The efficiency of human communication relies largely on the quality of the medium. Modern computing devices offer mediums such as keyboards to interact with information, yet those interactions are still primitive (*e.g.*, the user has to press every character to input a phrase) and often fail to facilitate different user needs (*e.g.*, a touch screen keyboard is not optimized for visually-impaired users). **As a human-computer interaction (HCI) researcher, I design, implement and evaluate intelligent methods to enable people to interact with information, and make systems accessible to users with different abilities.** My dissertation focuses on two research questions: 1) How can input methods utilize contextual information to understand a user's intention, lowering the barriers to efficient interaction? 2) How to model and evaluate the performance of intelligent input and output methods?

In my research, I combine **natural language processing (NLP), machine learning (ML), human factors, and user-centered design** to improve the communication experience on computing devices in three different strands of work: 1) inventing novel text entry methods to accelerate language production, 2) developing accessible tools to lower the barriers of interacting with information for people with disabilities, and 3) designing models and tools and conducting empirical studies to evaluate intelligent information input systems. In sum, I invent and build ways to address the persistent challenge of human communication with machines, and to scientifically validate that I have, indeed, improved it.

My research advances human-computer interaction (HCI) in both technical and behavioral aspects, leading to 15 publications in top-tier HCI conferences (*e.g.*, CHI, UIST, IMWUT). **I have also held internships at Google, Facebook and Apple, making an impact on real-world products**. For example, the input evaluation platform I developed (*TextTest++*) has been used by Facebook Reality Labs, and the Input team at Google also adopted my research to design features on Gboard and the Pixel phone, which are now shipped to billions of devices.

INTELLIGENT TEXT ENTRY METHODS FOR LANGUAGE PRODUCTION

Language, as a medium of information, is an essential feature of human history. Text entry, the task of entering language into computing devices, thus serves as a fundamental interaction in most human-computer systems. However, most text entry methods are command-based and only operate on the character level. The first contribution of my research is a set of **intelligent text entry methods** that utilize high-level semantic context to understand a user's intention, and use this to accelerate the text entry process.

PhraseFlow[1] is a mobile input system that performs phrase-level decoding, which enables the keyboard to correct previous text based on the subsequent input context (Fig. 1). For example, if a user types "he loces in Beijing", the word-level decoder corrects "loces" to "loves",

[1]Mingrui "Ray" Zhang, Shumin Zhai. (2021). PhraseFlow: Designs and Empirical Studies of Phrase-Level Input. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '21)*

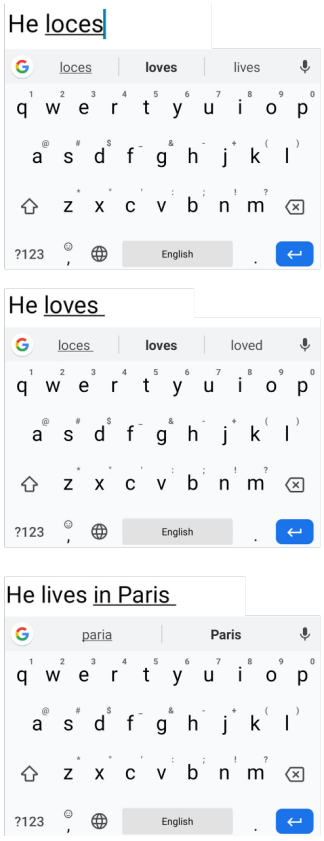

Figure 1. The flow of phrase level text input: "loces" is first corrected to "loves"; as further context is entered, it is then corrected to "lives".

while PhraseFlow corrects "loces" to "lives", yielding more semantically accurate results. To reduce users' cognitive load when using PhraseFlow, I iterated on the design of the keyboard interface and implemented a novel buffer-style correction interaction. The final version of Phraseflow reduced the error rate of autocorrection **by over 16%** compared to the Google Gboard, and users were able to reach better typing accuracy without losing speed. **The outcome of this work also helped the Google Input team to develop related features on Gboard, which is shipped to over 1 billion mobile devices.**

Along with PhraseFlow, which incorporated rich context to accelerate the *text entering process*, I also invented *Type, Then Correct (TTC)*[2], a set of intelligent correction interactions to facilitate the *text editing process*. Editing text on mobile devices often involves repetitive backspacing and cursor-moving actions, which are laborious and time-consuming. I developed three interactions that skipped these steps and allowed the user to type the correction first, then apply it to a previously committed error by dragging the text or pressing a button. The interactions *Drag-n-Throw* (Fig. 2) and *Magic Key* (Fig. 3) are backed by a recurrent neural network (RNN), which automatically detects errors given the input context. By utilizing the power of deep learning, the techniques can understand *what* and *where* the user wants to apply the correction. The concept of TTC has been adopted by other researchers and extended in other input interactions, including gesture typing and dictation.

Collectively, the input techniques I created demonstrate that by incorporating advanced methods in NLP and ML, and by breaking from the mouse-based desktop paradigm, text entry methods can incorporate context and lower the interaction barrier to facilitate language production in both text entry[1] and editing[2,3].

ACCESSIBLE TOOLS FOR COMMUNICATING WITH NON-TEXTUAL INFORMATION

Although most text information in digital form is accessible to blind or low vision (BLV) users, there are still severe barriers for them to interact with non-textual information, such as emojis and stickers. **My research contributes tools that help BLV users to understand and interact with non-textual information**, improving equal access in computer-mediated communication.

Emojis have become an essential element of online communication, and are widely used in everyday social interactions. The visual form of emojis makes them inherently difficult for BLV users to understand and use, causing social exclusion. To address the problem, I first conducted **surveys and interviews** with BLV users to understand the challenges they faced when interacting with emojis. Based on my findings, I created a voice-based emoji entry system, *Voicemoji*[4], to facilitate emoji input and

[2] Mingrui "Ray" Zhang, He Wen, Jacob O. Wobbrock. (2019). Type, Then Correct: Intelligent Text Correction Techniques for Mobile Text Entry Using Neural Networks. *Proceedings of the 32nd Annual ACM Symposium on User Interface Software & Technology (UIST '19)*

[3] Mingrui "Ray" Zhang, Jacob O. Wobbrock. (2020). Gedit: Keyboard gestures for mobile text editing. *Proceedings of Graphics Interface (GI '20)*

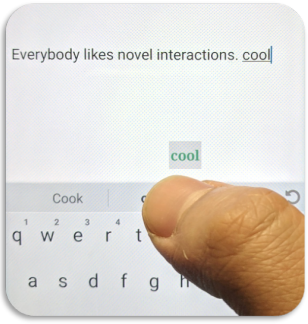

Figure 2. Drag-n-Throw lets the user drag a word and flick it into the general area of the erroneous word.

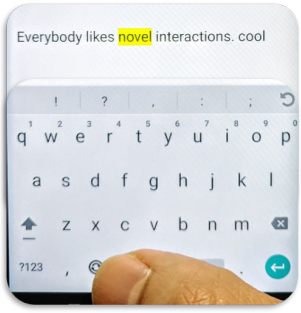

Figure 3. Pressing the Magic Key will highlight each possible error after the user types a correction.

[4] Mingrui "Ray" Zhang, Ruolin Wang, Xuhai Xu, Qisheng Li, Ather Sharif and Jacob O. Wobbrock. (2021). Voicemoji: Emoji entry using voice for visually impaired people. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '21)*

exploration. Voicemoji allows a user to query, input, and edit emojis via voice input with natural language, and provides context-sensitive emoji suggestions based on the current message (Fig. 4). To enter an emoji such as 🦯 *(man with probing cane)*, the user first issues a voice query such as, "a blind person emoji"; Voicemoji then searches over the internet and returns emojis with the closest descriptions. My studies showed that Voicemoji **reduced emoji entry time by 91.2%** over the current emoji keyboard, allowing BLV users to conveniently input, learn, and explore different emojis, enriching their daily communication experience. A feature like Voicemoji called "emoji transcription" now ships on the Google Pixel 6, which was released in Oct. 2021.

Figure 4. The interface of Voicemoji. The user inserts emojis through voice commands, and the system also provides relevant emojis suggestions based on the text.

Along with emojis, I also address accessibility challenges of another form of non-textual information: animated GIFs. Unlike static images, animated GIFs often lack adequate alternative text descriptions, and are not accessible to BLV users. By interviewing and co-designing with BLV users, I implemented *Ga11y* (pronounced "galley"), an automated GIF annotation system[5], which solves the annotation issue in two ways: **computer vision and crowdsourcing**. After a user issues an annotation request for a GIF, it is shown on a website, where volunteers can browse and provide annotations. If there is no human annotation available for a requested GIF, a description will be automatically generated through computer vision. This ensures that the user can get a timely response even if the GIF has not been annotated; on the other hand, the human annotations also reveal rich contextual and nuanced information for GIFs, which automated methods do not currently provide.

[5] Mingrui "Ray" Zhang, Mingyuan Zhong and Jacob O. Wobbrock. Ga11y: an Automated GIF Annotation System for Visually Impaired Users. *In submission*

MODELS, TOOLS AND STUDIES FOR EVALUATING INFORMATION INPUT SYSTEMS

As interaction researchers are creating novel experiences to facilitate the information input process, rigorous evaluations are needed to assess the quality, performance, and broader impacts of intelligent systems. Drawing on **Shannon information theory** and empirical user studies, I developed new input evaluation models and metrics, created a web-based tool called *TextTest++* for conducting evaluations, and generated an understanding of users' perceptions of intelligent prediction systems.

In nearly all human performance tasks, speed and accuracy trade-off against each other. Due to this trade-off, researchers often cannot draw firm performance conclusions when comparing two text entry methods. Therefore, I developed the model and mathematics for **text entry throughput**[6], a metric I derived directly from Shannon information theory, which unifies speed and accuracy into a single metric measured in bits/second. My key insight was to view **text entry as an information transmission process**, and the user and the entry method together as a **noisy channel**. From this perspective, I was able to quantify the throughput of the channel, thereby evaluating the text entry method in a way that **unifies speed and accuracy**. Through careful user studies, I

[6] Mingrui "Ray" Zhang, Shumin Zhai, Jacob O. Wobbrock. (2019). Text entry throughput: Towards unifying speed and accuracy in a single performance metric. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 19)*

showed that throughput was indeed the most stable metric across different speed-accuracy tradeoff conditions, and could serve as a robust performance metric.

Going further, to enable the evaluation of modern text input methods that use intelligent features such as auto-correction and word prediction, I updated the calculation of traditional error metrics by devising the **transcription sequence model**[7], and open-sourced the corresponding evaluation platform *TextTest++* (Fig. 5), which has since been used by Facebook Reality Labs, and in at least three other text entry and psychology research projects around the world.

I also conducted empirical studies to understand how input systems affect users. To investigate the impact of **intelligent emoji prediction** systems on online communication, I ran laboratory and field studies where participants composed text messages via a keyboard offering word- or semantic-level emoji predictions[8]. I found that the emoji suggestions reflected the message tones and aided users to better convey their intentions. While prior work has primarily focused on the role that emojis play in online communications, my findings shed light on how emoji suggestion mechanisms affect the communication experience.

FUTURE RESEARCH AGENDA

As an Assistant Professor, I will continue my goal of building and evaluating technology to facilitate human-machine communication. Specifically, I see opportunities and challenges for information communication systems with two trends: 1) the emergence of non-traditional computation platforms such as AR/VR and wearable devices, 2) the wide deployment of AI-infused information systems on commercial products.

What does a future information input interaction look like? The ubiquitous computing era is here, and devices today evolve rapidly into various form factors. We therefore need to design input interfaces to make them easy to use in mobile and wearable settings and support cross-device input tasks. As a first step towards this goal, I recently implemented a wearable text entry solution, *TypeAnywhere*[9], which allows a user to perform QWERTY **typing on any surface** without a hardware keyboard (Fig. 6). I will continue exploring this "ubicomp design space" with other modalities such as voice, gestures, and gaze.

Moreover, with powerful hardware processors and advanced machine learning algorithms, our devices are now able to detect our activities from multiple data sources. In the future, I plan to build **personalized context-aware input systems** by utilizing off-the-shelf sensor data. For example, a prediction system that provides replies with different linguistic styles based on the app the user is using (e.g., email vs. message apps), or different contents based on a user's current activity (e.g., suggesting locations when the user is using a map when driving).

[7] Mingrui "Ray" Zhang, Jacob O. Wobbrock. (2019). Beyond the Input Stream: Making Text Entry Evaluations More Flexible with Transcription Sequences. *Proceedings of the 32nd Annual ACM Symposium on User Interface Software & Technology (UIST '19)*

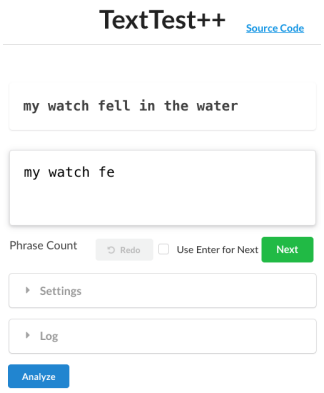

Figure 5. The TextTest++ interface for modern text entry evaluations.

[8] Mingrui "Ray" Zhang, Alex Mariakakis, Jacob Burke and Jacob O. Wobbrock. (2021). A comparative study of lexical and semantic emoji suggestion systems. *Proceedings of iConference 2021*

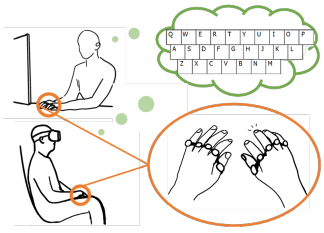

Figure 6. TypeAnywhere allows the user to type on any surface with finger-worn devices.

[9] Mingrui "Ray" Zhang, Shumin Zhai, Jacob O. Wobbrock. TypeAnywhere: A QWERTY-Based Text Entry Solution for Ubiquitous Computing. *In Submission*

How to design accessible tools for emerging forms of information? Today's world is full of new forms of information beyond traditional text and images, including videos, interactive content on mobile apps, and AR/VR content. However, most of these emerging forms are inaccessible to people with disabilities such as motor, vision, or hearing impairments. I plan to address this challenge in two ways: 1) to support the *understanding* of the information: Similar to my Ga11y work[5], I will combine machine learning and crowdsourcing to provide explanations of non-textual information sources. 2) to support the *interaction* with the information: Intuitive and practical interactions can be approached through *participatory design*, which gains insights from target users' experiences. I am applying the design methodology to an ongoing project which lets BLV users easily input text on large touch screens[10].

[10] Mingrui "Ray" Zhang, Jacob O. Wobbrock. BlindTyping: Text Entry Without a Keyboard on Touch Screens. *In Preparation*

How to design human-in-the-loop interactions for intelligent information systems, and how to evaluate their broader impacts? Advances in AI have empowered commercial systems to intelligently present, recommend and collaborate with users to interact with information. Those systems, including predictive text systems such as Gmail Smart Reply and conversational agents such as Alexa, are widely deployed and used by billions of users. However, while the emergence of Artificial Intelligence-Mediated Communication (AI-MC) presents clear practical benefits on communication efficiency, their broader impacts, such as how they affect a user's mental models and linguistic routines, and how they are perceived in various social contexts, still need to be investigated. I plan to design metrics that take psychological and communication theories to evaluate intelligent language production techniques. Furthermore, most intelligent algorithms are deployed as black boxes, lacking transparency to the user. I am interested in designing human-in-the-loop interactions to allow end-users to understand and adjust intelligent systems, including deriving active-learning algorithms for personalization and building user-feedback mechanisms to tune underlining models. By understanding, evaluating, and designing intelligent information interaction systems, I aim to realize the vision of Human-Computer Symbiosis for AI-MC.