

Case Report

Jan Schwarzer, Julian Fietkau* and Susanne Draheim

Exploring body tracking sensors in longitudinal ambient display studies in the wild

<https://doi.org/10.1515/icom-2025-0038>

Received September 29, 2025; accepted November 13, 2025;

published online November 28, 2025

Abstract: Body tracking sensors (i.e., depth cameras) such as Microsoft Kinect have been utilized in ambient display research for more than a decade. They facilitate a deeper understanding of phenomena occurring throughout interactions, aid the investigation of ambient displays within a broader context, and effectively complement existing qualitative methods such as on-site observations. Although these sensors have made significant contributions to research, there are still challenges with regard to data collection and analysis, particularly in light of recent advances in artificial intelligence. Further research is needed into how these sensors can contribute to a better understanding of how ambient displays are used in practice long term and how they can help to develop this understanding. In this article, we expand on the potentials and limitations of body tracking sensors that we experienced in our own in-the-wild research. To this end, we present insights from a small fleet of long-term, real-world installations of ambient displays we manage, which incorporate multiple body tracking sensors. The present article is concluded with a discussion of future directions for the field. In particular, the present study explores the potential contributions of body tracking sensors to recent methodological developments in the field of Human-Computer Interaction.

Keywords: ambient displays; body tracking; field studies; long-term research; quantitative methods; sensor deployments

1 Introduction

In recent years, there have been unprecedented changes in the field of Human-Computer Interaction (HCI), especially relating to data-driven applications in the wake of the ongoing hype surrounding applied artificial intelligence (AI) and its countless methodological developments, particularly in the field of machine learning (ML). Additional advancements include Extended Reality (XR) technology, namely Virtual Reality (VR) and Augmented Reality (AR), Internet of Things (IoT), blockchain, big data, pervasive or ubiquitous technology, and more.^{1–3} These technological advancements are fundamentally changing the way people interact with each other, raising the question of how we can understand their wider societal implications.⁴ Although traditional interaction issues remain relevant and warrant further research, many contemporary issues in HCI, such as ethics, privacy, and security, are in fact related to this shift to data intensity.⁴ As Weiser⁵ envisioned in the early 1990s, the modern world is now characterized by rapidly advancing ubiquitous technology,⁶ with which humans interact in their natural, built, and synthetic environments. Public and semi-public displays, or *ambient displays* as we refer to them, are part of this development. A wide variety of applications exist today, including examples, for instance, to visualize the production and consumption of electricity stemming from photovoltaic panels in private households,⁷ to highlight work progress in agile software development teams,⁸ to ease access to public transport systems,⁹ and to combine large-display installations with AR technology to address perspective distortion.¹⁰

One important area of HCI that can help us to understand how technological artefacts are used and adopted is behavior tracking. On a technological level, the term *behavior tracking* more broadly refers to the ability to provide researchers with in-depth insights into, for example, how people move around or collaborate. While the overall number of approaches to behavior tracking observed in HCI is increasing,⁴ the same is true for the domain of ambient display research. The first studies investigating

*Corresponding author: Julian Fietkau, University of the Bundeswehr Munich, Munich, Germany, E-mail: julian.fietkau@unibw.de.

<https://orcid.org/0000-0001-7264-8496>

Jan Schwarzer and Susanne Draheim, Hamburg University of Applied Sciences, Hamburg, Germany,

E-mail: jan.schwarzer@haw-hamburg.de (J. Schwarzer),

susanne.draheim@haw-hamburg.de (S. Draheim).

<https://orcid.org/0000-0002-6788-5072> (J. Schwarzer).

<https://orcid.org/0000-0001-7515-7473> (S. Draheim)

user behavior surrounding display installations were conducted during the 2010s. They focused on issues such as scrutinizing pedestrian traffic in public spaces [e.g., 11, 12] and on students' walking behavior around university canteen installations [e.g., 13], to name but a few. The main technical tools for this purpose were, and are to this day, body tracking sensors (i.e., depth cameras) from various vendors (e.g., Microsoft Kinect and Stereolabs ZED). These sensors utilize different approaches, such as time-of-flight, stereo vision, or structured light, to this end. Due to their cost efficiency in comparison to manual (human) observation, their ability to be readily adapted to other deployments, and the fact that they are easily integrated with other methodologies,^{11,12} body tracking sensors have become integral to this category of research in recent years. Notably, body tracking sensors have been used in the natural and situated environments of ambient display installations. Research in real-world places, or *in the wild* as we refer to following Williamson and Williamson,¹³ embodies an important paradigm shift in HCI and experiences increasing attention in the community.¹⁴ However, research in the wild is a messy and complex endeavor.^{13,15,16}

Some of the issues we touch on in this article are reflected in the seven grand challenges for the HCI community.⁴ These include, for instance that current sensor systems, big data analyses, and ML methods in general still need to fully adjust to human needs with respect to supporting and enhancing human abilities.^{4,17} Further limitations include a scale shortage in current research endeavors, a lack of long-term insights from real-world environments, and an insufficient theoretical background.^{18–20} Equally, appropriate evaluation and validation techniques remain an open issue.^{4,19} Applied to ambient display research, we still have limited knowledge of how to approach the methodological aspects of longitudinal *in-the-wild* research, particularly in light of recent advances in body tracking sensor technology and AI in general. Admittedly, data-driven applications and ML have made significant progress in the field of HCI in recent years. However, we need to establish a foundation with regard to the application of body tracking sensors in longitudinal research spanning multiple years. We argue that further fundamental work is necessary to enable the ready application of ML methods in our domain. Important overarching questions surround aspects such as automation and the combination of quantitative and qualitative methods into a solid theoretical framework, what type of data to collect, as well as which supervised and unsupervised ML approaches are fruitful candidates. From our perspective as HCI scientists, we would like to raise awareness of the challenges we have experienced in our research and the

opportunities we see for our community going forward. To this end, we concentrate on three central research questions in the present article:

1. How can we make effective use of modern body tracking sensors with their specific feature sets in mind?
2. In what ways do specific usage patterns manifest themselves in the collected data?
3. What are useful algorithmic means to analyze the data?

By doing this, we aim to highlight the complexities of *in-the-wild* research at different phases using body tracking sensors. While some issues concern aspects prior to or throughout data collection (e.g., hardware limitations), others draw attention to the foundations of the analysis process (e.g., automatically distill usage patterns). We do not claim that the three outlined questions elaborated on are exhaustive in any way, nor that they reflect the challenges other researchers might face. However, they summarize key issues that we came across in our own research.

The article is organized as follows: In Section 2, related literature is introduced. The different body tracking sensors employed in our research are presented in Section 3, while Section 4 elaborates on the various experiments and field research that we have conducted. In Section 5, we discuss the experimental and methodological outcomes in more detail and present guidance for future research, and Section 6 concludes this article.

2 Related work

This section highlights recent developments in the use of body tracking sensors in ambient display research, concluding with a reflection on the status quo.

2.1 State of the art

The field of computer-supported cooperative work (CSCW) has a long history rooted in the analysis and design of technology for collaboration, which would be much too expansive to summarize here. For a literature overview of how collaboration using ambient displays can be experimentally evaluated in general (not focusing on body tracking sensors), see Mateescu et al.²¹ The work to integrate the current wave of AI technologies into ambient display evaluation studies is still ongoing, but we have begun to see examples such as Atta et al.²²

For this article, we will focus our attention on the study of ambient display deployments using body tracking technology. In the following, selected examples for usage of body tracking sensors in real-world environments are presented.

Some of these examples represent short-term efforts, while others have used body tracking sensors in their work for up to a year. The most seminal work in this context is the study by Williamson and Williamson.¹¹ The authors analyzed pedestrian traffic in front of a public display installation by using a camera positioned three stories above the walkway on which the display was located. The principal motivation was to understand how such technology changes public spaces, how it is being used in authentic contexts, and how different interaction styles actually work. A custom computer vision based tool incorporating a variety of different diagnostic and visualization techniques was used to analyze the data. While data of over 900 pedestrians were collected, the study was very short-termed, as the data collected amounted to only about 4 h of video material. However, the work of Williamson and Williamson¹¹ paved the way for the use of camera sensors to track people's movement in front of display installations, thereby inspiring subsequent research in the following years. A few years later, Williamson and Williamson¹³ revisited their approach, now concentrating on the question of how experimenter interventions affect the evaluation process. Specifically, the authors sought to find out, how different types of experimenter presence (e.g., no visible presence of investigators in contrast to proactively intervening during interaction) introduce what kind of bias. The authors collected a total of 4 h of video data for each of the three types of intervention investigated, including more than 5,000 passers-by in total. A Microsoft Kinect v2 camera was installed to capture data right in front of the display, while another camera was placed 15 m above and 15 m behind the display installation to collect overhead video material. OpenNI libraries were used to analyze the data from the Kinect sensor and the authors utilized their own custom tool mentioned above to investigate pedestrian traffic. In addition, the authors collected data from manual observations and interaction logs. Williamson and Williamson¹³ found, for instance, that the presence of an observer significantly reduced interactions with the display installation. Fundamentally, they encourage the systematic control of experimenter roles in evaluations and the use of high-quality measurements such as pedestrian traffic data to quantify the observer effect. In the same year, Elhart et al.²³ introduced and evaluated a similar custom tool for tracking audience mobility. Their research motivation was similar to that of Williamson and Williamson,¹¹ while emphasizing that only a few low-cost tools exist to capture spatial and temporal behavior. In their setup, the authors used a Microsoft Kinect device and utilized a combination of open source computer vision and web visualization techniques (among others OpenNI and OpenCV).

The camera and display were installed in front of a university canteen. The tool itself was evaluated using 14 videos, each 5 min long. Subsequently, the tool was used over 52 days to collect and analyze data from approximately 41,000 passers-by. Their findings include the fact that, for instance, the highest number of passers-by occurred during lunchtime and that most people spent no more than 4 s in front of the display. Elhart et al.²³ believe that the main strength of their sensor-based approach lies in providing additional information on aspects such as content transitions and touch interactions. Another example is the study by Mäkelä et al.,¹² which, to the best of our knowledge, is the first long-term investigation of a real-world display setup using body tracking sensors. The authors' motivation concerned the overall process of data collection and analysis of depth-based camera data, resulting in a semi-automatic process to study public displays. Their research is based on data collected over the course of a year using a Microsoft Kinect camera in a university setting. This data includes information on over 100,000 passers-by. The introduced process consists of four principal phases: data collection, preparation, feature extraction, and analysis. While the first two phases are straightforward and can be automated to a large extent, the latter two phases require the most manual work, such as determining research questions. The authors ran analyses using Microsoft Excel and SPSS. In their setting, Mäkelä et al.¹² found that, for example, over 90 % of users were passive users (i.e., people who were not actively engaging with the installation). They also revealed that users entering the sensors's field of view from the front were significantly more likely to become direct users (i.e., people who interacted with the display). Overall, their study made notable contributions on a methodological level.

2.2 Research gaps and implications

We revisited the aforementioned studies and conducted a comprehensive literature review. The aim was to find new research on audience behavior in the wild that uses body tracking sensors in ambient display studies. While we initially set out with a forward reference search and an evaluation of resulting papers, we subsequently also focused more broadly on relevant literature repositories such as ACM Digital Library and IEEE Xplore. This process revealed that, to date, the studies by Elhart et al.²³ and Mäkelä et al.¹² are the only ones to have attempted longer-term research into this topic. While all of the studies mentioned here greatly informed our research, we find this fact surprising for two meaningful reasons.

First, given the progress made in recent years, more accurate and feature-rich cameras are available today. As

these newer cameras can produce richer insights (e.g., detection of a higher number of people simultaneously, a wider field of view, and generally better detection accuracy), they could also assist greatly in improving our understanding of a display's surrounding environment. Alongside qualitative methods such as interviews and observations, these sensors could help us to investigate group constellations, collaborative exchanges, and effects occurring during interaction (e.g., the honeypot effect) in more detail. By doing so, modern body tracking sensors could help to shed light on the wider implications of ambient displays in real-world settings. We could better reflect on when displays installations are being used and how, as well as how they fit into the existing toolset architecture of every modern company. To us, it seems that research in this area has stalled – i.e., transitioning from academic feasibility studies to real-world endeavors that utilize the full range of qualitative and quantitative methods. We believe that foundational research is necessary to pioneer methodological approaches in order to develop new, disruptive theories that can advance our field at its core. Such endeavors certainly require a great deal of time and resources, but we believe they are worthwhile. In other words, we think that our field requires more holistic and long-term research not only expanding on methodological questions, but also touching on the real-world implications of display deployments.

Second, despite the recent advancements in AI, it seems that our field is not capitalizing on them. In addition to the more traditional approaches, such as clustering and tree algorithms, many variants of neural networks are now widely available. Some of these are specifically designed to work with body tracking data, such as Graph Convolutional Network (GCN) models, and enable new approaches to analyzing large amounts of data. Instead, it seems that a large part of the HCI community's focus has shifted towards working with technologies such as AR and VR. Yet, we still do not know how ambient displays are utilized in practice, nor how we can even develop this understanding. We encourage researchers to experiment with AI's capabilities to gain a better understanding of its potential contributions. For example, supervised learning approaches could be explored to identify ways to automate certain aspects of qualitative work such as coding procedures. Others may explore ways to leverage GCNs to identify patterns of interaction in body tracking data. Overall, we believe that there is significant potential that remains unused, given the advances in AI and the need to handle large volumes of data produced by body tracking sensors.

3 Body tracking sensors

Building on the research gaps and implications mentioned above, we now turn our attention to the deployment of body tracking sensors in our research. First, we highlight how these sensors have been used and for what purpose. Second, we then provide a brief overview in terms of how these sensors work and operate. Finally, we present one custom-built tool that we use in our research for visualizing and analyzing the corresponding body tracking data.

3.1 Deployment settings

All of our past and current deployments have focused primarily on investigating long-term usage. The methodological toolset available at the different points in time has affected these individual investigations. For example, when we began our first multi-year deployment in 2014, we initially relied on interviews, observations and touch interaction logs. Building on this, we developed a fundamental theory about how the display installation had been used by agile software development teams back then.²⁴ Our motivation was driven by a discrepancy between the ideas proposed in the literature so far – such as the notion that ambient displays encourage communication and collaboration – and how these promises actually manifest in practice over time, if at all.

In 2016, we conducted our first experiments using a Microsoft Kinect v2 sensor and began intensifying our work with body tracking sensors. We quickly realized the opportunities that this new technology presented, both in amending existing methods and in contributing rich new nuances to the overall research. It suddenly became easy to expand on interactions in an installation's vicinity, thereby reducing the need for field observations. While these sensors undoubtedly have their limitations, such as a limited field of view, they nonetheless enable us to gain an initial understanding of a display's surroundings. The results of our initial experiments were published in an article in 2022.⁸ Over the coming years, we have set up different display deployments in both Hamburg and Munich. Our research matured in terms of, for example, specific research questions and the experience of conducting research in the wild. Overall, we have run deployment experiments mainly in two different semi-public contexts: a software development company and a university, both in Germany. For both contexts, we placed one or more interactive ambient displays into a room which is not open to the public at large, but which many

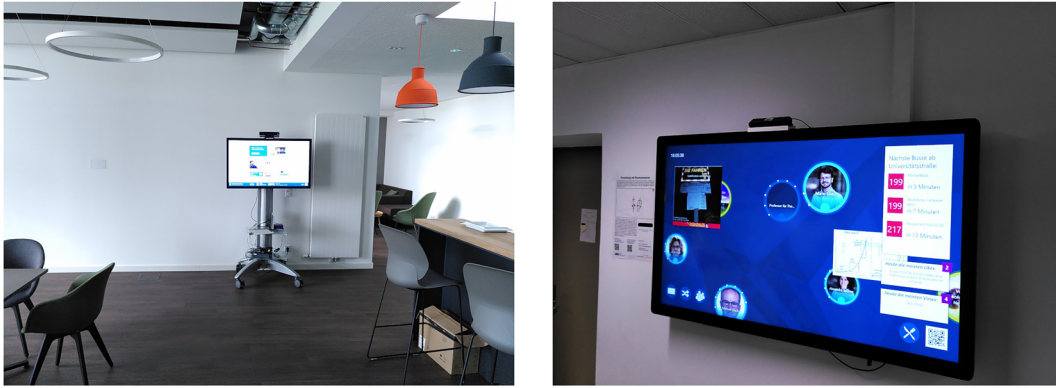


Figure 1: Two photos of representative ambient display deployments with body tracking sensors. Left: software company deployment, using a Microsoft Kinect v2 sensor. Right: university deployment, using a Stereolabs ZED 2 sensor.

people with access (company employees/university students) would pass by on foot every day. In fact, the amount of expected foot traffic was the main criterion for our placement decisions. For example, Figure 1 on the left shows one of our two currently deployed installations in the aforementioned company's New Work café, which is visited by many people during the day.

Each deployment consisted of a screen (rarely also more than one) with a body tracking sensor attached to the top of its frame. In this way, experiments revealed that the sensors could perform the body tracking task of the area right in front of them the best. Again, our goal is to capture and understand the usage of ambient displays “in passing” (as opposed to prolonged, focused use), and all facets of the deployment – not only the display placement, but also the interactive software running on the device – were oriented towards this goal. The detailed purpose and contents of the ambient displays are beyond the scope of this article, but further information can be found in Schwarzer et al.⁸ and Koch et al.,²⁵ respectively.

3.2 Technical approach

Our attention now turns to the question of how these sensors operate. Optical body tracking sensors are the main instrument used for detailed analysis of behavior, focusing on anonymized body tracking models. The optical sensors used in our research (Microsoft Kinect, Stereolabs ZED 2) work roughly as follows: First, an image of the environment is digitized by the camera sensors. However, the image material is not recorded immediately, but instead a body tracking algorithm is applied in real time, which marks individuals with their body postures in the image. While the Kinect v2 sensor uses a random forest algorithm for this purpose, the ZED 2 relies on a neural network.

The manufacturers of the two commercial sensors we use do not disclose the exact details of their respective recognition methods. However, we can deduce the basic process from studies [e.g., 26]. Using Kinect v2 as an example, we would like to describe the functionality in more detail below. It can be assumed that the random forest algorithm integrated in Kinect v2 was trained with several hundred thousand images to ensure its functionality. The processing chain of the camera can be divided into three parts, visualized in Figure 2. First, the Kinect sensor collects depth images using infrared, in which each pixel contains depth information accurate to within a few centimeters. The advantages of depth images include the ability to cope with poor lighting conditions, being invariant to color, texture, and body shape, and the ability to synthesize realistic images of people. Second, classification algorithms are used to determine probabilistic pixel-based body regions. Some of these parts are defined in such a way that they directly locate specific body points, while others fill in the gaps or can be used in combination to predict other joints. Finally, the specific positions of body points are specified in three-dimensional coordinates. The previously determined pixel-based information regarding the body regions must now be integrated across all pixels in order to make reliable suggestions for the positions of the body points. For the Kinect v2, this procedure results in a total of 26 individual body points per person. The body points determined analytically in this way are recorded with their positions in space. As a result, the sensor technology provides relatively accurate data on the position, posture, line of sight, etc. of the persons in the spatial area in front of the screens. Although it is not possible to recognize specific people on the basis of these abstract body models, conclusions can be drawn with regard to recurring individual or group behavior.

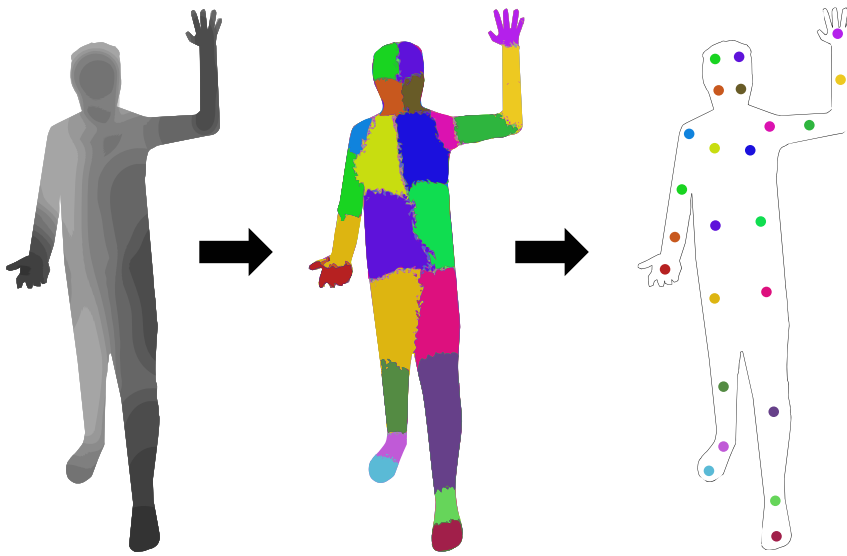


Figure 2: The body tracking algorithm used by the Microsoft Kinect v2 sensor, showing how a depth image gets processed into body regions and finally key points. Own visualization, inspired by Shotton et al.²⁶

The Stereolabs ZED 2 conceptually fulfills the same task, but there are nuanced differences in the technical approaches. It uses parallax depth detection based on two separate camera sensors instead of the Kinect v2's infrared technology. At a slightly higher off-the-shelf purchase price, it can support increased resolutions and frame rates as well as detect a maximum of 10 humans up to 20 m away compared to the Kinect v2's maximum of six humans up to 5 m away. The Kinect v2 has a fixed 25-point body model, while the ZED 2 supports several different body models with up to 38 key points. However, unlike the Kinect's, the ZED 2 software does not perform engagement estimations, which must be implemented by the data consumer if they are needed.

3.3 PoseViz

One major obstacle was the lack of an established format for storing and transmitting body tracking data. Existing data formats were either vendor-specific (e.g., Microsoft *Kinect Studio* recordings) or not suitable for stationary body tracking setups where passers-by may enter and leave the area of interest at any time (e.g., *Biovision Hierarchy* format). To be able to do non-trivial empirical work with body tracking data, we first had to design a format suitable for storing such data as well as transferring it in bulk or in real time, and then develop software tools to read, write, and visualize data in this format. With the goal in mind that future researchers should have as easy a time as possible to understand our recorded body tracking data if needed, we decided on a textual format that uses line-based fields

to delineate frames (specific moments in time) within a recording, persons within a frame, and key points (specific limbs and joints) within each person. This makes our format (dubbed *PoseViz*) fairly easy to parse algorithmically as well as to read in any text editor. In the process of designing the format, we implemented code to access our two sensor models' respective APIs and transform their body tracking data into our format to enable them to be stored and reviewed.

This gave us the ability to work on a browser-based playback software (see Figure 3) that also shares the name *PoseViz* with the file format itself. PoseViz is capable of reading one or more stored body tracking recordings, render them in a 3D visualization, and allow playback and

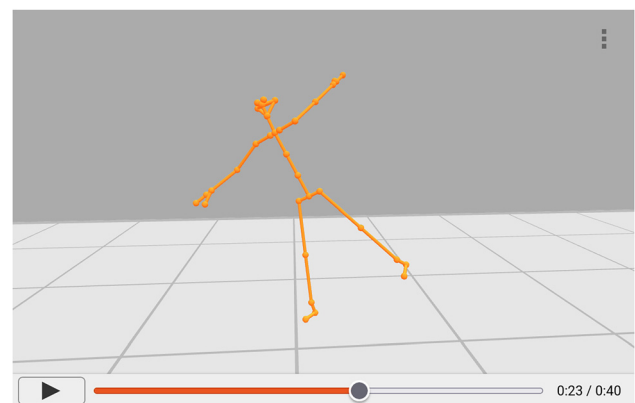


Figure 3: A screenshot of the PoseViz software showing a person moving inside an abstract 3D space with video-style playback controls at the bottom. Visit <https://poseviz.com/> to see this demonstration in motion.

scrubbing just like in typical video players. It is possible to look at interactions from different angles and to generally get an impression of the quality of the sensor data. PoseViz allows the user to toggle various display aspects including position markers, gaze estimations, and walking trajectories. The playback speed and rendering perspective can be adjusted, and PoseViz can display time-based annotations as well as engagement data over time, if such data is embedded in the recording. The planning and design process is described in more detail in Fietkau.²⁷ An interactive demonstration of PoseViz is available for free testing online.¹

Examining larger quantities of body tracking recordings for specific hypotheses requires additional bespoke tooling that can analyze the aspects relevant to those hypotheses. Our work examining two-dimensional walking paths as clustered time series data,²⁸ described further in subsection 4.3.2, serves as a practical example of the kind of quantitative insight that can be gained from body tracking data facilitated by tools like PoseViz.

4 Experiments and field research

We now draw attention to insights from experiments and field research we have carried out in the aforementioned settings. As outlined in the introduction, we focus on three central questions to this end.

4.1 How can we make effective use of modern body tracking sensors with their specific feature sets in mind?

To this day, we are using two types of body tracking sensors in our research: Microsoft Kinect v2 (released in 2014) and Stereolabs ZED 2 (released in 2019) sensors. Each camera has its own advantages and disadvantages. For instance, the ZED 2 sensor can detect up to 10 people simultaneously, whereas the Kinect v2 camera is limited to six people in total. However, the Kinect v2 sensor provides a specific feature which the ZED 2 sensor does not: To a greater or lesser extent, it can tell us whether or not a person is looking directly in its direction. Tests revealed that the Kinect v2 cameras successfully determined the looking-at gesture when mounted on top of each display (see Figure 1) even when we were looking at the displays instead. This aspect was fundamental to our research, as it enabled us to detect when people interacted with the displays in a passive way (e.g., by passing by without engaging with them

actively), as opposed to passing by inattentively. Overall, we refer to this behavior as *engagement* in our research. Before we started using body tracking sensors, our research relied heavily on touch interaction logs. This meant that, except from observations and interviews, we could not expand on passive interactions in the displays' surrounding areas. In the following, we shed light on how we replicated this detection mechanism of engagement with the ZED 2 sensor.

4.1.1 Parameters for engagement

We started by determining a list of body tracking features that, in combination, could allow us to establish a correlation with engagement.²⁹ We settled on the following list:

- Distance between a person and a display.
- Movement speed of a person (slowly walking or standing people are more likely to be paying attention).
- Body orientation (people facing towards the display are more likely to be paying attention to it).
- Gaze direction (same as before but measuring only the head instead of the full body).
- Direct interaction (people reaching or pointing towards the display are very likely to be paying attention to it).

Each of these five variables was measured, then clamped and normalized to a scalar value in the range between 0 and 1. For the distance and speed values, the input range was calibrated using practically sensible real-world values. For the body orientation and gaze direction values, the angular difference to the direction of the screen was calculated, with any angle $>90^\circ$ being assigned the value 0. The direct interaction value measured the time that people had their arms raised and pointed at the screen, with the value of 1 reached after 15 s. The engagement score for each person at each point in time was then calculated as the average of these five values. Accordingly, an engagement score of 0.8 indicates, for instance, that a person showed a higher level of engagement, while a score of 0.2 underlines the opposite.

4.1.2 Initial testing

We performed a pilot test to validate this method to calculate the engagement score.²⁹ To that end, 27 different constellations of individuals paying attention to the screen from different distances, at different movement speeds, and using different body and gaze angles were performed in front of our experimental installation. The resulting recordings were manually scored for their degree of attention and the

¹ See: <https://poseviz.com/>.

results compared with our engagement score. Based on this, we determined that there were still major issues with the measurement extraction code (e.g., inaccuracies in the measurement of directions), but that the correlation was good enough to prove the general feasibility of the approach.

Next, we took a sample from the non-staged body tracking data gathered from our long-term deployments for the sake of comparison. As expected, the real-world data was generally noisier and contained more distractions and irrelevant movements compared to our staged scenarios. To test the feasibility of large-scale asynchronous engagement scoring of body tracking data, we scored some 40,000 individual recordings. It was difficult to assess the validity of the scoring since there was not much accessible ground truth to compare it to, but the score distributions across the two deployment sites appeared plausible regarding their spatial circumstances (two screens in hallways predominantly used by passers-by, one with much more unaffiliated foot traffic than the other) and a brief qualitative analysis of randomly selected recordings revealed that the feature-based engagement score appeared to be generally suitable as an automated estimation for manually assigned engagement values. In a categorized comparison of the two measurements, we observed an average deviation of 15–20 % [29, Table 2]. The main conclusion of this experiment was that the engagement score could be a valuable instrument in determining attention from body tracking data without manual intervention.

4.1.3 Revised testing

To further refine the approach, we conducted a follow-up experiment, reported on by Filippov et al.,³⁰ in which the engagement score calculation was simplified to omit the gaze direction, which proved difficult to estimate accurately, and the presence of direct interaction, which unduly down-ranked passive but interested observers. The engagement score was thus calculated based on the body orientation, the distance to the screen, and the movement speed. Furthermore, this second experimental approach made a stronger effort to examine the body tracking recordings as time series data that has dynamics which can be analyzed instead of merely averaged out. By calculating the engagement score for each timestamp in a recording and plotting it against time, we can examine how a person's engagement changes throughout the recording. Detecting the local maximum allows us to identify individuals who paid attention for only a short period within a longer recording.

Similar to the initial test, we once again took a sample of real unsupervised body tracking data and manually labeled it on our engagement scale. Using this data as a

baseline, we were additionally able to derive a suitable engagement score threshold to differentiate between non-engaging passers-by and people who paid at least some attention to the installation. Testing the classifier with a different random sample from the data set, we arrived at an accuracy of just over 90 %, a notable improvement compared to the first iteration that further validates the feature-based engagement scoring approach.

4.2 In what ways do specific usage patterns manifest themselves in the collected data?

Alongside questions concerning interaction patterns, times of peak usage and information displayed at certain times, there is also the question of what specific usage patterns might occur and how they manifest in the collected data. The literature describes different patterns that can typically be observed in ambient display research such as the novelty effect³¹ and the honeypot effect.³² The novelty effect suggests that new technology is used more frequently in the period immediately following its deployment. The honeypot effect, at its most basic, refers to situations in which one person standing in front of a display installation attracts others to join them. In our work, we placed a strong emphasis on the honeypot effect because of the interesting collaboration constellations that it involves by definition. Arguably, the range of emerging collaboration patterns is vast. In the following, however, we expand on the work we carried out to investigate how the honeypot effect potentially manifests itself in the data. The overarching goal is to be able to automatically classify instances of the honeypot effect in the future.

4.2.1 Detecting the honeypot effect

To gain insight, we performed a study building on the earlier work on engagement measurement^{29,30} described in subsection 4.1, now looking beyond interactions by individual people and specifically examining constellations of multiple people appearing in the deployment areas simultaneously, in order to detect and classify instances of the honeypot effect.

Our investigation of empirical methods for the detection of instances of the honeypot effect is described by Bieschke,³³ in which the feature-based approach was extended to multi-person patterns and used to automatically filter for honeypot effect candidates in long-term body tracking recordings. Once again, a collection of archetypical honeypot effect situations was artificially enacted and recorded in front of a real interactive

installation as body tracking data. A manual examination of their commonalities followed by the iterative development of a feature-based classifier led to the conclusion that honeypot constellations always feature multiple people arriving at different times with overlapping presence windows (pre-condition), with everyone involved looking in the direction of the screen and approaching it for some amount of time. Using these criteria, a sample of 9,000 recordings covering one calendar month was classified into honeypot and non-honeypot scenarios, with five honeypot candidates emerging. These were subsequently verified through individual visual inspection and compared to a random sample of non-matches, substantiating the claim that the feature-based classifier can practically function as a honeypot constellation detector.

4.2.2 Limitations of this approach

It is worth noting that this approach as well as possible ML classifiers for honeypot constellations must contend with the fact that body tracking recordings give us access to people's movements, but not their intentions. According to the strict definition of the honeypot effect, the second person must give attention to the installation *because* someone else is already present. By analyzing body tracking data, we can only ascertain that someone else was already present, but we do not gain insight into people's actual motivations, which would require deeper qualitative analysis through methods such as interviews or on-site observations.

4.3 What are useful algorithmic means to analyze the data?

The focus now shifts to the analysis of the data itself. Over the last decade, our general analytical approaches have evolved and matured. We experimented with various methods, which we outline below. This development is also reflected to some extent within our community, as can be seen when we compare two fundamental studies that reflect on the current state of HCI.^{4,34} While AI played a rather minor role in their 2019 study, Stephanidis et al.⁴ make extensive references to it in the revised version from 2025.

4.3.1 Descriptive statistics

In 2017, we deployed a first Microsoft Kinect v2 setup and collected roughly 100,000 records over 4.5 months. The first challenge was to devise a way of analyzing this large amount of data. We asked ourselves fundamental questions about

how to approach this dataset, what to look for, the quality of the data, and others. Ultimately, inspired by related research [e.g., 13], we opted for descriptive analyses and published our results in an article.⁸ These analyses concentrated on spatial and temporal audience behavior using different visualizations and statistics to demonstrate our findings. For example, we identified the directions from which people approached the display installation. We also indicated the areas within the camera's field of view that showed the highest level of engagement.

Arguably, the most important lesson learned during this time was how to effectively collect, pre-process (e.g., filtering), and analyze depth-based camera data in a meaningful way. We spent a great deal of time visually inspecting the data and digging into its nuances. We also learned to accept some limitations of the Kinect sensor. For example, the camera sometimes accidentally lost tracking of people when they briefly left the rather narrow field of view. Additionally, the possibility that the camera will lose tracking of a person due to occlusion must be accounted for. Therefore, while these sensors could be vital components of a research design, their limitations must be taken into account.

In summary, this phase of our research was shaped by the fact of having some preexisting knowledge (e.g., potential ways to analyze the data) and assumptions (e.g., preliminary insights on usage patterns) of the data. This stemmed primarily from our extensive immersion in the data prior to conducting any analysis, as well as from our reading of related literature.

4.3.2 Unsupervised learning approaches: clustering algorithms

We gradually realized that we needed to shift away from making assumptions about expected results. To a greater or lesser extent, the goal was to replace at least some of the labor-intensive steps discussed above and to increase the level of automation during analysis. We focused on finding ways for algorithms to explain what we see in the data instead of us providing descriptive explanations. One of the central pieces of the analysis involved investigating groups of walking trajectories, which in our case are paths in a two-dimensional coordinate landscape that depict the movement of people from a bird's-eye view. At the most basic level, these trajectories show how people behave in front of a display installation, such as passing by or moving toward it. Walking trajectories therefore contribute important insights regarding passive usage. Although we found instances of similar walking trajectories in the data manually, we could not determine whether these were

representative at all or whether they were all potential instances in the data set.

As a result, we reviewed the literature to find ways to assist with this endeavor. Ultimately, we found that hierarchical clustering combined with dynamic time warping was an effective way to automatically group walking trajectories. Other studies, as we point out in our corresponding publication,²⁸ have used both algorithms to examine, for instance, flyways of birds³⁵ or household electric load curves.³⁶ We implemented the algorithms and ran evaluations on a subset of the data. Overall, we were able to categorize different types of walking trajectories into corresponding groups.

Although we experienced this unsupervised ML algorithm helpful, it has its limitations. One of the most obvious limitations is the computational performance of the algorithm: with datasets increasing in duration, calculating the results can quickly become practically untenable. Future work must address these performance issues for the implementation to be applicable to large data sets. Furthermore, because the nuances of the data gathered in the wild are so rich, there is often no clear distinction between two groups of walking trajectories, even when suggested by the algorithm. Additional work is necessary to fine-tune the clustering itself, as well as the similarity metric used in dynamic time warping.

4.3.3 Supervised learning efforts

While we are still working on refining the above-detailed clustering algorithm, we were also eager to find out if we could automatically identify patterns in the data that we already manually identified and were interested in. Hence, we were looking into supervised ML approaches. In our exploration of potentially suitable methodologies to detect and classify engagement and other underlying effects and sentiments in body tracking data, we did not want to limit ourselves to human-discernible features. Our initial foray into using supervised ML is documented by Lacher et al.³⁷ We posited that a ML model based on manually labeled training data may be able to perform the classification of individual body tracking frames into engaged and non-engaged states. In this cursory study design, tagged training data was generated by automatically labeling frames within a short time range as engaged surrounding any direct interaction event in the system logs. The downside is that passive engagement would not be accounted for by this approach, but on the upside, it would yield large amounts of highly reliable training data for active engagement. A model could then be trained using this data and used as a classifier for

future recordings. However, we ended up not carrying out this exact experiment on account of its suspected unsuitability for passive engagement.

Instead, the neural network approach was revisited by Ottenheim.³⁸ The study – as a bachelor thesis – investigated the applicability of GCNs for automating the interpretation of in-the-wild skeletal data. GCNs have produced promising results in the recognition of gestures and actions regarding skeleton data and are attracting increasing attention.³⁹ The study examines whether GCNs, in specific the Spatial-Temporal Graph Convolutional Network (ST-GCN), can effectively interpret gestures captured in uncontrolled environments. A dataset was created with data from one of our deployments, and a training environment was developed that incorporates transfer learning and data augmentation methods. The results show that GCNs can capture the spatial and temporal dynamics necessary for accurate gesture recognition in real-world scenarios and provide insight into the potential of GCNs to optimize automated gesture interpretation in heterogenous and uncontrolled in-the-wild environments.

The findings demonstrate that the ST-GCN model is effective in detecting specific gestures using in-the-wild skeleton data, thereby establishing its value as a tool for automating gesture recognition. Transfer learning has been demonstrated to be particularly advantageous, markedly enhancing model performance. Furthermore, the selection of optimizer, batch size, and weight decay is instrumental in attaining optimal accuracy. While pre-processing and data augmentation do exert an influence, it is less pronounced than initially anticipated. Ultimately, the complexity of the classification task, whether two-class or three-class, has a discernible impact on performance, with simpler tasks consistently yielding superior outcomes.

5 Discussion

Despite our general expansion of research opportunities through sophisticated sensor technology and large amounts of data, some of the traditional challenges remain when conducting research in the wild. This highlights the need for ongoing research into methodological approaches to deployment-based studies of digital collaboration tools and to empirical in-the-wild studies in general.⁴ In this discussion, we would like to take a step back and attempt to summarize our broader current view of the research field. At the same time, we would like to sketch further provisional ideas on how we, as HCI scientists, perceive ML and data-driven applications to be reshaping the landscape of conducting research

in our field. Ultimately, our discussion comes down to the central question of how these approaches can effectively support longitudinal, in-the-wild research involving body tracking sensors. In our view, there is no easy answer to this question. First, we point out the continuing importance of qualitative research approaches for in-the-wild HCI research. At the same time, the framework conditions for mixed methods in the post-COVID working world have become more complex due to the accelerated hybridization of work, as we explain in the second section. In the third section, we point out the fundamental qualitative and labour-intensive challenge of integrating ML approaches, a balancing act that must be very precisely tailored to the respective research purpose. In the last section, we discuss issues of research ethics related to process transparency for participants and the creation of the best possible anonymity in the analysis of sensor-based movement data.

5.1 The undisputed importance of qualitative data

Qualitative methods have always been and will always be crucial to any longitudinal endeavor seeking meaningful insights. We are convinced that only a holistic approach incorporating both quantitative and qualitative methods can provide rich and disruptive results. However, this adds complexity to processes such as data collection and preparation when compared to working solely with quantitative data from body tracking sensors, where these processes are described as straightforward.¹² We think that the data intensity of ML methods will not replace qualitative field work such as conducting on-site observations and interviews to a large extent. However, we do believe that ML can provide new, insightful nuances to an overarching research methodology (e.g., grounded theory) that aims to establish robust theoretical foundations for the field. For instance, ML methods can produce new abstractions of body tracking data that were previously impossible or would have required countless hours of manual labor. We believe that more research is necessary, focusing on the interplay of qualitative and quantitative methods, as well as how applied AI can assist effectively throughout this process to, as Stephanidis et al.⁴ point out, achieve robust assessments of relevant interactions. At its core, data itself can be conceptualized as a representation of a sociotechnical system, incorporating technology, social norms, and biases.⁴⁰ Undeniably, both qualitative and quantitative data add a rich layer of nuance to this picture. Depending on the research questions and on future developments, the focus may shift more toward one end of the methodological spectrum. However, we are

confident that qualitative methods will be integral to HCI field research in years to come.

5.2 Contextual implications: post-pandemic hybrid work arrangements

Simultaneously, new challenges have emerged regarding hybrid work practices post COVID-19. While some companies have returned to their pre-pandemic way of doing business (i.e., working in the office from 9 to 5), others are keeping the option of working from home available to their employees. In fact, working from home remains important for most employees, who have transitioned to a hybrid working model.⁴¹ As a result of this development, one of our research partners, for example, established the aforesaid New Work café to encourage communication between on-site employees and to make the office space more appealing in general. Hybrid work presents some unique challenges to in-the-wild research, although ethnographic approaches are certainly taking up this challenge and methodologically re-exploring the field of hybrid organizations, including innovative documentation methods and media formats.^{42–46}

For example, certain in-office practices have been replaced by digital alternatives (e.g., online meeting formats), making them difficult to investigate and understand in the first place, as well as their wider impact. Furthermore, at certain times during the week, there may only be a few employees in the office. This could make it impossible to collect meaningful data using body tracking sensors. When deploying HCI artefacts, researchers must therefore anticipate audience fluctuation or change. Although issues relating to hybrid work only indirectly touch upon the topic of data-driven applications, it is nonetheless crucial to consider them, as they may determine whether an HCI research project is successful or not.

5.3 Algorithmic choices, a researcher's headache

It is also important to remember that none of the supervised or unsupervised algorithms presented here are a universal solution, nor are they all-encompassing for their corresponding research questions in any sense. For instance, there are still performance-related issues (e.g., clustering compute complexity), we do use the algorithms for specific purposes (e.g., grouping similar walking paths), and a notable amount of manual work is still necessary (e.g., pre-processing steps). There is clearly more work to be done, but these approaches are helpful for our research. We believe that we have now reached a point where we can use ML algorithms in a more meaningful way with rich, long-term

body tracking data. Our extensive experience of conducting research in the field over several years plays a significant role in this observation (e.g., understanding the research contexts, having on-site contacts, resolving hardware issues, and more). We have become increasingly accustomed to using different algorithmic methods such as clustering algorithms and neural networks such as GCNs. The central question is how to capture complex social behavior by considering a multitude of data sources, such as body tracking sensors, and which ML algorithms are appropriate for this purpose.

Notably, systems capable of ingesting full video feeds have recently been emerging in the field of large language models (LLMs) such as VideoLLM-online.⁴⁷ These approaches promise semi-automated interpretation of video data. However, in their current form, they would be unsuitable for our context as we need to guarantee privacy by prohibiting the use of full video for empirical analysis. The abstraction and removal of personally identifiable information from the raw data was one of the reasons that led us to body tracking sensors to begin with (see the following section). Nonetheless, video LLMs are an interesting emerging application of AI technology and their applicability to different contexts may improve in the future.

5.4 Ethical considerations

Apart from the technical concerns relating to body tracking data, ethical and regulatory challenges also deserve attention. Ethical concerns generally arise from the collection, usage, and storage of data.^{48,49} As Stephanidis et al.⁴ vividly summarize, AI allows for the use of personalized data in many unforeseen ways. The ostensible anonymity of body tracking data, with its absence of physically identifying features as one would see in video recordings, can fall apart when you consider ways to identify individuals from their specific movements (e.g., gait analysis,⁵⁰ characteristic gestures) or interactions with external data, such as correlating people's presence in body tracking recordings with vacation dates or lab sign-in sheets. Even though we had initially planned to publish the raw body tracking data recorded at our deployment setups, a deeper investigation of the potential for deanonymization caused us to reverse that decision and publish only summarized statistical data. This is also why we practice transparency when collecting data with body tracking sensors during field deployments, such as through handouts and discussions. We believe that ethical concerns will become increasingly relevant as AI is adopted in HCI research.

Another ethical issue arises when we consider body tracking sensors and their vendor-specific limitations. We must not underestimate the normative role of body tracking models in the data collection phase. Simply by virtue of categorizing image areas into “humans” and “not humans”, the way the body tracking process works entails making specific assumptions about what constitutes a “valid” human body, which, depending on the inference approach, may be implicit or even completely unknown. For example, we can conjecture that deviations from the average human body (e.g., usage of mobility aids such as canes or wheelchairs, limbs that have atypical proportions or that are missing altogether, generally atypical body shapes) may lead to a higher rate of detection errors and thereby introduce data integrity errors rooted in accessibility and inclusivity. In the most egregious outcome, people who do not sufficiently “fit” the training data may be quietly and unintentionally omitted from recordings. Improvements on this issue would be predicated on, for example, more open and inclusive training data sets for image recognition, something that would require significant resources and likely additional regulation of industry efforts. We have discussed these challenges in more depth in Fietkau and Schwarzer.⁵¹

6 Conclusions

In this article, we have reported on our connected in-the-wild deployment studies with body tracking sensors. We have summarized the different data-driven methodological approaches that our experiments have pursued, covering manual feature-based analysis as well as the use of ML classifiers. Following our report, we have discussed several challenges encountered in the course of these experiments, which have not been conclusively solved, but for which our approaches may offer guidance for future experiments of a similar nature.

At time of writing, the research community is experimenting with AI methods in a huge variety of contexts. Our experiences suggest opportunities for AI/ML methods to assist in evaluating quantitative sensor data in a way that can reduce the burden on researchers, but only with newly developed tooling for specific questions. For example, our evaluation of 2D walking trajectories required the development of bespoke software tooling on top of the general tools for recording and visualizing body tracking data that we had already built. The same has been the case for each of our individual research questions regarding human behavior.

In the future, a fundamental analysis of work scenarios in the wild must increasingly include support for hybrid work, which has become a regular working mode in many

companies. Our future research will develop and evaluate a methodological framework to improve understanding of collaboration in authentic hybrid work environments. This framework, briefly outlined in Schwarzer et al.,⁵² will focus on automation (i.e., interpretations based on algorithms and ML models) and data triangulation (i.e., a range of research methods) to understand the wider implications of the evaluated technologies. We also intend to utilize digital and on-site ethnographic approaches, combining methods to observe location-based, remote, and hybrid work activities, and to learn how these modes of interaction influence each other, and what types of interaction may emerge. Building on the insights from our concluded research project,⁵³ we intend to continue to approach the empirical work through the technological context of ambient displays. These artifacts follow the leitmotif of “physical windows” into digital spaces, such as bidirectional camera setups for real-time collaboration between teams, or “metaphorical windows” that display contextual information to promote insights into remote work. At the same time, our research will focus on a specific hybrid work practice: coordination between agile software development teams or cross-team coordination.

Acknowledgement: This article summarizes the activities from multiple long-term research projects. The authors would like to thank all the participating parties, especially the cooperating company, as well as all the students who contributed empirical work.

Research ethics: Not applicable.

Informed consent: Not applicable.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Use of Large Language Models, AI and Machine Learning

Tools: The authors used the *DeepL Translator* tool to identify spelling errors and gather wording suggestions. AI and machine learning methods were used in several of the above experiments as described in Section 4. Generative AI/large language models were not used.

Conflict of interest: The authors state no conflict of interest.

Research funding: Much of the work presented here was part of research project “Investigation of the hon-eypot effect on (semi-)public interactive ambient displays in long-term field studies,” which was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, DOI: 10.13039/501100001659) – project number 451069094.

Data availability: Because body tracking data carries a moderate risk of deanonymization, the raw data for these experiments is subject to privacy regulations. Full or partial access

may be granted by the corresponding author upon request depending on legal evaluation.

References

1. Clinch, S.; Ward, J. A. Augmented Cognition. *IEEE Pervasive Comput.* **2023**, *22*, 6–7.
2. Kraus, S.; Kumar, S.; Lim, W. M.; Kaur, J.; Sharma, A.; Schiavone, F. From Moon Landing to Metaverse: Tracing the Evolution of Technological Forecasting and Social Change. *Technol. Forecast. Social Change* **2023**, *189*, 122381.
3. Mamykina, L.; Epstein, A. D.; Klasnja, P.; Sprujit-Metz, D.; Meyer, J.; Czerwinski, M.; Althoff, T.; Choe, E. K.; De Choudhury, M.; Lim, B. Grand Challenges for Personal Informatics and AI. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*; ACM, 2022, pp. 1–6.
4. Stephanidis, C.; Salvendy, G.; Antona, M.; Duffy, V. G.; Gao, Q.; Karwowski, W.; Konomi, S.; Nah, F.; Ntoa, S.; Rau, P.-L. P.; Siau, K.; Zhou, J. Seven HCI Grand Challenges Revisited: Five-Year Progress. *Int. J. Human–Comput. Interact.* **2025**, *41*, 11947–11995.
5. Weiser, M. The Computer for the 21st Century. *Sci. Am.* **1991**, *265*, 94–104.
6. Boucher, M.; Blumenstein, K.; de Jesus Oliveira, V. A.; Seidl, M. BYOD – Bringing your Own Device into Single-Surface Interaction Models. *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*; ACM, 2021, pp. 1–13.
7. Menon, A. R.; Zapico, J. L. Ambient Awareness: Experiencing Always-On Displays in the Life of PV Households. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*; ACM, 2025, pp. 789–805.
8. Schwarzer, J.; Draheim, S.; von Luck, K. Spatial and Temporal Audience Behavior of Scrum Practitioners Around Semi-Public Ambient Displays. *Int. J. Human–Comput. Interact.* **2022**, *39*, 3567–3585.
9. Sattar, S. K.; Curtis, H.; Neate, T. Public Assistive Displays: Employing Public Interactive Displays to Improve Public Transport Access 4All. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*; ACM, 2024, pp. 1–8.
10. Raza, M.; Ketsoi, V.; Malloch, J.; Bashbaghi, S.; Purmehdi, H.; Reilly, D. PerspectAR: Addressing Perspective Distortion on Very Large Displays with Adaptive Augmented Reality Overlays. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*; ACM, 2025, pp. 1–21.
11. Williamson, J. R.; Williamson, J. Analysing Pedestrian Traffic Around Public Displays. In *Proceedings of The International Symposium on Pervasive Displays*; ACM, 2014, pp. 13–18.
12. Mäkelä, V.; Heimonen, T.; Turunen, M. Semi-Automated, Large-Scale Evaluation of Public Displays. *Int. J. Human–Comput. Int.* **2017**, *34*, 491–505.
13. Williamson, J. R.; Williamson, J. Understanding Public Evaluation: Quantifying Experimenter Intervention. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*; ACM, 2017, pp. 3414–3425.
14. Chamberlain, A.; Crabtree, A., Eds., *Into the Wild: Beyond the Design Research Lab*; Springer International Publishing, 2020.

15. Effert, J.-S.; Günther, F. No Field Without Access: Recruitment as a Core Component of HCI Field Studies. In *Mensch und Computer 2025 – Workshopband*; Gesellschaft für Informatik e.V., 2025.
16. Sellmann, H.; Reiz, N.; Gerken, J.; Mörike, F. Field-Based Research on Hybrid Work: Approaching More Inclusive Research Practices for Broader Perspectives? In *Mensch und Computer 2025 – Workshopband*; Gesellschaft für Informatik e.V., 2025.
17. Draheim, S.; von Luck, K.; Putzar, L. Working with the Gaps in Understanding: Revisiting an Approach to Affective Computing Research in the EmotionBike Project. In: *Proceedings of the 18th ACM International Conference on Pervasive Technologies Related to Assistive Environments*; ACM, 2025, pp. 596 – 601.
18. Blandford, A. HCI for Health and Wellbeing: Challenges and Opportunities. *Int. J. Human-Comput. Studies* **2019**, *131*, 41–51.
19. Roffarello, A. M.; De Russis, L. Achieving Digital Wellbeing Through Digital Self-control Tools: a Systematic Review and Meta-analysis. *ACM Trans. Comput-Human Interact.* **2023**, *30*, 1–66.
20. Sanches, P.; Janson, A.; Karpashevich, P.; Nadal, C.; Qu, C.; Daudén Roquet, C.; Umair, M.; Windlin, C.; Doherty, G.; Höök, K.; Sas, C. HCI and Affective Health: Taking Stock of a Decade of Studies and Charting Future Research Directions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*; ACM, 2019, pp. 1–17.
21. Mateescu, M.; Pimmer, C.; Zahn, C.; Klinkhammer, D.; Reiterer, H. Collaboration on Large Interactive Displays: A Systematic Review. *Human-Comput. Interact.* **2021**, *36*, 243–277.
22. Atta, M.; Neumayr, T.; Hirschmann, F.; Schönböck, J.; Augstein, M. From Observation to Automation: Ai-Enhanced Analysis of Hybrid Collaboration. In *Proceedings of the 23rd EUSSET Conference on Computer Supported Cooperative Work*, European Society for Socially Embedded Technologies (EUSSET), 2025.
23. Elhart, I.; Mikusz, M.; Mora, C. G.; Langheinrich, M.; Davies, N. Audience Monitor: an Open Source Tool for Tracking Audience Mobility in Front of Pervasive Displays. In *Proceedings of the 6th ACM International Symposium on Pervasive Displays*; ACM, 2017, pp. 1–8.
24. Schwarzer, J.; Draheim, S.; von Luck, K.; Wang, Q.; Grecos, C. Spontaneous Utilization: a Classic Grounded Theory of Utilizing Ambient Displays in Professional, Large-Scale Agile Software Development Environments. *Int. J. Human-Comput. Interact.* **2021**, *38*, 1440–1463.
25. Koch, M.; Fietkau, J.; Stojko, L. Setting up a long-term Evaluation Environment for Interactive Semi-public Information Displays. In *Mensch und Computer 2023 – Workshopband*; Gesellschaft für Informatik e.V.: Rapperswil, Switzerland, 2023.
26. Shotton, J.; Fitzgibbon, A.; Cook, M.; Sharp, T.; Finocchio, M.; Moore, R.; Kipman, A.; Blake, A. Real-Time Human Pose Recognition in Parts from Single Depth Images. In *CVPR 2011*; IEEE, 2011; pp. 1297–1304.
27. Fietkau, J. A New Software Toolset for Recording and Viewing Body Tracking Data. In *Mensch und Computer 2023 – Workshopband*; Gesellschaft für Informatik e.V., 2023.
28. Schwarzer, J.; Fietkau, J.; Fuchs, L.; Draheim, S.; Von Luck, K.; Koch, M. Exploring Mobility Behavior Around Ambient Displays Using Clusters of Multi-dimensional Walking Trajectories. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*; ACM, 2023, pp. 1–6.
29. Mathes, L. K. P.; Cabalo, C. C.; Gatzemeyer, L. Evaluating the Engagement of Users from Public Interactive Displays. In *Mensch und Computer 2023 – Workshopband*; GI, 2023.
30. Filippov, V.; Lowack, I.; Merseburger, J.; Ell, K. *Semi Automatic Differentiation Between Passing and Reacting Users*, *Experiment Report*; University of the Bundeswehr Munich: Unpublished, 2024.
31. Koch, M.; von Luck, K.; Schwarzer, J.; Draheim, S. The Novelty Effect in Large Display Deployments – Experiences and Lessons-Learned for Evaluating Prototypes. In *Proceedings of the 16th European Conference on Computer-Supported Cooperative Work – Exploratory Papers*, 2018.
32. Wouters, N.; Downs, J.; Harrop, M.; Cox, T.; Oliveira, E.; Webber, S.; Vetere, F.; Vande Moere, A. Uncovering the Honeypot Effect: How Audiences Engage with Public Interactive Systems. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*; ACM, 2016, pp. 5–16.
33. Bieschke, L. N. *Analyse von Body-Tracking-Daten vor großen Wandbildschirmen hinsichtlich Honeypott-Konstellationen*; University of the Bundeswehr: Munich, 2024. Master's thesis.
34. Stephanidis, C.; Salvendy, G.; Antona, M.; Chen, J. Y. C.; Dong, J.; Duffy, V. G.; Fang, X.; Fidopiastis, C.; Fragomeni, G.; Fu, L. P.; Guo, Y.; Harris, D.; Ioannou, A.; Jeong, K.; Konomi, S.; Krömker, H.; Kurosu, M.; Lewis, J. R.; Marcus, A.; Meiselwitz, G.; Moallem, A.; Mori, H.; Nah, F. F.; Ntoa, S.; Rau, P. P.; Schmorow, D.; Siau, K.; Streitz, N.; Wang, W.; Yamamoto, S.; Zaphiris, P.; Zhou, J. Seven HCI Grand Challenges. *Int. J. Human-Comput. Interact.* **2019**, *35*, 1229–1269.
35. Kazempour, D.; Beer, A.; Herzog, F.; Kaltenthaler, D.; Lohrer, J.-Y.; Seidl, T. F. A. T. B. I. R. D. A Tool for Flight and Trajectories Analyses of Birds. In *2018 IEEE 14th International Conference on e-Science (e-Science)*; IEEE, 2018, pp. 75–82.
36. AlMahamid, F.; Grolinger, K. Agglomerative Hierarchical Clustering with Dynamic Time Warping for Household Load Curve Clustering. *2022 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*; IEEE, 2022, pp. 241–247.
37. Lacher, J.; Bieschke, L.; Michalowski, F.; Münch, J. Using Machine Learning to Determine Attention Towards Public Displays from Skeletal Data. In *Mensch und Computer 2023 – Workshopband*; GI, 2023.
38. Ottenheim, T. *Graph Convolutional Network Based Gesture Interpretation of Skeleton Data*, *Bachelor's Thesis*; Hamburg University of Applied Sciences, 2024.
39. Wang, W.; Zhang, Y.-D. A Short Survey on Deep Learning for Skeleton-based Action Recognition. *Proceedings of the 14th IEEE/ACM International Conference on Utility and Cloud Computing Companion*; ACM, 2021, pp. 1–6.
40. Jarrahi, M. H.; Memariani, A.; Guha, S. The Principles of Data-Centric AI. *Communications of the ACM* **2023**, *66*, 84–92.
41. Xu, T.; Sarkar, A.; Rintel, S. Is a Return to Office a Return to Creativity? Requiring Fixed Time in Office to Enable Brainstorms and Watercooler Talk May Not Foster Research Creativity. *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work*; ACM, 2023, pp. 1–12.
42. Dourish, P. Reading and Interpreting Ethnography. In *Ways of Knowing in HCI*; Springer: New York, 2014; pp. 1–23.
43. Przybylski, L. *Hybrid Ethnography: Online, Offline, and in Between*; SAGE Publications, Inc., 2021.

44. Liu, R.-F. Hybrid Ethnography: Access, Positioning, and Data Assembly. *Ethnography* **2022**, 14661381221145451; <https://doi.org/10.1177/14661381221145451>.
45. Karhapää, A.; Pöysä-Tarhonen, J.; Rikala, P.; Hämäläinen, R. Being an Ethnographic Researcher in a Modern Workplace: Advantages and Challenges of Digital Ethnography. *J. Organ. Ethnogr.* **2025**; <https://doi.org/10.1108/joe-11-2024-0085>.
46. Reunamäki, R. Living with and Living like those we Study: Organizational Hybrid Ethnography as a Research Approach in Organization Studies. *J. Organ. Ethnogr.* **2025**; <https://doi.org/10.1108/joe-10-2024-0076>.
47. Chen, J.; Lv, Z.; Wu, S.; Lin, K. Q.; Song, C.; Gao, D.; Liu, J.-W.; Gao, Z.; Mao, D.; Shou, M. Z. VideoLLM-online: Online Video Large Language Model for Streaming Video. In: *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 18407–18418.
48. Bansal, G.; Nah, F. F. H. Internet Privacy Concerns Revisited: Oversight from Surveillance and Right to Be Forgotten as New Dimensions. *Inf. Manag.* **2022**, 59, 103618.
49. Draheim, S. Can Research Ethics Serve as a Platform for Negotiating Digital Transformations of Society? In *Proceedings ICERI 2024 — 17th Annual International Conference of Education, Research and Innovation, IATED Vol. 1*, 2024, pp. 8239–8245.
50. Sethi, D.; Bharti, S.; Prakash, C. A Comprehensive Survey on Gait Analysis: History, Parameters, Approaches, Pose Estimation, and Future Work. *Artif. Intell. Medicine* **2022**, 129, 102314.
51. Fietkau, J.; Schwarzer, J. Herausforderungen Menschengerechter Forschung Mit Body-Tracking-Sensoren in Langzeit-Feldstudien. In *Mensch und Computer 2024 — Workshopband*; Gesellschaft für Informatik e.V., 2024.
52. Schwarzer, J.; Koch, M.; Draheim, S.; Fietkau, J.; von Luck, K. An Integrated Methodological Framework to Investigate Hybrid Work Technologies. In *2nd International Workshop on Hybrid Collaboration — EUSSET Conference on Computer-Supported Cooperative Work (ECSCW)*, 2025.
53. Koch, M.; Draheim, S.; Fietkau, J.; Schwarzer, J.; von Luck, K. **2024. Untersuchung Des Honeypot-Effekts an (halb-)öffentlichen Ambient Displays in Langzeitfeldstudien.** tech. rep.; Universität der Bundeswehr München, Hochschule für Angewandte Wissenschaften Hamburg; Hannover, Germany.