Mohit Jain | Research Statement

My research work **explores the introduction of emerging technologies to new populations**. I examine it through many research approaches, including novel interfaces and interactions, technology prototypes, qualitative and quantitative usability studies, and longitudinal deployments. For my Ph.D. research work, I focused on conversational systems (*chatbots*) as the emerging technology. I have conducted studies to identify usability issues experienced by first-time users of text-based chatbots [C16], explored novel interfaces to enhance the usability and acceptability of these chatbots [C21, C24, C25, W8], proposed methods to increase the security of speech-based chatbots [C20, U3], identified novel interactions with and applications of smart speaker conversational devices [C17], and even broadened the applicability of chatbots to new demographics including low literate users [J2]. I developed these varied solutions by drawing on my **diverse set of skills**, including mobile computing, data analysis, game design, signal processing, software engineering, and designing and conducting user studies. My research is additionally informed by perspectives in design, information visualization, communication theory, and behavioural, social and environmental psychology. Prior to my Ph.D., my research focused on designing, building, and evaluating **technological solutions for developing world** and **novel input and interaction techniques**. Thus, I have strong background in human-computer interaction (*HCI*), ubiquitous computing (*UbiComp*), and information and communication technologies for development (*ICTD*).

My research is inherently **interdisciplinary**. I frequently collaborate with not only experts in computer science and electrical engineering, but also broadly with designers, sociologists, farmers, teachers, and doctors. I also partner with various government and non-government organizations (NGO) to conduct the necessary fieldwork to ensure that the computing systems I build are appropriate for the target population. With my **strong 7+ years of industrial research experience** (as I did part of my Ph.D. while employed at IBM Research), I possess skills to contribute and work in large teams and impact product development. Moreover, I have led small teams for multiple internal and client-facing projects, have mentored several undergraduate and graduate interns, and have successfully collaborated with several universities. I am experienced in working and delivering on multiple projects at the same time. I am highly flexible and willing to learn new technologies based on the industry's requirement.

Over the course of 10+ years of my research career, I have published *25+ research papers* at top-tier computer science journals and conferences, filed *15+ patents*, received several *awards* (including 1 best paper award, 1 best paper nomination, invited to attend the Heidelberg Laureate Forum, a Wolfond Fellowship, an Aegis Graham Bell award, 8 IBM Accomplishment awards, and 2 IBM Outstanding Technical Achievement awards), won several *competitions* (including DFS Tech Chatbots Challenge 2017 and Microsoft Imagine Cup 2009), and contributed to multiple *client deals* with my work at IBM.

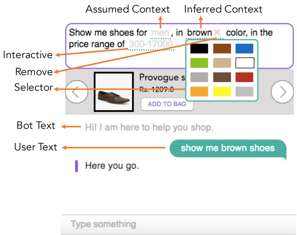
CONVERSATIONAL SYSTEMS

Conversational interfaces are fascinating, as conversation is the most familiar mode of interaction, requiring little learning or literacy. Chatbots have appeared on a variety of mobile and ubiquitous platforms, including smartphones, VR/AR devices, smart speakers and smart watches, with three major interaction modalities: speech, text, and their combination. However, only 16% of Internet users have interacted with a chatbot. This could be due to several reasons: usability issues with text-based chatbots, security concerns with speech-based chatbots, or targeting of chatbots specifically towards the Internet-savvy technically-advanced users. In my Ph.D. work, I explored these issues:

Usability Issues with Text-based Chatbots

In order to understand the problems with current text-based chatbots, I conducted a user study in which 16 first-time chatbot users interacted with 8 chatbots over multiple sessions [C16]. Analysis of chat logs and user interviews revealed several issues. As a solution to the identified issues, I proposed, developed and evaluated three chatbot systems, which are described below.

First, there is a mismatch between the chatbot's state of understanding (also called *context*) and the user's perception of the chatbot's understanding. To reduce this gap, I proposed **Convey** [C21], short for CONtext View, which is a window added to the chatbot interface that displays the inferred and assumed context of the conversation to the user (Fig 1). Convey's content gets updated as the conversation proceeds, always showing the latest understanding of the chatbot. Convey also provides intuitive interactions for users to modify context values in a simple and efficient manner.



Show me shoes for men, in brown ✕ color, in the price range of 300-17000

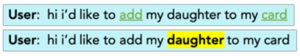
Hi! I am here to help you shop.

show me brown shoes

Here you go.

Figure 1. A shoe shopping chatbot with **Convey** at the top, where in the user is modifying the desired shoe color *context* using Convey.

Second, the state-of-the-art natural language understanding technologies are limited, leading to dialog failures. A *dialog failure* happens when a chatbot is unable to understand the user's input. **Resilient Chatbot** [C25] explores user preferences between eight repair strategies to recover from failures. We proposed three novel repair strategies explaining characteristics of the underlying machine learning algorithms (Fig 2), informed by recent work in transparency and explanation of AI systems. The remaining five strategies (such as confirmation, providing options) were adopted from commercial chatbots and were guided by grounding in communication theory. A scenario-based study deployed to 340 Mechanical Turk workers found that providing options and explanations were generally favored, as they manifest initiative from the chatbot and facilitate breakdown recovery. The findings of this research will inform the next version of IBM Watson Assistant.



User: hi i'd like to add my daughter to my card

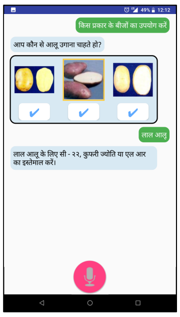
User: hi i'd like to add my daughter to my card

Figure 2. **Resilient Chatbot**: Two of the eight repair strategies: Keyword Highlight Explanation and Out-of-Vocabulary Explanation.

Third, users have unreasonably high expectations with chatbots. To teach users about chatbots, their internal functions, conversational failures and recovery strategies, we developed a fun learning experience, **BigBlueBot** [C24, W8]. BigBlueBot uses various methods, including *role-reversal*, where users need to act as chatbots. In a Mechanical Turk evaluation with 88 participants, they learned strategies for having successful human-chatbot interactions, reported feelings of empathy toward chatbots, and expressed a desire to interact with chatbots in the future.

Chatbots for Low Literate Users and Novice Smartphone Users

Current chatbots specifically target technically-advanced users. However, chatbots are well-suited for novice smartphone users as it offers direct information access without the need to navigate complex information paths required by GUIs. Also, chatbots enable users to formulate queries as if they were talking to another person, thus making chatbots suitable for low literate users. Thus, I designed **FarmChat,** a multi-modal, multi-lingual chatbot to support the farmer's information needs using natural speech interactions (Fig 3) [J2]. The knowledge base of FarmChat was informed by agricultural-experts working for Transform Rural India, an NGO in Ranchi, India. Through a formative study, we found that farmers were interested in seed selection, crop disease treatments, weather forecast, *etc*. An evaluative study with 34 farmers showed that FarmChat was highly usable and the information provided was acceptable to the farmers. FarmChat won a $20,000 grant (DFS Tech Chatbots Challenge 2017) funded by the Gates Foundation. In the future, my goal is to release FarmChat as a free and open-source tool for use by development organizations across the world.



किस प्रकार के बीजों का उपयोग करें

आप कौन से आलू उगाना चाहते हो?

लाल आलू

लाल आलू के लिए सी - २२, कुफरी ज्योति या एल आर का इस्तेमाल करें।

Figure 3. User interface of **FarmChat** (input: speech, button; output: audio, text, image).

Smart Speakers Security and Limited Interactions

The rapid growth of speech-based conversational interfaces has made them vulnerable to potential security threats. We developed an end-to-end system to detect replay attack–replaying a recording of the target speaker's voice—using deep convolutional networks [C20], which achieved an equal error rate of 0% on the evaluation set of the ASVspoof'17 dataset. Moreover, we explored a method to auto-generate usable and secure audio CAPTCHAs, to secure them against attacks by a text-to-speech system [U3]. We evaluated its usability with 60 sighted and 19 visually impaired participants.

Current smart speakers are limited to speech-only interaction. My intern and I exploited the in-built microphone array in smart speakers for opportunistically sensing gestures and tracking exercises. This was achieved by measuring the Doppler shift caused by a moving human body from an inaudible 20 kHz pilot tone (Fig 4). Beamforming at the microphone array enabled accurate detection till 3.5 meters away from the device [C17]. A neural network recognized 10 different exercises from 17 participants with 96% accuracy. This work not only enables multimodal interactions with smart speakers, but also supports diverse non-conversational applications of smart speakers.

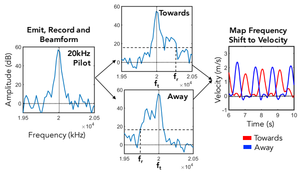


Figure 4. Signal processing pipeline to transform the audio signal received at 6 microphone channels to a velocity-vs-time plot.

These chatbot works were **inspired by my research experience prior to my Ph.D.** FarmChat was influenced by my active interest and prior work in building technologies for the developing world. My work on smart speaker-based interactions was due to my interest in building novel user interactions. Finally, BigBlueBot was motivated by my prior deep exposure to game design for teaching children. Here I will briefly discuss those prior works.

TECHNOLOGIES FOR THE DEVELOPING WORLD

For almost a decade now, I have been working on high-impact social problems: **education, healthcare, environmental sustainability** and **agriculture**. I have spent substantial time in the field, understanding the problem, and conducting longitudinal studies to ensure that the systems I have helped build were appropriate and usable. Building technological solutions for developing world not only requires core technical skills, but also a deep understanding of the complex cultural, social, economic and infrastructure challenges that impact the research process.

Education and Game Design

My first exposure to the research area of ICTD was during my undergraduate internship, working on an interdisciplinary collaborative project between UC Berkeley, Microsoft Research India, and the Azim Premji Foundation. Schools in developing countries tend to have a high student-to-computer ratio (ranging from 2:1 to 10:1). In such cases, learning benefits appear to accrue primarily to the child with the mouse, with the other children misses out (Fig 5). As such, we sought to provide each child with a mouse and cursor on screen, thus integrating every student into the active and collaborative learning experience at the cost of a few extra mice. Starting with paper prototypes, I redesigned the existing multimedia educational content to account for multiple inputs while minimizing changes to its pedagogical design [W1]. The design process was propelled by discussions with educational technology experts, teachers, and children. Our prototype was deployed in a low-income school, and observations during field studies led to several *design iterations*.

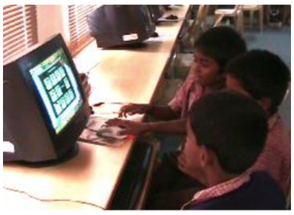


Figure 5. Note the proximity between the mouse and the hands of right and center children, as the left child watches.

Instead of redesigning existing content, I believed that educational content designed specifically for multiple mice would result in more engagement, greater learning, and higher retention. Hence, I designed **DISHA (DISease and Health Awareness)**, a multi-mouse-enabled collaborative learning platform that uses a narrative-interactive loop format of story-telling and multiple-choice Q&A to teach the symptoms, prevention methods, and cure of malaria (Fig 6). DISHA was evaluated for immediate learning and retention [C1, W2], and won the Microsoft Imagine Cup 2009.

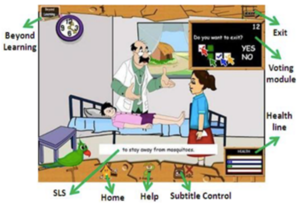


Figure 6. The **DISHA** interface while teaching users about prevention against malaria.

DISHA's field studies raised questions about the restrictive nature of computers on students' gameplay behavior. Thus, I explored two mobile-based techniques for using co-located collaborative gameplay to supplement English learning: (1) **Single Display Groupware**: a pico-projector connected to a phone with a handheld controller for each child to interact, and (2) **Multiple Display Groupware**: a phone for each child (Fig 7). Both were deployed in two rural schools. A longitudinal study showed that the single display groupware led to more learning-based interactions among the players, while the multi display groupware was preferred by them due to the added mobility [C19].

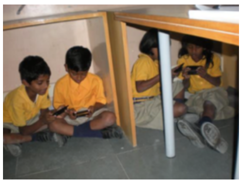


Figure 7. The **Multiple Display Groupware** uncovered interesting ways that children would leverage mobility. Here, children were found playing under the table in groups.

Healthcare

According to the WHO, high blood pressure (BP) causes roughly 12.8% of all deaths worldwide per year. Frequent monitoring of BP at home has been shown to improve the management of elevated BP. Hence, my colleagues and I developed and evaluated a **smartphone-based blood pressure monitoring application called *Seismo*** [C22] (Fig 8). Seismo relies on the measurement of pulse transit time (PTT), the time it takes for a pulse wave to travel from the aortic valve to a peripheral arterial site. Seismo uses the smartphone's accelerometer to measure the vibration caused by the heart valve and the smartphone's camera to measure the pulse at the fingertip. The system uses signal processing and machine learning techniques (described in Fig 9) to identify the peaks, find timing between them (PTT), and to estimate BP. Seismo was evaluated in a nine-person, four-session longitudinal BP study, and achieved an average RMSE of 5.2 mmHg across participants.

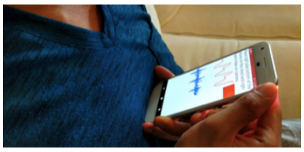
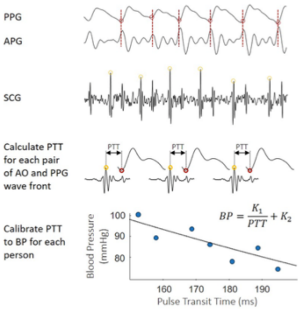


Figure 8. Measuring blood pressure using **Seismo**.


PPG
APG
SCG
Calculate PTT for each pair of AO and PPG wave front
PTT
PTT
PTT
Calibrate PTT to BP for each person
$BP = \frac{K_1}{PTT} + K_2$
100
90
80
Blood Pressure (mmHg)
160
170
180
190
Pulse Transit Time (ms)
Figure 9. Seismo: Photoplethysmogram (PPG from camera) and Seismocardiogram (SCG from accelerometer) are used to measure the PTT. To convert PTT to BP, an individualized calibration is generated based on BP cuff reference.

Along the same vein of developing novel inexpensive healthcare solutions, my team demonstrated that **gaze tracking can be used to identify people with autism** while they watch videos on a computer. Using a Random Forest classifier, our system achieved 100% accuracy in classifying autistic children across 60 participants [U1]. Moreover, a multi-layer perceptron-based regressor achieved a mean absolute error of 2.03 in estimating the patient's Child Autism Rating Scale score (range: 15 to 60) [U1]. This work was done in close collaboration with doctors.

Sustainability

In 2012, I found that most of the sustainability solutions proposed in the HCI literature were based on studies conducted in the developed world. Thus, I conducted several studies in collaboration with universities (CMU, UNCC and IIIT-D), exploring consumer's perception, their beliefs and attitudes, towards **environmental sustainability at homes, dormitories and workplaces** in India. Studies included surveys, photo-elicitation based qualitative studies, and energy feedback deployment studies. Results from these studies highlighted a culture of deep conservation for residential users and identified new opportunities for technology designs [C5, W5]. Integrating findings from social and environmental psychology, I found *daily individual paper feedback* encouraged the maximum conservation for people living in dormitories, resulting in ~20% energy reduction [C13]. Despite strong motivations to conserve in workplace settings, employees' conservative actions were limited due to lack of controls, knowledge, and responsibility [C8].

In offices, HVACs typically account for 40% of the energy consumption. HVAC loads vary as a function of occupancy; therefore, inferring occupancy is vital to optimizing energy. Hence, I worked on two novel, zero-cost, automated indoor localization techniques: using existing opportunistic context sources (such as calendar data, WiFi data, *etc*.) [C15] and using images captured from phones' cameras to identify the unique ceiling structure of any particular location in the office building [C19].

Focusing on rural settings, a significant portion of the population does not even have access to a reliable electricity supply. Concurrently, the rapid penetration of portable computing devices generates a significant amount of electronic waste. I proposed and developed a prototype of **UrJar**, *a device which uses re-usable lithium-ion cells from discarded laptop battery packs to power low-energy DC devices* (Fig 10) [C11, W7]. To understand its usability, I deployed UrJar at five street-side shops in India that did not have access to the electrical grid. UrJar as an ecosystem can provide a mechanism for DC electrification of rural areas, channeling e-waste towards the alleviation of energy poverty. UrJar was mentioned in 100+ media articles, including BBC and MIT Technology Review.

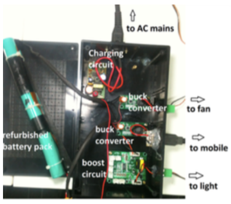


Figure 10. The hardware design of **UrJar**, all of which runs on discarded laptop batteries.

Apart from these academic-oriented projects, I also worked on **several client projects** at IBM. One such project was with Kuala Belalong Field Studies Centre (KBFSC), a research center located in a remote tropical evergreen rainforest in Brunei. It is visited by biologists and ecologists to study the flora and fauna of the surrounding areas. Power is available at the center for 8-10 hours/day from a diesel generator. I helped in developing an energy management software to reduce their fuel consumption while improving power availability [C12]. One of its key feature was a collaborative scheduler that provided visitors at the center with the choice of scheduling appliance usage. It has a persuasive interface to inform and motivate users' scheduling behavior. The system optimizes the generator's active hours using a customized optimizer technique to minimize diesel consumption as per the diesel generator efficiency characteristic (Fig 11). My deployed system reduced diesel consumption by a third and the total cost by 20% while making power available 24x7.

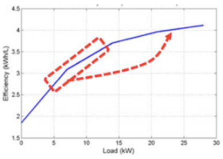


Figure 11. **Diesel generator efficiency** is highest when it is loaded close to its capacity.

NOVEL USER INTERACTION

Additionally, I have explored novel input and interaction techniques with various devices including **smartphone** [C3], **smart television** [C7], and **AR/VR devices** [J1], which led to my interest in smart speaker interaction [C17].

Smartphones: Gestures initiated from the bezel–the touch-insensitive frame surrounding a touchscreen display–can enable eyes-free interaction with smartphones as they can be learned into the muscle memory (Fig 12). I first examined the performance of different bezel menu layouts. Based on the results, I designed a bezel-based text entry application. In a lab-based longitudinal study, participants achieved 9.2 words/minute in eyes-free interaction [C3]. After only one hour of practice, participants transitioned from **novice to expert users**.

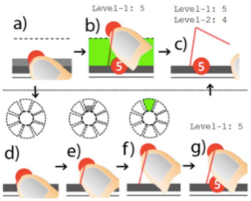


Figure 12. **Bezel Menu**
Expert usage: a→b→c,
Novice usage: a→d→e→f→g→c

Television: Interaction with distant large-screen displays (*e.g.*, interactive TV) requires **active pointing and target selection**. I empirically compared four phone-based distal pointing techniques – Continuous Touchpad, Continuous In-air (using motion sensing), Discrete Touchpad, and Discrete In-air – for point-and-select tasks. Participants preferred Discrete Touchpad, a discrete pointing technique using the phone's touchscreen as a touchpad [C7]. It also achieved the highest accuracy.

AR/VR: Existing input techniques for head mounted AR/VR, either lack expressive power or may not be socially acceptable. As an alternative, thumb-to-finger touches present a promising input mechanism that is subtle yet capable of complex interactions. We proposed **DigiTouch**, a novel reconfigurable glove-based input device that enables *thumb-to-finger touch interaction* by sensing continuous touch position and pressure on resistive fabric [J1] (Fig 13). This enabled a variety of interfaces, such as a split-QWERTY keyboard layout that could be mapped to the user's fingers. Participants achieved a mean typing speed of 16 words/minute at the end of ten 20-minute sessions.

Figure 13. **DigiTouch**: A touch-sensitive glove that explores thumb-to-finger interaction for input and text entry for AR/VR devices.

FUTURE RESEARCH PLANS

My mission is to bring a positive change through the technology that I develop. Currently, I am working with multiple NGOs to scale FarmChat to 10,000+ farmers in India. I am also working on deploying BigBlueBot to the general public to help people have successful conversations with chatbots and generate feelings of empathy towards chatbots. Extending my work on developing systems for people in developing regions, in the future, I am eager to design new technologies for underserved communities in the developed world. Apart from these, I am interested in exploring the following diverse research areas in the future:

Embodied Agents and Robotics: My work in chatbots are applicable to related domains, including embodied agents and human-robot interaction. I plan to explore these domains. How to make the user aware of the robot's state of understanding? What are the novel applications of embodied agents and robots for people in the developing world (*e.g.*, in the domain of education)? Expanding on the work of programming robots through natural language, how to use conversations to program robots?

Data Collection: Labeled data is critical to train machine learning models. While developing FarmChat, I considered enabling users to provide image-based input for plant disease diagnostics. However, the lack of a dataset containing infected and healthy Indian crop images restricted the development of this feature. Recently, I have been working on using a chat-based interface to crowdsource labelled data from farmers. *E.g.*, a chatbot could ask the farmer to '*send an infected potato leaf image*' while she/he describes the potato disease. Moreover, I plan to work on chatbot games for large-scale deployment, which apart from fun experiences, could also be used to incentivize human data collection. This labeled data can be used to train the chatbots.

Healthcare and Wellbeing: The growing adoption of ubiquitous devices presents an opportunity to have always available health monitoring capabilities. A few examples include detecting breathing and heart rate using smartphones, tracking and coaching for pranayama exercises using smart speakers, and in-house step count and fine-detailed activity recognition using a combination of smart speakers and phones. The approach can be summarized as *teaching old sensors new tricks*. Such systems can also act as a coach, asking users to take a break from work and perform on-the-spot exercises to avoid repetitive strain injury (RSI).

Security and HCI/ICTD: I am excited to explore the intersection of security/privacy and HCI, primarily video and audio privacy concerns regarding mixed reality in collaborative AR. Moreover, privacy concerns and its perception among people in rural communities with low digital literacy raises a host of interesting research questions at the intersection of security and ICTD. Finally, designing future systems that better inform target users in simple terminology about where their data will get stored, how it will be used, and who all will access it are challenging future research questions that I aim to pursue.

To conclude, my research *explores the introduction of emerging technologies to new user populations, ranging from first-time chatbot users, first-time smartphone users, first-time renewable energy users, to first-time users using smartphone as a medical device*. My approach is HCI-based, *i.e.*, iterative, often beginning with formative studies and, crucially, concluding with field deployments of working prototypes. I have a broad field of interest, leading me to publish not only at top HCI conferences like CHI [C3, C5, C21, C22] and UbiComp/IMWUT [C8, J1, J2], but also in other disciplines of CS, such as DEV [C10, C11], eEnergy [C18, W7], BuildSys [C12, C13], *etc*. In the past, I have developed innovative solutions, collaborated effectively with multiple partners, led (client-facing) projects, delivered products, learned technology as needed, mentored undergraduate and graduate students, and learned from the field experts including farmers, teachers, and health-workers. In the future, I plan to continue working on high-impact globally relevant problems with interdisciplinary collaborations.

REFERENCES: Please refer my CV (https://homes.cs.washington.edu/~mohitj/pdfs/MohitJain-CV.pdf)