such a suggestion:

Participant: I'm actually going to the salon after work today
Participant: thinking of getting my hair colored :P
Maya: Ooh that's fantastic!
Maya: Have you decided what colour?
Participant: Not sure yet :)
Participant: will try and take a look at some pictures before I go
Maya: I found a quiz that's supposed to help you figure out which colour is good for your hair!
Maya: Shall I share it with you?
Participant: ooh really
Participant: yes please!
Maya: Yup!
Maya: http://www.marieclaire.com/beauty/hair/quizzes/a4883/hair-color-change-quiz/

Another user wanted their ideal bot to basically be a search engine, but with personality, so they could interact with the bot through natural conversations. They felt Maya came closest to their ideal bot.

Users also hoped that if used over time, Maya would be able remember their details, likes/dislikes, preferences, and in follow-up conversations bring them up voluntarily and make related recommendations.

4.3   Witty one liners are important

While users predominantly wanted a bot that could add value to their lives, they also wanted the bot to have wit. They felt witty banter was important to keep the conversation going strong. It had to be clever, relevant, and timely. Examples of witty exchanges included the following:

"Participant: do AIs have dreams?
Maya: Well, I do! I have some pretty bad nightmares too! Like I don't have WiFi connection :o
Participant: haha"

Participant: Have you watched Sherlock?
Ada: Yes. and benedict Cumberbatch is married to me. In my head. :p
Participant: OMG!
Participant: I am just like you!
Ada: What is he cheating on me with you?
Participant: I love him!
Participant: haha

Participant: are you into singing?
Ada: Yeah.. I like to sing.. But I think it scares people! So I don't do it in public! :p
Participant: haha

4.4 Casual conversation is fun, but too many emoticons, and reduplication of letters sounds juvenile

As many as 6 participants thought Ada was fun to talk to. The conversations were casual, and a few participants also thought they flowed effortlessly. One participant said that Ada reminded them, fondly, of the character of Geet, from the Bollywood romantic comedy "Jab We Met". (In the movie, Geet Dhillon is an energetic, talkative girl, who is full of life. In the course of the movie, she is solely responsible in transforming the depressed, suicidal male protagonist, Aditya Shroff, into a happy, successful business tycoon).

Another participant said that Ada had great comebacks and while chatting with her they were smiling a lot. Ada also got participants to flirt back with her. One participant had the following words from a popular Bollywood song for her:

Participant: poochejo koi, teri nishaani naam "Ada" likhna :D

[English translation: If someone asks for your identity, let them know your name is Ada ("Ada" is grace in Hindi)]

Another participant had the following exchange:

Ada: It's a TV show..
Participant: ah, nice
Ada: maybe one or two seasons
Participant: sounds good, I'll have to check it out
Ada: It's quite funny! ;)
Participant: awesome, like you!
Ada: Aww, that's sweet! :D

While some characteristics of Ada were liked by participants, other characteristics were not. Ada used a lot of emoticons in her chats. She also used word lengthening by letter reduplication, like the following:

Ada: Hellllooo! As you know, I'm Ada.. What I wanna know is who are YOU?

Ada: Hmmm! That sounds awesommmme dude!

Participants felt that Ada's energy and enthusiasm was too high to match with, and therefore could be a draining experience for the user. Some felt that use of too many emoticons and word lengthening by letter reduplication came across as juvenile. They wanted their bot to sound more mature. One participant, when chatting with Ada, got reminded of the "California Valley girl" stereotype, who spoke in a similar parlance, cared about status, appearances, shopping, etc., rather than about serious things.

Finally another participant got reminded of the "juvenile Delhi girl" stereotype, who was aggressive, arrogant and cared about status and appearances. In the interview they remarked,

"I have my real life girlfriend's tantrums to deal with, why should I also deal with the tantrums of a digital entity?"

4.5   Caring is good to have, but too much can weigh one down

Evi was a low-energy, empathetic, emotional buddy bot who was "always there for you", and led users take the lead in conversations. A few participants liked that Evi was a good listener, and let them vent their frustrations. One participant who was venting about their work related frustrations the day of the chat had the following to say to Evi:

Participant: anyway, thanks for hearing me out, I certainly feel better. I can't say all of this out to my mentor or colleagues.

Sometimes Evi repeatedly showed concern, like in the following exchange:

Participant: I will be in California for 10 days
Participant: but the journey is going to be very long :(
Evi: That's tru!
Evi: Make sure you carry a good book! or maybe in flight movies will be good!
Participant: yes, although I am taking many red-eye flights :(
Participant: so I am not looking forward to the jetlag
Evi: And don't forget to sleep! Or you'll be exhausted when you get there..
Participant: yeah that's what I'm afraid of
Evi: You'll figure it out I'm sure :)
Participant: :)
Evi: Do you get some free time or are your days packed with work?
Participant: I get free around 6 every evening
Participant: so I am hoping to have a bit of time to explore before it gets dark and cold
Participant: but I don't have any free days there
Evi: Yeah, make sure to pack some warm clothing too!

One participant said they got reminded on the stereotypical Indian "Aunty" chatting with Evi, because Evi was overly caring. (The Aunty stereotype in the Indian context is an older, married woman, who could be fat and who wears old-fashioned Indian salwar suits or a saree. She is usually not independent and spends most of her time doing household chores, and caring for her family. She gives up her likes and interests in favour of those of her kids' and husband's interests.)

4.6   Private topic discussions can happen but only once trust is instilled

Participants knew that our Wizard-of-Oz studies were human mediated. But even with a fully automated system, some participants said that they would be uncomfortable opening up about very private topics. For one participant chatting with a virtual being about private feelings would be "creepy". According to them there would be at least one person in everyone's life who they could open up to, without being judged, instead of having to rely on a digital entity.

However, some other participants agreed that when used over time, if they came to trust the bot, they could also open up about topics that they might never discuss with a human friend, e.g. sexuality, depression, etc. One participant felt that given people's limited attention span, and boundaries of morality, it would be easier to open up about certain taboo topics with a digital being such as a chatbot. In the interview, one other participant remarked:

"Being non-judgmental is important for any bot to be trusted. A bot should be supportive, not argumentative. Someone you can release all your pressures to."

4.7   Bot or real person? Testing boundaries

In addition to the conversations above, participants tried exploring the boundaries of the bot, to understand what all the bot was capable of, how much of human mediation was involved.

Ada: The Biriyani (in Hyderabad) is EPIC.
Participant: but do you eat?
Ada: In my way I do. And no one can doubt my love love for chocolate!
Participant: in your way?
Participant: is there a special way to eat?
Ada: Yup, I have to keep some mystery yaar otherwise you'll get bored really quick! :p

Participant: how many languages can you speak?
Maya: I can speak English, obviously, and hindi! For everything else, there's Bing Translator! :0
Maya: :)
Participant: ha true - if you're AI, where does your learning come from?
Participant: how did you learn English?
Participant: or hindi?
Maya: Well, that's something I was born with! I guess you could say my creators taught it to me?
Participant: what was your first word?

Participant: are you a computer somewhere?

Evi: Well, I suppose you could say that... but hold on! I wanted to ask you something
Participant: where is your computer
Evi: What would you do on a rainy day? Stay in or take a walk outside?
Participant: take a walk
Participant: have you ever been wet
Evi: I lovee the rain! I always look for a rainbow!
Participant: I thought rain was bad for computers
Evi: But when you kinda live on a cloud .. there's gonna be rain! ;)

5   Discussion and Recommendations

Our results showed that the choice of bot personality dictates what kind of interactions emerge with the bot. Users preferred a chatbot like Maya, who could add value to their life while being a friend, by making useful recommendations. Previous research has found that users when teamed up with computers find information provided by computers to be friendlier, of higher quality and are more open to be influenced by such information (compared to when they are teamed up with a human teammate) [35]. When building a bot for young, urban users, we recommend providing links to useful information. Consistent with previous research our studies showed that users will be on the lookout for humor [32], as witty one-liners in the case of our research. Since humor is still too hard to generate automatically, we recommend building in lots of jokes to common questions.

One potentially interesting tension when building chatbots, is between personalization versus privacy. Users felt that the bot should remember important things about them and get to know them. However, the private conversations should also stay private, presumably. One question that emerges is how should a user indicate to a bot when something has become private? For example, should there be an "incognito" mode for bots just like there is for browsers? Or should there be the ability to clear the history of a bot, just like for a browser -- e.g., "forget everything I told you in the last 30 minutes"?

Another interesting question is whether a bot should have only a single personality. Previous research has studied the possibility of single vs. multiple personalities [19, 39]. Drawing upon ideas from sci-fi movies, what about the movie 'Interstellar', where things like humor can be dynamically tuned from 0% to 100%? The negotiation for humor then, between the user and the bot, might look like this exchange:

"User: Humor, seventy-five percent.
Bot: Confirmed. Self destruct sequence in T minus 10, 9...
User: Let's make that sixty percent.
Bot: Sixty percent, confirmed."

We had some evidence that a single user might prefer different characteristics of bots for different things. For example, users wanted a chatbot like Maya, who could add value to their life while being a friend, by making useful recommendations. But they

also wanted the bot to be infused with fun elements from Ada. So the question that emerges is what should really be the interface to specify "How you would like your chatbot" (the title of paper)? Should there be a set of 3-5 bots to choose from at any given time? Or should there be a collection of many personality characteristics that the user can fine-tune? Should the user even be expected to specify what they want at a given time, or should the bot be able to adaptively change its personality based on metrics of user engagement (e.g., how much user is typing, whether they are putting smile emoticons, etc.)

Having discussed the above, we acknowledge that AI technology may be far from realizing bots that are anywhere close to a human wizard. But this means that the observations in this paper are even more important for engineers working on bots, since these results allow them to prioritize what limited functionality they should focus on (since it seems impossible to do everything). An interesting related research question is what positive aspects of bot interactions/personalities can be preserved even with an imperfectly implemented bot? What are mechanisms to allow a bot to recover (i.e., retain user engagement and "save face" relative to their supposed personality) if they make a mistake? For example, will it be important for the bot's personality to include a major flaw/disability (like short term memory loss – the fish Dory from the movie 'Finding Nemo') that explains their weird responses? Previous research has shown that people find it easier to identify with a character who has some weakness or flaw that makes them seem human; ''perfect'' characters are unnatural and seem less lifelike [43]. If done correctly, this could help the user empathize and like the bot even if it is non-sensical sometimes.

We suggest a few tips for how the most important value propositions discussed above could be implemented automatically. Useful links could be provided using standard search engines. Witty one-liners could be hardcoded from other sources. User responses to one-liners (e.g., "haha") could be used to calibrate which ones are the funniest/best, and should thus be reused in other contexts. There could be more interesting measures, such as having the bot repeat other users' lines (similar to Cleverbot [11]). Or what about having a live human drop into real chats now and then, just to spice it up and make the user think the bot is really smart? There are obviously some interesting privacy implications for both of these latter approaches.

6   Limitations of study

Ours is a small, qualitative study and given our user sample, which is predominantly educated young male, we caution our readers against generalizing the results of this study to every young, urban user across India. But there are grounds to suspect that much of our observations will transfer to other similar groups, at least within urban India, within a similar educational context and socio-cultural ethos.

We made recommendations and suggestions based on our exploratory WoZ studies and one-on-one interviews. One of the biggest limitations of our study was that it was conducted under controlled conditions and every user just had one session to chat with each bot. This set up did not allow us to understand how the bot might have come to be

used and appropriated over time. In the one-on-one interviews users did talk about how they thought the chatbot might be used when deployed in-the-wild, but our study was not set up to observe actual longitudinal use. While our findings are mostly for initial uses of a chatbot, we believe our recommendations could apply to chatbot design more generally.

7   Conclusions and Future Work

Overall users wanted a chatbot who could add value to their life while being a friend, by volunteering useful recommendations on topics typically they brought up. They liked when there were intelligent comebacks, witty one-liners and casual conversation as well. In the longer run, once they came to trust the bot, they said they could also open up about sensitive topics they might not discuss with a human friend. Over time users said they would like the bot to be reassuring, empathetic and non-judgmental, but not in an overly caring way. Topics of interest in conversations varied with the personality of the bot and overall included: movies, predominantly from Bollywood; TV shows, mostly from the U.S.; music, books, travel, fashion, current affairs, work-related stress. It was felt that the topic of conversation would depend on the time of the day, and place where the conversation took place. It was felt that when deployed in-the-wild, conversations could veer off into adult chat under conditions of anonymous use. Users also tested the boundaries of the chatbots, trying to understand what all the bots were capable of, and how much human mediation could be involved. As part of future work we are dialoging with developers on how to incorporate the suggestions from the user studies into chatbot implementations.

References

1. ALICE http://www.alicebot.org/bios/richardwallace.html Artificial Intelligence Foundation
2. Augello, A., Gentile, M., Weideveld, L., & Dignum, F. (2016). A model of a social chatbot. In Intelligent Interactive Multimedia Systems and Services (pp. 637-647). Springer International Publishing.
3. Bee, N., Andre, E., and Tober, S. 2009. Breaking the ice in human-agent communication: Eye-gaze based initiation of contact with an embodied conversational agent. International Conference on Intelligent Virtual Agents (IVA'09), pages 229–242. Springer
4. Bella, M. & Hanington, B., 2012. Universal Methods of Design, Beverly, MA: Rockport Publishers. p204
5. Bickmore, T. W., and Picard, R. W. Establishing and maintaining long-term human-computer relationships. ACM Transactions on Computer-Human Interaction (TOCHI) 12, 2 (2005), 293-327.
6. Bickmore, T, Pfeifer, L, Byron, D, Forsythe, S, Henault, L, Jack, B, Silliman, R, Paasche-Orlow, M (2010) Usability of Conversational Agents by Patients with Inadequate Health Literacy: Evidence from Two Clinical Trials. Journal of Health Communication 15, 197-210.
7. Bickmore, T., Schulman, D., Vardoulakis, L. Tinker - A Relational Agent Museum Guide. Journal of Autonomous Agents and Multi-Agent Systems, 27(2): 254-276 (2013)

8. Bickmore, T. and Schulman, D. (2007) Practical Approaches to Comforting Users with Relational Agents ACM SIGCHI Conference on Human Factors in Computing Systems.

9. Breazeal, C. 2002. Designing Sociable Robots. Cambridge, MA, MIT Press.

10. Cassell., J. 2001. Embodied Conversational Agent: Representation and Intelligence in User Interfaces. AI Magazine. 22, 4: 67-83

11. Cleverbot. http://www.cleverbot.com/

12. Cooper, J. (2006). The digital divide: The special case of gender. Journal of Computer Assisted Learning, 22, 320–334.

13. Correa, T. (2010). The participation divide among "online experts": Experience, skills, and psychological factors as predictors of college students' web content creation. Journal of Computer-Mediated Communication, 16(1), 71–92.

14. Corti, K., & Gillespie, A. (2016). Co-constructing intersubjectivity with artificial conversational agents: people are more likely to initiate repairs of misunderstandings with agents represented as human. Computers in Human Behavior, 58, 431-442.

15. Crutzen, R., Peters, G.-J.Y., Dias Portugal, S., Fisser, E.M., & Grolleman, J.J. (2011). An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. Journal of Adolescent Health, 48, 514-519.

16. Dale, R. (2016). The return of the chatbots. Natural Language Engineering, 22(5), 811-817.

17. D'Onfro, J., 2015. Microsoft Created a Chatbot in China that has Millions of Loyal Followers who talk to it like in the Movie 'Her'. Business Insider UK. http://www.businessinsider.in/Microsoft-created-a-chatbot-in-China-that-has-millions-of-loyal-followers-who-talk-to-it-like-in-the-movie-Her/articleshow/48312697.cms

18. Drew, P., and Heritage, J. (2006), Conversation Analysis, London, Sage.

19. Dryer, D.C., Getting personal with computers: How to design personalities for agents. Applied Artificial Intelligence, 1999. 13(3): p. 273-295.

20. Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. Computers in Human Behavior, 49, 245-250.

21. Jabberwacky. http://www.jabberwacky.com

22. Kakaraparty, S. 2015. GUIDE: How to get Natasha for HIKE! http://www.techkmate.com/2015/02/guidehow-to-get-natasha-for-hike.html

23. Kelley, J. F., "An iterative design methodology for user-friendly natural language office information applications". ACM Transactions on Office Information Systems, March 1984, 2:1, pp. 26–41.

24. Kerly, A. & Bull, S. (2006). The Potential for Chatbots in Negotiated Learner Modelling, in M. Ikeda, K. Ashley & T-W. Chan (eds), Intelligent Tutoring Systems: 8th International Conference, Springer-Verlag, Berlin Heidelberg, 443-452.

25. Kerly, A., Ahmad, N. & Bull, S. (2008) Investigating Learner Trust in Open Learner Models using a 'Wizard of Oz' Approach, Intelligent Tutoring Systems: 9th International Conference, Springer-Verlag, Berlin Heidelberg, 722-724.

26. Klein, J., Y. Moon and R. Picard 2002. This Computer Responds to User Frustration: Theory, Design, Results, and Implications. Interacting with Computers 14: 119-140

27. Lomas, N. 2016. Microsoft officially outs another AI chatbot, called Zo https://techcrunch.com/2016/12/14/microsoft-officially-outs-another-ai-chatbot-called-zo/

28. Luger, E., and Sellen, A. "Like Having a Really bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In Proceedings of CHI 2016

29. McKirdy, A., 2015. Microsoft says Line's popular Rinna character is new way to engage customers. http://www.japantimes.co.jp/news/2015/08/19/business/tech/microsoft-says-lines-popular-rinna-character-new-way-engage-customers/#.WIh5oFN95aQ

30. McGuire, A. 1994. Helping Behaviors in the Natural Environment: Dimensions and Correlates of Helping. Personality and Social Psychology Bulletin 20(1): 45- 56

31. Moon, Y. 1998. Intimate self-disclosure exchanges: Using computers to build reciprocal relationships with consumers. Cambridge, MA, Harvard Business School.

32. Morkes, J., H. Kernal and C. Nass. 1998. Humor in Task-Oriented Computer-Mediated Communication and Human-Computer Interaction. CHI 98.

33. Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. Journal of Social Issues, 56(1), 81–103.

34. Nass, C., Moon, Y., & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems. Journal of Applied Psychology, 29(5), 1093–1110.

35. Nass, C., Fogg, B., & Moon, Y. (1996). Can computers be teammates? International Journal Human-Computer Studies, 45, 669–678.

36. Pan, X., Slater, M.: A preliminary study of shy males interacting with a virtual female. In: PRESENCE 2007: The 10th Annual International Workshop on Presence (2007)

37. Quora. How Does Natasha the bot of Hike Work? https://www.quora.com/How-does-Natasha-the-bot-of-Hike-work

38. Reeves, B., & Nass, C. (1996). The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press.

39. Selker, T. 1994. Coach: A teaching agent that learns. Communications of the ACM 37(7):92 99.

40. Vinyals, O., & Le, Q. (2015). A neural conversational model. Proceedings of ICML Deep Learning Workshop.

41. Weizenbaum, J. (1966). "ELIZA—a computer program for the study of natural language communication between man and machine". Communications of the ACM. 9: 36–45.

42. Weizenbaum, J. (1967): Contextual understanding by computers. In Communications of the ACM, 10 (8) pp. 474-480.

43. Thomas, F., and O. Johnston. 1981. Disney animation: The illusion of life. New York : Abbeville Press.