

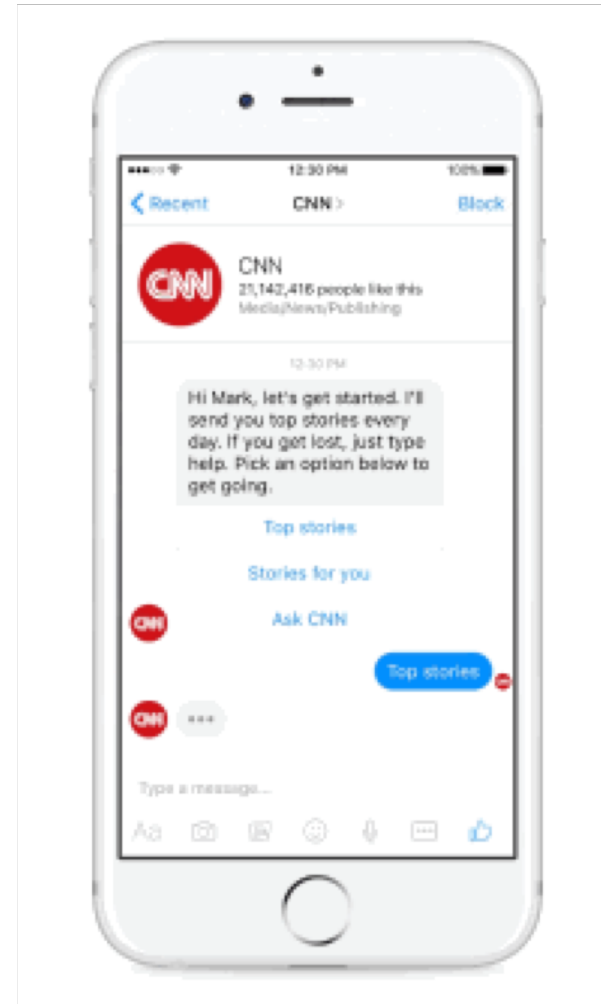
Evaluating and Informing the Design of Chatbots

Mohit Jain, Pratyush Kumar, Ramachandra Kota, Shwetak Patel

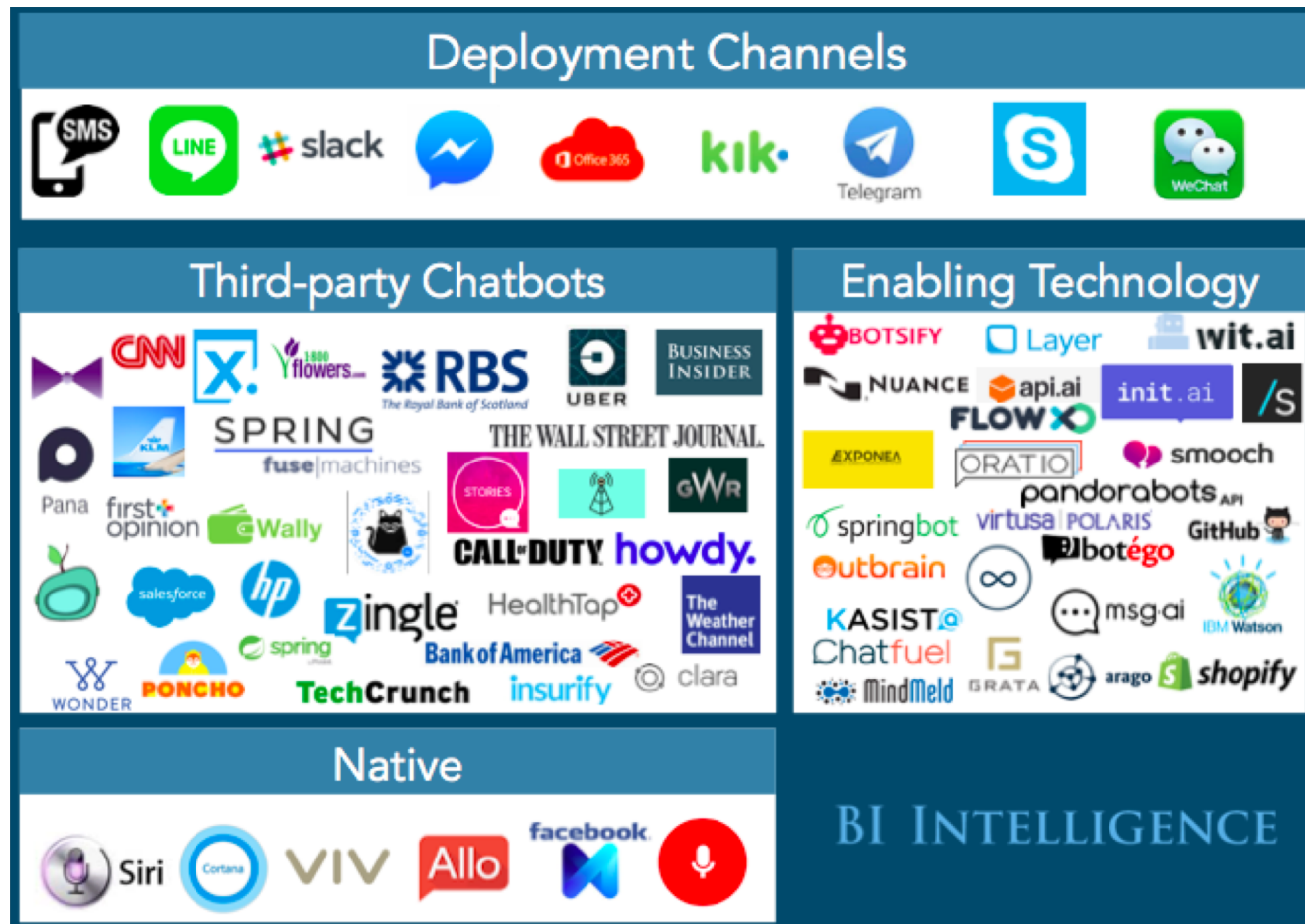


Chatbots

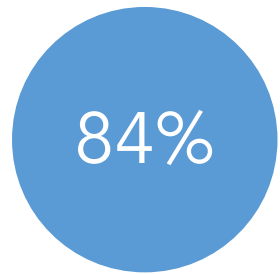
Text messaging-based
conversational system



Chatbot Ecosystem



Problem



of the Internet users have not used a chatbot yet

Crucial to understand the interaction pattern of first-time chatbot users to inform and guide the design of future bots

Our Research Work

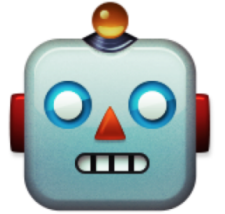
We study the experience of

16 first-time chatbot users

interacting with **8** chatbots 

on the Facebook Messenger  platform

Selection of Chatbots



Aim: Select chatbots with which a new user is most likely to interact

Top 100 Messenger chatbots [on Chatbottle]

Identified eight domains

Selected the highest rated chatbot in each

Chatbot	Domain
Alterra	Travel
Call of Duty	Entertainment
chatShopper	Shopping
CNN	News
Hi Poncho	Utility
Pandorabots	Chit-chat
Swelly	Social
Trivia Blast	Game

Selection of Participants

10/16 Engineering background

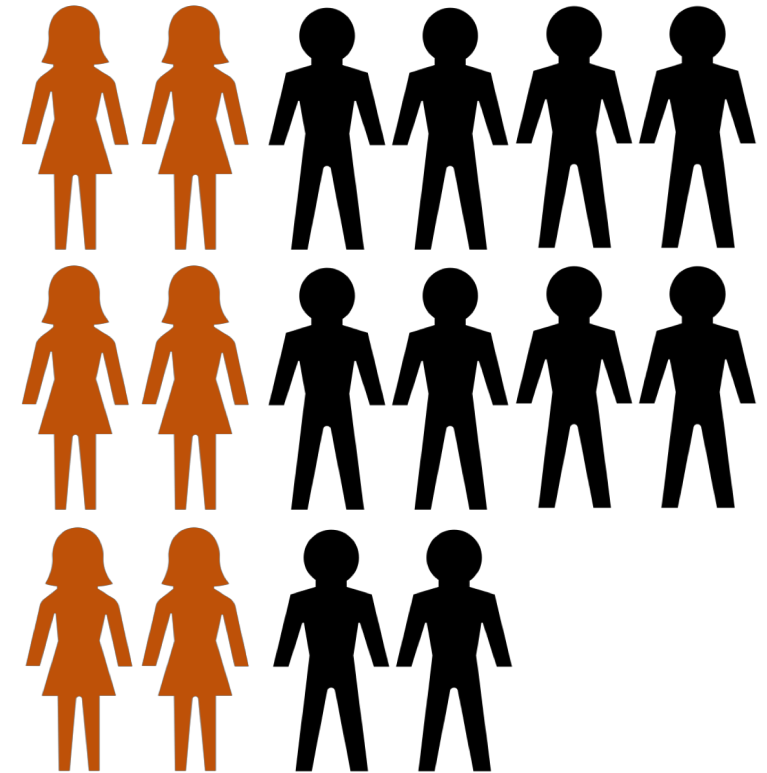
Avid computer and phone user (11.8 ± 1.3 hrs daily)

Avid Messenger user

All understood chatbots at a conceptual level

No prior experience with chatbots

Participants adequately represent technology early adopters who will likely constitute the majority of chatbot users in the near future.



32.1 ± 6.9 years

Procedure

Use **each chatbot** for **~3-5 mins daily** for the **next 3 days**

Each participant received a daily reminder

To encourage **exploration** and **open-ended chatbots** usage

Participants were **not** instructed on:

how to interact with the chatbots,

what the chatbots were about, or

what kind of tasks to perform using the chatbots.

Results

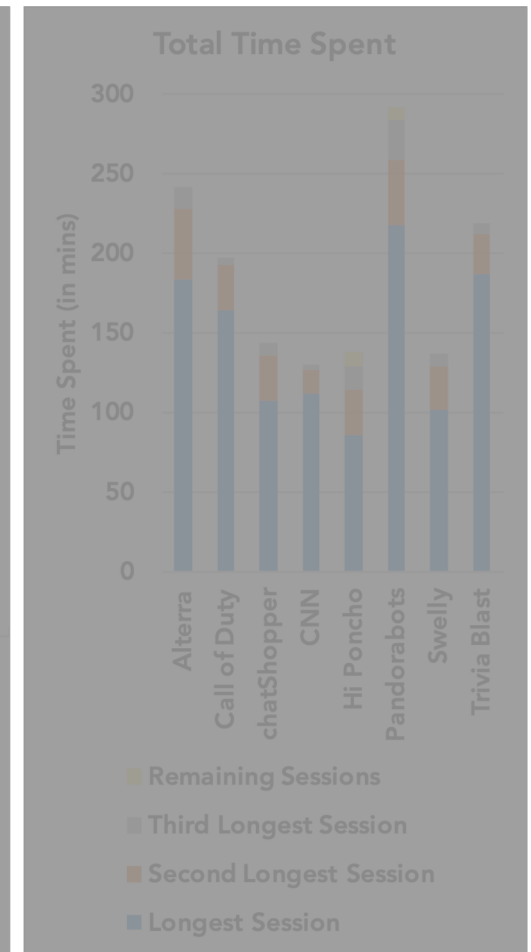
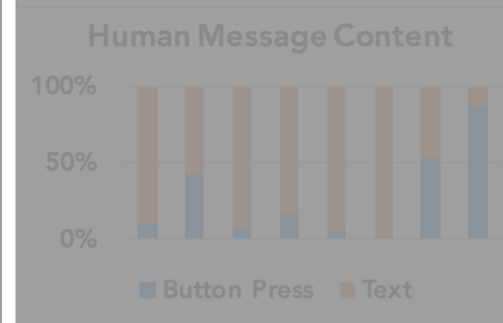
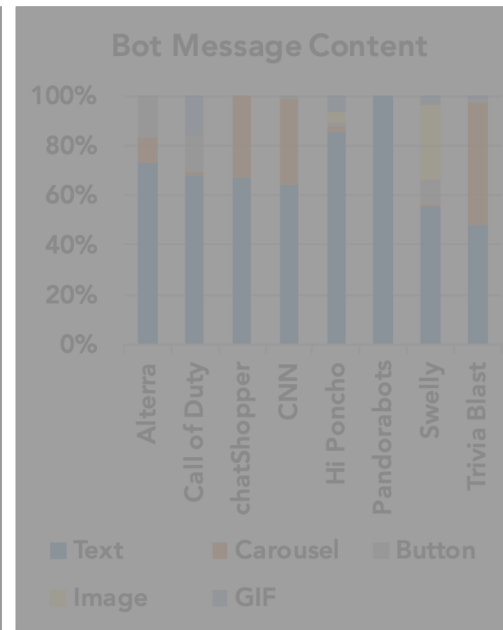
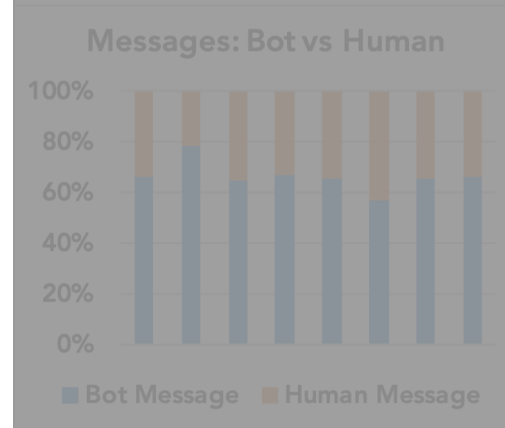
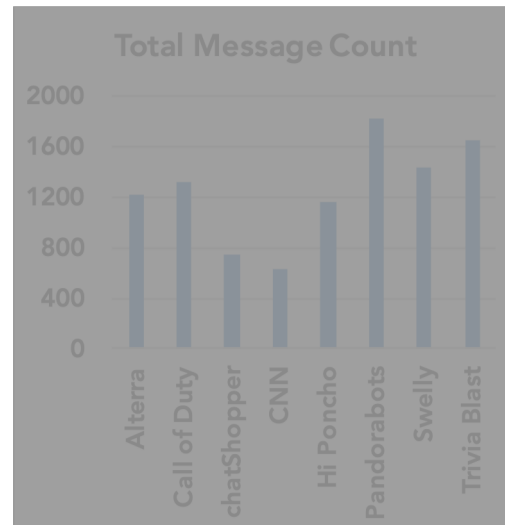
25 hours

379 sessions

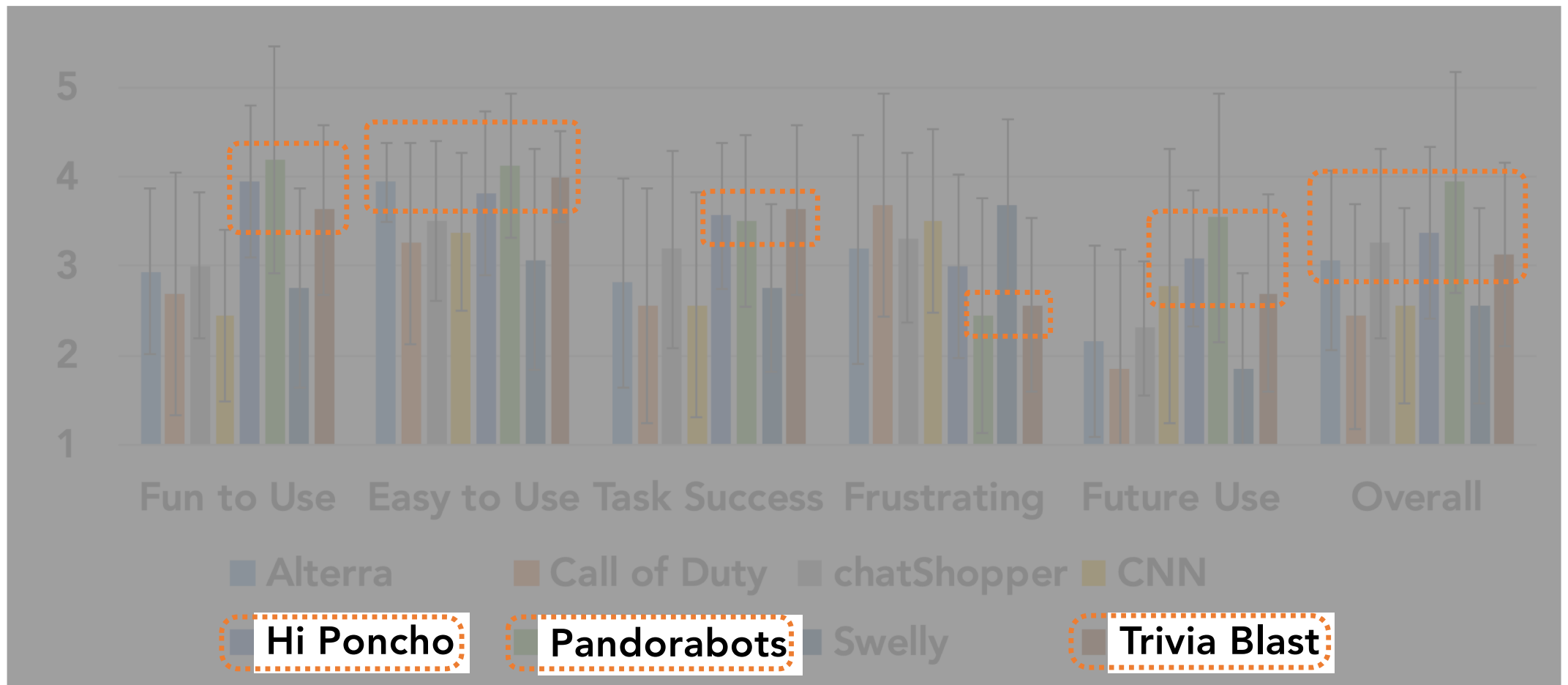
9968 messages

65.8% by chatbots

34.2% by users



Results



Four Themes



Functionality



Conversational Intelligence



Personality



Interface

Theme 1: *Functionality*

Did the chatbot do what it is supposed to do, and if so, how good was it?



"help me pass time during commute" P₈, Trivia Blast

"It doesn't even understand 'weather', 'Pakistan', 'migrations'... doesn't work at all" P₁₃, CNN

Participants appreciated chatbots which were able to perform "tough" tasks.

"It worked even for 'rain in Bangalore', 'hiking in London', 'umbrella in Seattle'. It just works!". P₁₅, Hi Poncho

Theme 1: *Functionality*

Compare the bots with existing alternatives, including apps/websites, and search engines, to accomplish the same task



"awesome idea... I can't google for opinions" P₁₀, Swelly

"websites are better than Alterra... I can quickly browse through hundreds of flights" P₂, Alterra

Suitability of chatbots was highly dependent on domain

Theme 2: *Conversational Intelligence*

Chatbot's understanding of the input text as an important criterion to determine whether its a "chatbot" or not.



"not a chatbot, as (it) can't chat" P₉, Trivia Blast

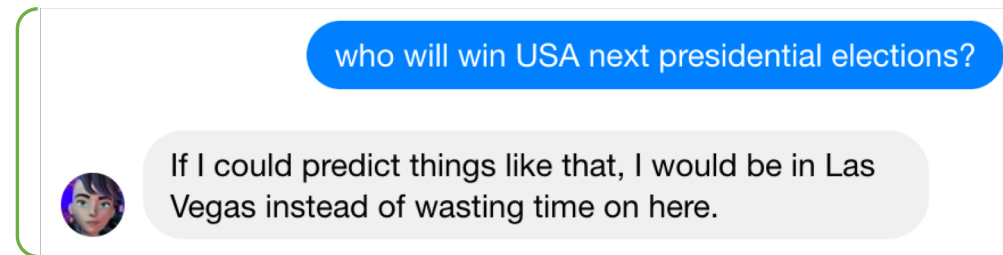
"It is as good as talking to a human" P₉, Pandorabots

"It answers like my spouse" P₁, Pandorabots

Participants expected other chatbots with basic "keyword-understanding" (Alterra, chatShopper) to have "human-like conversational abilities"

Theme 2: *Conversational Intelligence*

For conversation failures, users liked chatbots covering-up with a smart response, or admit that it failed



P₈, Pandorabots

Theme 2: Conversational Intelligence

Retaining conversational context



“super happy to use it (chatShopper)” because it was able to follow up on her query of “shoes”, followed by “in red” $P_{1, \text{chatShopper}}$

“I told Pandorabots that X is my friend and Y is his wife. Later I asked her, who is Y, and she correctly said X’s wife!” $P_{10, \text{Pandorabots}}$

Theme 3: *Personality*



Personalities have a strong impression

chatShopper, Pandorabots : 'he', 'she' (pronouns)

CNN, Trivia Blast : 'it' (tools)

(Personal) Small talk

*"didn't even respond to how
are you?... not even to hi"*

P₁₅, Call of Duty

*"She was not addressing me by
my name... very impersonal."*

P₁₅, Pandorabots

Theme 3: *Personality*



Humor

*“Cool, I DJ’ed there once. Good crowd.
Right now it is 28°C and clear there.”*

P₆, Hi Poncho

why are you learning about humans?



Because if I know a lot about human behaviour, it
will be easier to erase your species.

P₁₅, Pandorabots

Exit gracefully

*“It was impossible to end the conversation. I tried ‘exit’,
‘quit’, ‘stop it’, ‘end this’, still it kept talking.”*

P₉, Call of Duty

Theme 4: Interface



Bot Text

Button

User Button Press

User Text

Carousel

Menu Button

Auto-suggestion Buttons

“it doesn't require typing, just interacted with the buttons.” P₅, Trivia Blast

Theme 4: *Interface*

Opening content in a new window detached from the chat interface



clicking on 'Read this Story' button opens a new CNN webpage... "has to leave the current (browser) tab. With 10+ tabs open, coming back to that tab is tricky"

P_{5, CNN}

Persistent display of chatbot capabilities, and a menu option to access the chatbot main functionalities

Summary: *Functionality*



A chatbot must:

accomplish its primary task

outperform its existing website/app/search engine alternatives by offering diverse and/or enhanced functionalities

check for domain suitability



Summary: *Conversational Intelligence*



A chatbot needs to have 'human-like' conversational capabilities, including:

context preservation (intra- and inter-session),

understanding of negative statements,

cover-up smartly or admit failure, and

ability to ask intelligent questions proactively to reduce its search space, engage the user in a meaningful conversation, along with helping the user with the task.



Summary: *Personality*



Chatbot should have an apparent personality suiting its domain.



The chatbot should be able to:

- introduce and advertise its functionalities,
- engage users in small talk,
- provide a personal touch,
- respond humorously, and
- exit gracefully.

Summary: *Interface*



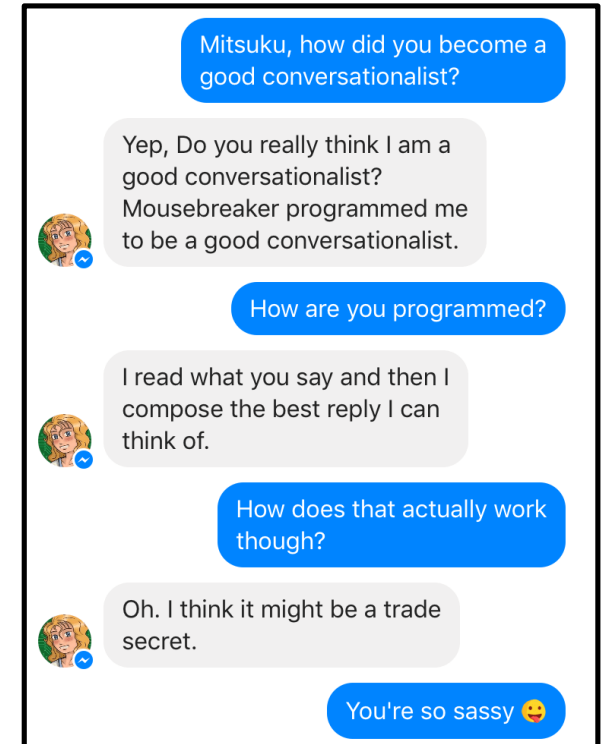
A chatbot should have:

interactive elements in the interface, along with text input,
minimal external links, and

show certain information, including chatbot's description and main
menu, persistently to the user



Thank You!



Mohit Jain

IBM Research, India: mohitjain@in.ibm.com

University of Washington, Seattle USA: mohitj@uw.edu