We Are the Robots: Tapping Into the Lived Experiences of Wizards of Oz

Mafalda Gamboa
Interaction Design, CSE
Chalmers University of Technology, University of Gothenburg
Gothenburg, Sweden
mafalda.gamboa@chalmers.se

Sofia Thunberg
Interaction Design, CSE
Chalmers University of Technology, University of Gothenburg
Gothenburg, Sweden
sofia.thunberg@chalmers.se

Patricia Alves-Oliveira
Robotics
University of Michigan
Ann Arbor, Michigan, USA
robopati@umich.edu

Meagan B. Loerakker
TU Wien
Vienna, Austria
meagan.loerakker@tuwien.ac.at

Abstract

After 40 years of use, Wizard of Oz is still a popular method, particularly amongst Human-Robot Interaction researchers. In this Late-Breaking Work, we look into how the lived experiences of those performing the wizardry are represented in literature, and combine it with first-person vignettes from two experienced wizards. These two accounts surface a variety of ethical tensions and issues with the practice of wizarding—reflecting on their impact on the method itself. Through this research, we plan to identify what skills and sensibilities are honed through the practice of emulating imagined social robot agencies. By exposing these ideas, we seek to find other wizards who would like to contribute with their lived experiences to our call to understand and make explicit the obscured dimensions of Wizard of Oz felt by the researchers behind the curtain.

CCS Concepts

• **Human-centered computing → HCI design and evaluation methods**.

Keywords

Wizard of Oz, Lived Experiences, Human-Robot Interaction, First-Person Methods

ACM Reference Format:
Mafalda Gamboa, Sofia Thunberg, Patricia Alves-Oliveira, and Meagan B. Loerakker. 2025. We Are the Robots: Tapping Into the Lived Experiences of Wizards of Oz. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3706599.3720149

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
CHI EA '25, Yokohama, Japan
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3720149

1 Introduction

Along with low-fidelity prototyping, researchers have found ways to explore futures when current technology is not yet stable enough to be tested reliably [5, 22, 36]. The aim is to find potential issues and opportunities by taking a leap of faith. A popular method to this end is Wizard of Oz (WoZ) [8]—a deceptive approach in which study participants are convinced they are interacting with an automated system, when, in truth, a human is controlling the machine.

In this Late-Breaking Work (LBW), we look critically into what has been consistently obscured in the reporting of WoZ studies: the wizards themselves. While we are not the only ones turning the attention to the wizard [4, 7, 18], we suggest focusing not only on their performance but on how the method affects them and how they affect the method through their positionality. We argue that the lived experiences of wizards should be further investigated. Each wizard's sensibilities and their imaginaries of robots are potentially mirrored in their practice—which will influence the research results and perpetuate certain views on the future of robotics. In short, we seek to (a) identify the skills and sensibilities of a 'good' wizard along with the ethical risks of the practice to the wizards themselves, and (b) identify which assumptions about the future of artificial beings that are made material through their performances.

First, we structure the paper by presenting a short description of the WoZ-method, some reflections on its implications, followed by recent examples of its application. Paired with this, we offer first-person vignettes by two wizards. We close with a discussion on revisiting the issues of WoZ, while re-directing and re-surfacing the value of the experience of those who have extensive practice of the method.

1.1 What is Wizard of Oz

WoZ-method was first formulated in 1984 (then called the *OZ* paradigm [17] inspired by [15]), and was mainly used for studies in Natural Language Processing [e.g. 8, 23]. The method takes its name from the classic novel 'The Wonderful Wizard of Oz' by Frank Baum, where the giant head that appears to be a powerful wizard is just an ordinary man pulling levers behind a curtain. Throughout the years, the method has been picked up by fields like Human-Computer Interaction (HCI) and Human-Robot Interaction (HRI).

The main argument for using this method in HRI is to simulate *natural behaviour* that autonomous robots of today are *unable to perform*. The wizard, fully or partially, controls the robots' behaviours (e.g. speech, movements, lights, gestures, sounds), mediated through the robot, from a different physical space than where the interaction is taking place. This often means that the wizard experiences the world through the robot: sees through the cameras, and hears through the microphones. They also actuate by monitoring the joints, triggering preprogrammed sequences of movements, and writes or speaks what the robot is saying. In a nutshell, the wizard is an invisible puppeteer of the robot.

1.1.1 Reflections on the Method in the Literature. In a systematic review on how WoZ is used in HRI, Riek [27] found that few studies reported wizard error, wizard training, or wizard recognition. She proposed new reporting guidelines for the field, including questions about the robot, user, and wizard to demonstrate validity. Other guidelines have focused on the design of the WoZ setup, including a control interface allowing the wizard to freely trigger or fix robot behaviours instead of using a fixed protocol, and wizard action logging [21]. Despite these efforts, the guidelines are seemingly rarely used in the HRI community.

There are a number of research projects admittedly considering how wizarding is a difficult task and attempting to support it [e.g. 11, 26]. Porcheron et al. [26] offers a detailed description of the wizard toolkit and how a control setup can be managed through interaction between coordinating participants, researchers, and a robot vacuum cleaner. While this highlights the *complexity of the task*, especially when systems fail, it also points out that the wizard needs to *'maintain the fiction'* [26]. This notion relies on the wizard to not act as a human conversational partner, but rather as an algorithm by following a protocol to keep the simulation honest and respond in a manner that closely aligns with what could reasonably be implemented [23]. Another study by Brown et al. [4] discusses that the wizard's role in itself is part of learning how users wish to interact with robots in public spaces. Here, the authors instructed the wizards by asking them to interact *'naturally'* with the users, and gave them the flexibility to choose by themselves how to operate the robot [4]. In the field notes to this study, tensions with disclosure were reported, and seemingly circumvented by re-directing bystander's questions to an offline conversation to avoid disclosing the deception [6]. Mitchell and Mamykina [24] report on a WoZ study in the wild with expectations of 24/7 connection, 'which would put undue burden on the wizard due to a biological need for sleep and food, as well as other familial and academic obligations.' [24, p. 3]. However, this wizard remained anonymous and was only referred to as 'working as a full-time PhD student', offering no further reflections on their role [24]. These examples show how the experiences of the wizards should not be neglected in research, including how they keep the fiction, interpret the notion of interacting naturally (whatever that may mean), and deal with the burden of the task. This resonates with Suchman's argument that robotics discourse seems to indicate a displacement of an individualist conception of agency with a relational one so much as to displace the biological individual with a computational one [30]. She illustrates with her encounter with the Kismet robot, developed by Breazeal [e.g. 3]:

> 'The contrast between my own encounter with Kismet and that recorded on the demonstration videos makes clear the ways in which Kismet's affect is an effect not simply of the device itself but of Breazeal's trained reading of Kismet's actions and her extended history of labors with the machine. In the absence of Breazeal, correspondingly, Kismet's apparent randomness attests to the robot's reliance on the performative capabilities of its very particular "human caregiver".' [30, p. 246]

Kismet's agency is created from sociomaterial arrangements, through the caregiver. Often, descriptions of the wizards seem detached from the human realities of their lives (the wizard is described as a machine or animal) and does not account for the impact of practising the method, such as dealing with difficult work conditions, the hours spent wrangling (labours with the machine), or the cognitive load of maintaining the fiction.

1.2 The Importance of First-person Lived Experiences of Wizards

Within feminist theory of science, it is fundamental to acknowledge the situated knowledges at play [16, 31]. For example, through the notion of standpoint theory, we find reasonable argument to be more attentive to the experiences and backgrounds of researchers themselves [2]. WoZ is a method particularly sensitive to this, as it involves one or more researchers performing highly demanding tasks (as described above) in a manner that will both mirror their own previous experiences but also have an impact on their lives and research paths. This calls for further attention to the lived experiences of wizards, which we aim to support in this paper.

First-person methods and approaches have gained traction in HCI—which has naturally led to, for example, workshops [13, 20] and a special issue [9]. As researchers, our personal histories and experiences inevitably shape our work, influencing it both epistemologically and materially. Within HRI, a lot of the work happens in tinkering with robots, familiarising ourselves with the technology at hand, and creating a personal practice of wrangling [29]. Takayama suggested during a keynote at the HRI'22 conference that we should study robot wranglers [32]. While it is of value to research wrangling and wizarding practices through mainstream methods of user research, much of the depth of their experiences are more likely to be articulated through their first-person own accounts.

While many researchers question whether first-person methods can be rigorous enough to constitute credible knowledge or research— we argue instead that without acknowledging, recognising, and reflecting on the wizard, the method is not rigorously described. Dealing with robots becomes an everyday encounter for wizards. It is necessary to weave the wizard's mundane experiences into the reporting of the method. Approaches such as autoethnography have already demonstrated their potential to provide rich, authentic insights into our lived experiences with technology. For example, Gamboa [12] exposes intricate tensions of living with a number of toy robots in her home, alongside her children. The use of the method allowed her to give an honest account of the intimate and conflicting issues the presence of those robots created.

Simultaneously, these methods introduce distinct challenges. When researchers immerse themselves in articulating their own experiences, they may confront blurred boundaries between personal and professional identities, often requiring a level of vulnerability that could feel daunting. However, this vulnerability is part of the process [25]. Furthermore, others portrayed in those stories or involved in those studies should be considered carefully, such as in the example above [12]. The practice of wizarding offers a unique opportunity to develop such a sensibility towards the development of robots—what Garrett et al. [14] calls 'felt ethics'. To exemplify some of these tensions we present stories from two wizards where some of these complex feelings surfaced.

2 Stories of Two Wizards

To unpack the lived experiences of wizards, the first author interviewed the second (see interview guide in Appendix A). Whereas the first author (Mafalda) is a specialist in first-person methods, the second (Sofia) is a well-oiled wizard. After this interview, both authors interviewed a second wizard and third author (Patrícia).

The questions were centred on eliciting memories of specific stories or events, focusing on the felt dimensions exhibited by the wizard. The interviews tapped into the earliest and most recent memories of wizarding, alongside considerations for what makes a 'good' (or conversely 'bad') wizard, and lasted about one hour each. The stories are told in the first person written by the authors, as the interviews served primarily as an elicitation method to aid the second and third authors in articulating their experiences. However, in future studies, the stories of wizards can be told and analysed in other ways.

2.1 Wizard 1: Sofia

I describe my role as a wizard, first as a journey of using wizarding as a proof-of-concept, to a more restricted research environment, up until today, where I am questioning using this method again.



Figure 1: NAO robot flirting on the talent show as described by Sofia.

2.1.1 Wizarding for Entertainment. I have been the wizard hundreds of times during lab and school visits, on television, on radio, on stage, on exhibitions, on fairs and in public spaces. My first experience was with the NAO robot. I was introduced to the robot only a week before my first wizard experience. I recall the robot looking deep into my eyes during this first encounter—a common reaction to NAO—and finding myself unwilling to stop interacting with and relating to this artificial being, who seemed to have agency. This was of course the result of good wizarding. A week later I was behind the robot on the show Sweden Got Talent where four friends and I participated. The production team was keen on emphasising that it was the first time robots competed 'independently' in a televised talent show. Backstage, when the other performers were mingling and preparing, we shot some background stories where the robots were cute and flirtatious with people, saying things as 'I like you, are you single?' and 'You have very soft hands' when the host helped the robots on stage (Figure 1). I enjoyed making people laugh through those robot characters.

Another time, I was the wizard of a Pepper robot in an apartment, showcasing for an expo in a new neighbourhood for several weeks. I sat behind a curtain in the apartment, in a room where guests were not allowed, but I could still hear the conversations outside. Pepper was located most of the time right outside but could also move around in the apartment and on the balcony. This was my first experience wizarding Pepper this extensively in a public setting, and I had a setup where I could see through the robot's cameras, write what the robot could say, trigger gestures, and use an Xbox controller for navigation. When someone entered, the robot would wave 'hello' and my go-to question became to ask their name. When they answered, I would repeat their name. That sold them the idea that the robot was autonomous—'Oh it can remember me'. Then I would have long conversations with people, approach some and say something spontaneous about what they were talking about and ask for help, like getting the robot from the balcony.

I was having so much fun interacting with people and demo what robots *are able* to do. I even tricked friends with a background in robotics that we had gotten Pepper to work this well. I started to feel connected to Pepper: the robot's limitations were my limitations. I received feedback from the few that knew about the wizarding that *I was a really good robot*: Perhaps my way of interacting through the robot matched people's imaginary of how robots could act.

A few years later, a local museum had sold two date nights with the Furhat robot and wanted me to wizard for the night. The setup at the museum was that I was located behind a one-way mirror with Furhat placed on a table in front of the mirror. The first date was with two men in their 60s and the second date was with two teenage girls. The museum host had not prepared what the date would be about and I was very anxious before the guests arrived. Sitting in the dark room behind the mirror, I watched the guests sit down at the table facing Furhat (and me), and then the museum host left for over an hour. It was now up to me to make this an interesting date. With NAO, one can always make the robot dance, fall and stand up again, sneeze, laugh and do something funny, but with Furhat, one is limited to talking. I mostly demonstrated the robots' features by using different faces and voices, and asked them personal questions. The older men focused on asking questions about AI, future possibilities of technology development and limitations about the robot, which was easy to entertain. But the girls did not say anything. After 10 minutes I had run out of features to show off. I was now in panic mode since they had paid for this

date and probably expected more then what they were given. Eventually during both dates I let the robot say it needed to sleep and turned off the mask, and stepped out of the room. The confusion, disappointment and—in the case of the girls—shame I faced, was horrendous but I no longer wanted to hide behind the robot and found it unethical to not disclose the wizarding.

2.1.2 Wizarding for Research. During my bachelor thesis, I conducted my first study with a WoZ design [34]. In this experiment, I controlled the robot from outside a room with the door ajar so I could overhear the conversations in the room. I monitored NAO and Pepper respectively through their cameras, and prompted their speech while some autonomous features were activated. After the study, I told the participants that I had been using WoZ and I noticed a process starting within me where I found it more and more difficult to declare the design setup. I felt like I had fooled them, pulling away the curtain and showing it was all a lie. I sensed a disappointment in my participants. Not that any of them withdrew from the study or questioned my design, but *I was so used to being the magician, rather than the wizard*, that I did not feel comfortable in this situation any more.

The last time I was wizarding an experiment, I used Furhat and investigated older adults' perception of the robot [33]. I had developed a control system for the robot through several iterations and had an overview of the camera feed, speech, facial expressions and gestures. Unfortunately, the sound feed did not work properly so I was located outside of the room where the interaction took place but with the door open to overhear the conversation. We quickly came to the understanding that our participants were sceptical toward the robot and explicitly participated to express that they did not like robot development. As I declared the dreadful message of me being the wizard after the study procedure, several of the participants seemed positively surprised and expressed that it was good that development had not come further for the robot to be autonomous. However, this did not decrease my stress level or anxiety. I often relied on my research partner to deliver the message.

2.1.3 Moving Away from Being the Wizard. I belonged to a group that actively wanted to deceive people in believing in robots, the idea of robots as part of our world, tricking them that robots were far more functional and developed than they were. Even when asked a direct question about the wizard—e.g. 'There is someone controlling the robot right, this level of autonomy cannot be true?'—we answered with some level of uncertainty.

When being faced with the research context, I delved into the ethics behind the wizarding and was struck by my embodied reaction, realising that I was actually impersonating the robot, ending my desire to make robots work. That realisation moved me towards rather wanting to share the truth of the state-of-the-art of robots. During the last past years I have emphasised the limitations of robots instead of trying to entertain.

2.2 Wizard 2: Patrícia

I describe my role as a wizard in a long-term project researching robots in a school environment (see Figure 2). The stories I share, come from the development of the robot behaviours through a restricted wizarding protocol with limited sensors and visibility [28],



Figure 2: Patrícia sitting behind the screens wizarding the robot.

followed by an equally intensive study where the pre-defined robot behaviours were tested, and finally some remarks on the last phase where I observed the robot I had developed interact autonomously [1].

2.2.1 Creating the Robot. When I first started wizarding, it was during a school project in Lisbon. I vividly remember my first experience—it was intense and overwhelming. I was stationed in a small room next to the classroom, completely cut off from the interaction happening outside. I remember just being in panic because I did not know what was going on. I had limited feedback from the interaction between the robot and the students as a fundamental feature of the experiment design. The stress was so high that I found myself clicking randomly on behaviours for the robot, hoping the kids would not disengage. That is when we decided to add a 'panic button' to the interface, something I could press just to release the tension without it impacting the robot's behaviour.

Looking back, those early days were all about trial and error, both for me and the interface I was working with. It was the first step in creating an empathic robot tutor, a character meant to facilitate learning rather than delivering knowledge. I created all the behaviours for the robot. It was supposed to be an empathic robot tutor to facilitate knowledge more than delivering it. Drawing from my background in psychology and education, I poured over strategies and content that could inspire this robot's character—practically embedding parts of me into the robot.

2.2.2 Being the Robot. As we moved forward, the focus shifted to refining the behaviours I had designed. These were longer interactions with groups of children, where the robot took on the role of an educational tutor. I had created a detailed 'character bible' for the robot. When I wrote a sentence that the robot would say, that sentence would convey non-verbal behaviours. Everything, from gazes to gestures, had to be pre-coordinated to make the robot appear seamless and natural. But even with preparation, being a wizard was not easy. The experience was a mix of high stakes and constant decision-making under pressure. It is like you are trying to interact with someone that does not speak the same language as

you, so you overdo it. It was all too easy to second-guess my choices, as I could never be sure if the interaction was truly working.

Interestingly, as I triggered different behaviours for the robot, I began to feel like I could recognise individual kids based on how they interacted. In a weird way, I felt like I could distinguish the participants even without knowing them. It was as though the patterns of interaction revealed their personalities, and I started forming a connection with them through the robot. Even though I never met the children directly, this sense of familiarity stayed with me, adding another layer to the experience.

2.2.3 What Matters for Wizarding. The final stage was observing the robot deployed autonomously with the behaviours I had crafted. This was the culmination of months of effort, but it brought its own surprises. It was like seeing a movie about what I created before, it felt surreal to see my work come alive in an autonomous robot.

The robot interacted with the same children weekly for two months, building a rapport that I had not anticipated. When the two months were over and the robot said, 'This is our last session together,' some children became visibly sad. It felt kind of unfair to elicit an interaction that is pleasant, and then to say goodbye and remove the robot. What struck me the most was how the children created their own rituals with the robot. They started doing peace signs and other gestures that were not part of the programmed behaviours. I observed that I did not program that, but it showed bonding. Seeing this emotional connection made me realise the depth of the relationship that had been formed, but it also left me questioning the ethics of such interactions.

3 Discussion

The aim of our LBW is to start the conversation on the future of WoZ studies, through unpacking two first-person accounts. The richness of just two stories is already a testament to the importance of this work. We suggest future work to complement already existing guidelines [21, 27] with a dimension that takes into account the input from the wizards themselves.

3.1 Felt Ethics between Disclosure, Disappointment, and Distress

The experiences of Sofia and Patrícia show the process of *being and becoming a robot*, through learning how to wrangle robots, but also how to create custom fictions or personalities for each wizarding interaction. This is a practice of care [14] which we cannot fully unpack in this LBW. Sofia believes that it is a good experience for students to become wizards as it will force them to face those issues first hand [23]. Initially, Sofia found a lot of joy in the wizarding practice. While this is food for thought in future work, we can already note how the hesitancy in engaging in further wizarding expressed by Sofia, along with Patrícia's questioning of the ethics of such studies, are important to lift to the limelight. Being a wizard is portrayed both as an enjoyable process of honing a sensibility towards the implications of the method, but also a practice that brings concerns towards lack of genuineness of HRI. In short, the practice of wrangling and wizarding may help to attune researchers to the potential ethical issues they could be dealing with in a manner that is embodied and felt as an emotional resistance—something that can hardly be reproduced only through procedural ethics assessments.

Sofia also mentioned how she would be more open to being a wizard again if she would not be responsible for disclosing the deception, or if she could transmit some of the embodied shame of deception to other entities such as Large Language Models (LLMs) or other tools of assistance—offloading morality to technological systems. Indeed, Artificial Intelligence (AI) tools like LLMs are being developed for wizards [e.g. 10, 35]. There is evident danger in the fact that the ethical issues do not disappear per se, but if they are not *felt* by a wizard, they may more easily fade to the background. Patrícia is convinced that these technologies cannot replace a wizard entirely—that LLMs are just a tool. She remarked: 'A human wizard brings timing, emotional depth, and context that AI struggles to replicate. Ultimately, wizarding is about creating meaningful connections, and for that, you still need a human touch.' It begs the question if the continued need of the wizard is a sign that robots should not have social roles. If a wizard is always required to achieve a meaningful connection, but also 'align with what could reasonably be implemented' [23], it is sensible to wonder if the method should continue being applied at all, reinforcing the imaginary of such a far-fetched possibility.

The stories of these two wizards show that a WoZ study must include detailed and rich descriptions of the positioning and experiences of the wizards themselves. Without this, the method assumes an undefined methodological dimension which otherwise profoundly compromises the described results. The increasing focus on positionality within HCI is a strong indicator of recognition of the researcher's role as a point of rigour [2]. It is not difficult to imagine that disclosing Sofia's sensibilities towards wizarding are relevant to the reporting of her future projects. Yet, HRI studies employing the WoZ methodology fail to systematically report the wizards' experiences and perspectives [27].

3.2 'Good' Wizards

Both Patrícia and Sofia mentioned a plethora of skills needed to be a 'good'—meaning well performing—wizard of social robots. Both agreed it is a job that requires someone to be calm. Patrícia noted: 'Wizarding is not just about operating a robot; it requires a specific mindset. A good wizard must make decisions under pressure and know the tools and tasks intimately. It requires a large capacity for preparation.' Sofia mentioned that these skills can only be honed with practice—she mentioned how training in improvisation theatre would likely help in staying alert, being creative, and improving the ability to deal with unexpected situations. And these two cases only refer to a human per robot case. In other cases, the complexity of the robot—e.g. Sophia [19]—requires multiple people to be able to handle the task. In that case, one can assume that capacity to collaborate is also fundamental.

These notions tie back to the idea of wrangling as a pre-requisite for a good wizarding practice. Practical knowledge of the capacity of the machines is not only part of the method but an important facet of the first-person accounts. According to Sofia and Patrícia, this intimate connection takes time to achieve, and it is unreasonable to assume just anyone can become a 'good' wizard by following a

protocol—hence wizarding should be seen under the lens of best fit rather than best practice.

Ultimately, there is a need to unpack what 'good' means. While a well-performing wizard needs all these skills, one important point is how through their practice, they encountered issues with the ethics of their performance. Perhaps, based on these two examples, a 'good' wizard is one that no longer wants to do their magic if it means social deception or results in distress.

3.3 Future Work: Wizards Unite!

The two stories we tell scratch only the surface of the immense depth and variety of embodied experiences wizards have. Being a machine, or impersonating a robot, appears to be a difficult and transformative experience. Our two wizards revealed issues with stress and, at times, poor working conditions. Wizards seem to have to face ethically frail situations which they may or not be prepared to handle appropriately. Working as a wizard is not exclusive to research: at Prosper a butler robot is controlled for making 'trickier tasks' while training data of households is gathered. There are also initiatives for people with disabilities or mental health issues who are housebound but can work as robot wizards to help them stay on the job market. One example is the Dawn Café, where robots fulfil different tasks (e.g. greeting customers, taking orders and delivering food).

Considering humans are working as professional robots, it is imaginable we could collectively formulate a bill of rights for wizards. We suggest as a thought exercise to fund the fictional *Union of Wizards of Oz*, to foster conversations between wizards and articulating what are the characteristics of the work. To that end, we imagine workshops, interviews, and observations with wizards, researchers or otherwise, to share experiences and expectations.

4 Conclusion

While WoZ studies are fairly common, the focus is seldom turned towards the wizard themselves. Through two detailed first-person accounts of two knowledgeable wizards, we suggest the need for future work in articulating the lived experiences and felt tensions in this practice. We propose turning the limelight to the ethical issues surfaced, making explicit the missed opportunity in troubling both the method and its impact on those practising it. Ultimately, in a time of great advancements in AI, it is urgent to consider what notions of the future role of robots are being perpetuated in our epistemological tools.

Acknowledgments

This work is partially funded by the Wallenberg AI, Autonomous Systems and Software Program — Humanity and Society (WASP-HS).

References

[1] Patrícia Alves-Oliveira, Pedro Sequeira, Francisco S Melo, Ginevra Castellano, and Ana Paiva. 2019. Empathic robot for group learning: A field study. *ACM Transactions on Human-Robot Interaction (THRI)* 8, 1 (2019), 1–34.

[2] Shaowen Bardzell. 2010. Feminist HCI: taking stock and outlining an agenda for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, GA, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1301–1310. doi:10.1145/1753326.1753521

[3] Cynthia Breazeal. 2004. *Designing Sociable Robots*. The MIT Press.

[4] Barry Brown, Fanjun Bu, Ilan Mandel, and Wendy Ju. 2024. Trash in Motion: Emergent Interactions with a Robotic Trashcan. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 591, 17 pages. doi:10.1145/3613904.3642610

[5] Jacob T. Browne. 2019. Wizard of Oz Prototyping for Machine Learning Experiences. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3290607.3312877

[6] Fanjun Bu, Alexandra W.D. Bremers, Mark Colley, and Wendy Ju. 2024. Field Notes on Deploying Research Robots in Public Spaces. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 162, 6 pages. doi:10.1145/3613905.3651044

[7] Andreas Butz, Annika Kaltenhauser, and Malin Eiband. 2020. The Expert of Oz: A Two-sided Study Paradigm for Intelligent Systems. In *Companion Publication of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS' 20 Companion)*. Association for Computing Machinery, New York, NY, USA, 269–273. doi:10.1145/3393914.3395874

[8] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz Studies - Why and How. In *Proceedings of the 1st International Conference on Intelligent User Interfaces*. 193–200.

[9] Audrey Desjardins, Oscar Tomico, Andrés Lucero, Marta E. Cecchinato, and Carman Neustaedter. 2021. Introduction to the Special Issue on First-Person Methods in HCI. *ACM Transactions on Computer-Human Interaction* 28, 6, Article 37 (Dec. 2021), 12 pages. doi:10.1145/3492342

[10] Jingchao Fang, Nikos Arechiga, Keiichi Namikoshi, Nayeli Bravo, Candice Hogan, and David A. Shamma. 2024. On LLM Wizards: Identifying Large Language Models' Behaviors for Wizard of Oz Experiments. In *Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents* (Glasgow, United Kingdom) *(IVA '24)*. Association for Computing Machinery, New York, NY, USA, Article 16, 11 pages. doi:10.1145/3652988.3673967

[11] Juliano Franz and Derek Reilly. 2017. TangiWoZ: A Tangible Interface for Wizard of Oz Studies. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 2584–2590. doi:10.1145/3027063.3053254

[12] Mafalda Gamboa. 2022. Living with Drones, Robots, and Young Children: Informing Research through Design with Autoethnography. In *Nordic Human-Computer Interaction Conference* (Aarhus, Denmark) *(NordiCHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 27, 14 pages. doi:10.1145/3546155.3546658

[13] Mafalda Gamboa, Claudia Núñez Pacheco, Sarah Homewood, Andrés Lucero, Janne Mascha Beuthel, Audrey Desjardins, Karey Helms, William Gaver, Kristina Höök, and Laura Forlano. 2024. More Samples of One: Weaving First-Person Perspectives into Mainstream HCI Research. In *Companion Publication of the 2024 ACM Designing Interactive Systems Conference* (IT University of Copenhagen, Denmark) *(DIS '24 Companion)*. Association for Computing Machinery, New York, NY, USA, 364–367. doi:10.1145/3656156.3658382

[14] Rachael Garrett, Kristina Popova, Claudia Núñez Pacheco, Thorhildur Asgeirsdottir, Airi Lampinen, and Kristina Höök. 2023. Felt Ethics: Cultivating Ethical Sensibility in Design Practice. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 1, 15 pages. doi:10.1145/3544548.3580875

[15] John D Gould, John Conti, and Todd Hovanyecz. 1983. Composing Letters with a Simulated Listening Typewriter. *Commun. ACM* 26 (1983). Issue 4.

[16] Donna Haraway. 1988. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies* 14, 3 (1988), 575. doi:10.2307/3178066

[17] J. F. Kelly. 1984. An Iterative Design Methodology for User-Friendly Natural Language Office Information Applications. *ACM Transactions on Office Information Systems* 2 (1984), 26–41. Issue 1.

[18] Heike Christiane Kotsios, Philipp Michael Markus Peter Asteriou, and Philipp Wintersberger. 2024. Only Trust a Hidden Wizard: Investigating the Effects of Wizard Visibility in Automotive Wizard of Oz Studies. In *Proceedings of the 16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Stanford, CA, USA) *(AutomotiveUI '24)*. Association for Computing Machinery, New York, NY, USA, 74–82. doi:10.1145/3640792.3675738

[19] Nikolaos Kouravanas and Andreas Pavlopoulos. 2022. Social Robots: The case of Robot Sophia. *Homo Virtualis* 5 (may 2022), 136–165. Issue 1. doi:10.12681/homvir.30320

[20] Andrés Lucero, Audrey Desjardins, Carman Neustaedter, Kristina Höök, Marc Hassenzahl, and Marta E. Cecchinato. 2019. A Sample of One: First-Person Research Methods in HCI. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion* (San Diego, CA, USA) *(DIS '19 Companion)*. Association for Computing Machinery, New York, NY, USA, 385–388. doi:10.1145/3301019.3319996

[21] Nikolas Martelaro. 2016. Wizard-of-Oz Interfaces as a Step Towards Autonomous HRI. http://robotwebtools.org
[22] Danića Mast, Alex Roidl, and Antti Jylha. 2023. Wizard of Oz Prototyping for Interactive Spatial Augmented Reality in HCI Education: Experiences with Rapid Prototyping for Interactive Spatial Augmented Reality. In *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery. doi:10.1145/3544549.3573861
[23] David Maulsby, Saul Greenberg, and Richard Mander. 1993. Prototyping an intelligent agent through Wizard of Oz. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands) *(CHI '93)*. Association for Computing Machinery, New York, NY, USA, 277–284. doi:10.1145/169059.169215
[24] Elliot Mitchell and Lena Mamykina. 2021. From the Curtain to Kansas: Conducting Wizard-of-Oz Studies in the Wild. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 46, 6 pages. doi:10.1145/3411763.3443446
[25] Kristina Popova, Rachael Garrett, Claudia Núñez Pacheco, Airi Lampinen, and Kristina Höök. 2022. Vulnerability as an ethical stance in soma design processes. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 178, 13 pages. doi:10.1145/3491102.3501994
[26] Martin Porcheron, Joel E. Fischer, and Michel Valstar. 2020. NottReal: A Tool for Voice-based Wizard of Oz studies. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (Bilbao, Spain) *(CUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 35, 3 pages. doi:10.1145/3405755.3406168
[27] Laurel Riek. 2012. Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. *Journal of Human-Robot Interaction* 1 (2012), 119–136. Issue 1. doi:10.5898/jhri.1.1.riek
[28] Pedro Sequeira, Patrícia Alves-Oliveira, Tiago Ribeiro, Eugenio Di Tullio, Sofia Petisca, Francisco S Melo, Ginevra Castellano, and Ana Paiva. 2016. Discovering social interaction strategies for robots from restricted-perception Wizard-of-Oz studies. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 197–204.
[29] Aaron Steinfeld, Odest Chadwicke Jenkins, and Brian Scassellati. 2009. The oz of wizard: simulating the human for interaction research. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction* (La Jolla, CA, USA) *(HRI '09)*. Association for Computing Machinery, New York, NY, USA, 101–108. doi:10.1145/1514095.1514115
[30] Lucy Suchman. 2006. *Human-Machine Reconfigurations: Plans and Situated Actions* (2 ed.). Cambridge University Press.
[31] Lucille Alice Suchman. 2007. *Human-machine reconfigurations: Plans and situated actions*. Cambridge university press.
[32] Leila Takayama. 2022. Putting Human-Robot Interaction Research into Design Practice. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction* (Sapporo, Hokkaido, Japan) *(HRI '22)*. IEEE Press, 1.
[33] Sofia Thunberg, Maria Arnelid, and Tom Ziemke. 2022. Older Adults' Perception of the Furhat Robot. In *Proceedings of 10th International Conference on Human-Agent Interaction ('HAI22)*. 4–12.
[34] Sofia Thunberg, Sam Thellman, and Tom Ziemke. 2017. Don't judge a book by its cover: A study of the social acceptance of NAO vs. Pepper. In *HAI 2017 - Proceedings of the 5th International Conference on Human Agent Interaction*.
[35] Tom Williams, Cynthia Matuszek, Ross Mead, and Nick Depalma. 2024. Scarecrows in Oz: The Use of Large Language Models in HRI. *ACM Transactions on Human-Robot Interaction* 13 (jan 2024). Issue 1. doi:10.1145/3606261
[36] J.D. Zamfirescu-Pereira, David Sirkin, David Goedicke, Ray LC, Natalie Friedman, Ilan Mandel, Nikolas Martelaro, and Wendy Ju. 2021. Fake It to Make It: Exploratory Prototyping in HRI. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 19–28. doi:10.1145/3434074.3446909

A Interview Guide

The following guide served as a prompt for the two vignettes portrayed in the Late-Breaking Work. The interviews were transcribed but served only as an elicitation method for the interviewees to write their own accounts in the first-person. Hence the results are not analysed as a traditional interview, but the interview is instead a method to prompt reflections.

A.1 Research Questions

(1) What are the skills necessary for a good wizard in Wizard of Oz studies?
(2) What are the assumptions about social robots wizards make in their performances?

A.2 Interview Guide

(1) Introductions
(2) Goal of this work
    - The goal of our work is to better understand what are the characteristics of a good wizard in Wizard of Oz studies. To this end, we are interviewing experienced wizards in order to gather their insights and lived experiences of being in such a role.
    - Along with this work, we seek to find what are the assumptions made by wizards on the nature of social robots.
(3) Recording of the session
    - Do you give permission to record this session so it can be transcribed as a basis for the accounts you will write yourself?

A.3 Starting Questions

(1) How would you define Wizard of Oz as a method?
(2) How would you describe how extensive your experience is as a wizard?

A.4 First Wizarding Experience

(1) What was your first wizarding experience?
(2) How were you recruited?
(3) Can you tell me more about the set-up? (Context, who else was involved, etc)
(4) What did you learn?
(5) What character/personality were you playing?
(6) Is this study published?

A.5 Wizarding Stories (Repeat according to time)

(1) Would you mind telling us about another story?
(2) How were you recruited?
(3) Can you tell me more about the set-up? (Context, who else was involved, etc)
(4) What did you learn?
(5) What character/personality were you playing?
(6) Is this study published?
(7) Why did you pick this story to tell us?

A.6 Wizarding Skills

(1) In your opinion, what are the skills necessary for a good wizard?
(2) How do you inform the robot behaviour you are enacting? (is the behaviour based on assumptions, sci-fi tropes, AI developments, human behaviour?)
(3) How knowledgeable are you on current AI developments?

A.7 Complementary Questions

(1) Do you know any other wizards that may be interesting for me to interview?
(2) Are there other Wizard of Oz studies you would recommend to us?