RESEARCH-ARTICLE

Robots as Hosts in Autonomous Buses: A Field Trial

AGNES JOHANNA AXELSSON, Delft University of Technology, Delft, Zuid-Holland, Netherlands

BHAVANA VADDADI, KTH Royal Institute of Technology, Stockholm, Stockholms, Sweden

CRISTIAN BOGDAN, KTH Royal Institute of Technology, Stockholm, Stockholms, Sweden

DEIRDRE TOBIN, KTH Royal Institute of Technology, Stockholm, Stockholms, Sweden

GABRIEL SKANTZE, KTH Royal Institute of Technology, Stockholm, Stockholms, Sweden

Open Access Support provided by:

KTH Royal Institute of Technology

Delft University of Technology

Published: 28 October 2025
Online AM: 05 August 2025
Accepted: 25 July 2025
Revised: 28 May 2025
Received: 24 July 2024

Robots as Hosts in Autonomous Buses: A Field Trial

AGNES AXELSSON, Division of Speech Music and Hearing, KTH Royal Institute of Technology, Stockholm, Sweden and Intelligent Systems, Delft University of Technology, Delft, Netherlands
BHAVANA VADDADI, Integrated Transport Research Lab, KTH Royal Institute of Technology, Stockholm, Sweden
CRISTIAN BOGDAN and DEIRDRE TOBIN, Division of Media Technology and Interaction Design, KTH Royal Institute of Technology, Stockholm, Sweden
GABRIEL SKANTZE, Division of Speech Music and Hearing, KTH Royal Institute of Technology, Stockholm, Sweden

In Autonomous Public Transport (APT), particularly with shuttle buses, passengers travel in smaller, more intimate vehicles—and in the future, such vehicles may operate without an authoritative driver or host. This setup may lead to potential safety concerns, as passengers are left alone together. Additionally, this future absence of a driver or host means that there is no one to address questions or uncertainties that may arise. One proposed solution is introducing a robot onboard the bus, serving a similar role to a human host. To explore this solution, an experiment was conducted in Barkarby, Stockholm, Sweden. Passengers, generally unfamiliar with APT or social robots, experienced two short rides on a bus equipped with either an embodied Furhat robot as the host or a disembodied voice agent in the ceiling. Data were collected from passenger-agent interactions, post-questionnaires, and semi-structured focus group interviews. Results indicate a division in passenger preferences, with some favoring the robot and others the voice assistant. Passengers asked more questions to the robot, suggesting a clearer affordance for interaction. While the questionnaires did not show significant differences, passenger behaviors indicated that they anthropomorphized the robot more. The interviews revealed that passengers felt more secure with a human operator and doubted the robot's authority during incidents with aggressive passengers or accidents. Our findings show that social robots can help make autonomous buses feel more welcoming and interactive. Future APT systems have many design issues that need to be resolved before riders can find them safe and appropriate to use, and social robots can play a role in resolving such issues—both the ones we see today, and potentially ones that will appear in the future.

CCS Concepts: • **Human-centered computing → User studies**; **Field studies**; *Usability testing*; Participatory design; User centered design;

Additional Key Words and Phrases: autonomous, public transport, APT, robot, clerk, assistant, self-driving, bus, passenger, guide, shuttle, wizard

This study is part of the project "Social Robots Accelerating the Transition to Sustainable Transport", funded by the Swedish Energy Agency and Furhat Robotics (P2020-90133). Part of this research was financed by *STandUP for Energy*.
Authors' Contact Information: Agnes Axelsson (corresponding author), Division of Speech Music and Hearing, KTH Royal Institute of Technology, Stockholm, Sweden and Intelligent Systems, Delft University of Technology, Delft, Netherlands; e-mail: a.axelsson@tudelft.nl; Bhavana Vaddadi, Integrated Transport Research Lab, KTH Royal Institute of Technology, Stockholm, Sweden; e-mail: bhavana@kth.se; Cristian Bogdan, Division of Media Technology and Interaction Design, KTH Royal Institute of Technology, Stockholm, Sweden; e-mail: cristi@kth.se; Deirdre Tobin, Division of Media Technology and Interaction Design, KTH Royal Institute of Technology, Stockholm, Sweden; e-mail: deirdret@kth.se; Gabriel Skantze, Division of Speech Music and Hearing, KTH Royal Institute of Technology, Stockholm, Sweden; e-mail: skantze@kth.se.

ACM 2573-9522/2025/10-ART19
https://doi.org/10.1145/3759158

ACM Reference format:
Agnes Axelsson, Bhavana Vaddadi, Cristian Bogdan, Deirdre Tobin, and Gabriel Skantze. 2025. Robots as Hosts in Autonomous Buses: A Field Trial. *ACM Trans. Hum.-Robot Interact.* 15, 1, Article 19 (October 2025), 42 pages.
https://doi.org/10.1145/3759158

1 Introduction

Social robots have been deployed and tested in many public-facing settings, including healthcare [30], elder care [11, 23], in-vehicle assistance [44, 49, 51], and public transport [29, 80]. Many human-AI collaborative settings require close collaboration with humans, who can provide goals for a robot [60, 62], or observe a robot [77] or vehicle [5] so they do not cause dangerous or unsafe situations.

The advent of autonomous technologies has sparked a paradigm shift in our current transport system. Amidst this transformative landscape, **Autonomous Public Transport (APT)** emerges as a promising solution to the increasing challenges urban mobility systems face today [14]. APT solutions, such as full-size buses and smaller shuttle buses, operate on fixed schedules or via on-demand apps. They provide a cost-effective, adaptable, and safe alternative to personal cars [27, 59]. Currently, APT services such as shuttle buses operating on public roads have reached SAE [18] Level 4, which means that they can drive themselves without human intervention in certain conditions, but often require a safety operator on board for legal or technical reasons. Fully autonomous buses, classified as SAE Level 5, which can operate without any human driver or safety operator, are still limited to pilot testing and have not yet been deployed on a large scale on public roads [18, 38, 66].

Concerns regarding trust and safety within APT vehicles remain prevalent among users and must be addressed to ensure full acceptance and utilization of these systems [32, 54]. In traditional public transport, the *driver* plays an essential role in the vehicle, acting as an authority figure. Drivers help ensure passenger safety, assist, address concerns, navigate unexpected situations, and maintain order [4, 32]. Without a human driver or host on board, passengers in future SAE Level 5 APT will not have anywhere obvious to turn to ask questions or get guidance if an unexpected event happens during the ride. Integrating social robots as hosts in autonomous buses could potentially address these issues by giving the passengers someone to turn to.

By providing communication, assistance, and personalized interactions, social robots have the potential to reshape how passengers relate to and utilize APT [22]. By imbuing these robots with traits reminiscent of hospitable hosts—such as warmth, empathy, and adaptability [26]—APT systems could potentially cultivate a sense of comfort and familiarity among passengers.

However, the integration of robots as hosts in autonomous buses is not without its challenges and ethical considerations. Questions surrounding privacy, data security, and the ethical implications of human–robot relationships underscore the need for careful deliberation and thoughtful design [56]. The integration of social robots into APT cannot resolve all problems and user concerns with such vehicles, and rather must happen alongside addressing safety and reliability issues with the vehicles themselves. All aspects must be worked on for social robot integration to be efficient and accepted.

Therefore, a critical question arises: if a robotic agent is placed in a bus, should it be integrated as an inherent part of the bus design, or should it be considered a separate, independent entity? We aimed to investigate how social robots can address feelings of insecurity and lack of safety in APT. This article explores the multifaceted role of robots as hosts in APT through a field trial at Sweden's unique Autonomous Bus Trial Site in Barkarbystaden, north of Stockholm. Questions about how the agent should be framed as part of its environment relate to *situatedness* [52] and *embodiment* [21, 28], leading us to our first research question:

RQ1: What are the differences between a disembodied and an embodied agent serving as a host on a self-driving bus, in terms of how passengers perceive it and interact with it?

In a regular, human-operated bus, it is clear to passengers that the driver and a ticket inspector fulfill different functions in the vehicle. We, as human passengers, learn from previous human–human interaction [16] to trust that a driver or ticket inspector will step in if a violent situation arises, or can provide basic information about delays or traffic situations. These assumptions about provided services are not at all clear when a robot clerk is present on the bus. What should it do, and what do passengers expect it to do? This leads us to our second research question.

RQ2: What services could an agent (embodied or not) provide in APT and what would passengers want to use it for?

Above, we described how safety and trust in APT are key factors for passengers to accept and use the vehicles. Thus, our third research question explores the robot's effect on perceived safety:

RQ3: Does the presence of an agent in APT address passengers' feelings of lack of safety or insecurity prompted by the absence of an authoritative human or smaller vehicles?

In a within-subject experiment, we contrasted a physical robot with a disembodied agent (a smart speaker in the bus ceiling). This setup allowed us to explore how passengers interact with, perceive, and prefer the agent based on its embodiment. We previously published a video paper [6] showing the nature of the interactions during the rides, but without analyzing any of the data we collected. In this article, by employing both qualitative and quantitative analysis methods, we address the three mentioned research questions, triangulate the results (Section 4), and provide design guidelines (Section 5) derived from on-site experimentation (Section 3.2). This article thus presents a valuable insight into how—and if—social robots can be integrated into the public transit systems of the future.

2 Background

2.1 Autonomous Public Transport and Passenger Perceptions

While most research on autonomous technologies has focused on personal vehicles, interest in applying these technologies to public transport has been growing [59]. Autonomous systems have already been successfully implemented in various public transport modes, such as metros and trains, in Europe, the USA, and Japan for several years. Recently, there has been a significant increase in testing autonomous technologies in road-based public transport. As with personal vehicles, APT can bring great benefits to everyday public transport users, service providers, operators, and society. According to the literature, some of these benefits include reduced human errors and accidents [81], efficient travel [27], inclusive and accessible alternatives [47], reduced operational costs [10], and a potential decrease in energy consumption and emissions [59]. These benefits can enhance the public transport experience and support the shift toward a sustainable transport system. However, it is imperative to recognize user expectations and attitudes toward APT vehicles to understand their acceptance and ultimately their usage.

Recent studies on APT acceptance indicate that, compared to personal **Autonomous Vehicles (AVs)**, users generally view APT vehicles more favorably [55]. However, they have identified crucial concerns related to safety [32, 36] and trust in the technology [15, 54], which impact user acceptance toward and willingness to use APT. While several other concerns impact participant willingness to use APT, this article focuses on factors associated with safety and lack of trust and how social robots can address them. Among these factors are:

(1) Users might feel that they do not have control during emergencies when there is no host on board and due to lack of sufficient information [32, 36].
(2) Users might feel that the technology is not at its full potential due to lower speeds, and performance in heavy vehicular, or pedestrian traffic [15, 54].
(3) Users might believe that human intervention is crucial in difficult situations, given the immature state of the technology. In the studies conducted by Dong et al. [24] and Mouratidis and Serrano [54], it was observed that people were more willing to use APT vehicles if there was an onboard host who could supervise the operations and help with the ride. The presence of the on-board host can be a determining factor for the users to continue using APT in the future [24, 41, 42, 54].
(4) Users might feel a potential threat to themselves due to inappropriate activity within a small space, lack of safety equipment within the bus, or fear of software hacking [25, 48].

In the following sub-section, we explore how a social robot can address the above-listed concerns through its social presence, facial mimicry, and oral communication.

2.2 Social Robotics and Embodiment

An *embodied* agent has a physical presence in a space [21]. Industry robots or delivery robots must exist in the space together with the objects that they assemble or deliver. Purely communicative agents, however, can be present only in audio (voice assistants), in text (chatbots), or be embodied in virtual form on a screen, or in VR. Thus, the embodiment is not the same thing as *situatedness*, which refers to the robot's integration into the environment in which it is placed, having a physical relationship to the users interacting with it [52].

Social robots are robots that are designed to interact with humans, either as their main task or as part of their interaction with the physical environment (e.g., "cobots" [33]). Social robots are often argued to benefit from embodiment for social purposes [76], but it is not clear if physical presence leads to humans feeling a greater degree of social presence with the robot or social influence from the robot [72]. An embodied robot can be co-located and co-present with a human in a space [21], while a disembodied agent can have neither of these properties.

Anthropomorphism, or human-likeness, is another design aspect of robots that can affect how human interaction partners perceive the robot as a social partner [69, 84]. By making a robotic agent more human-like, human interaction partners can make use of more of their social behaviors toward other humans to inform how to interact naturally with the agent [39]. Social robots can be perceived as warmer and more competent by acting more accommodating [26]. The mere presence of an embodied social robot in a social space can lead to positive conversations and jokes about the robot in, for example, an elderly care scenario [11]. However, developing social robots that can reliably interact with sensitive, especially elderly, users requires a great deal of adaptation to the individual capabilities of the users, and the novelty is not guaranteed to last over time [23].

In a meta-study on anthropomorphism, Roesler et al. showed that anthropomorphism interacts with embodiment in such a way that human-like systems shown in a picture are generally evaluated as less socially present than those a participant has interacted with personally. This suggests that the perception of anthropomorphism interacts with situatedness and embodiment in a complex way [64].

Fischer et al. found that humans who interact with differently embodied agents talk to those agents in ways that depend on the embodiment. Participants are more likely to address a robot by name in a mechanical task if it is physically embodied, as opposed to on a monitor. Additionally, the more physical movement the robot shows, the more ready participants are to expect it to be able to perform complex mechanical movements when teaching it a physical task [28].

One specific aspect of embodiment is whether the agent can display its focus of attention in an anthropomorphic way through its gaze. Wiese et al. showed that an embodied agent with human-like gazing behaviors (turning the head as well as moving the pupils) could direct a human interaction partner's attention to specific objects [78]. In multi-human interactions, gaze can indicate turn-taking, with participants gazing at others to pass the turn to the next speaker [2]. This behavior can be replicated with an embodied gazing agent [46, 67], but less so with an agent on a screen [74, 75] and not at all with a disembodied agent. These differences are similar to those in human–human interaction when comparing in-person meetings, video calls, and audio-only phone calls [73].

Embodiment and situatedness can be designed so that users have different experiences depending on what they prefer. A virtual agent can change where it exists and in what form if the designers think that this is beneficial in some way. Luria et al. performed an experiment on how participants perceived screen-based agents that could *re-embody* themselves by moving to a new monitor, or *co-embody* a screen together with other agents. The authors found that participants were not comfortable with the same social presence being capable of different tasks requiring different expertise, like a receptionist agent that also played the role of a doctor [50]. Some participants also thought of re-embodiment as if the agents were *calling in*, remotely, from their actual location, rather than moving locations to the new embodiment [50].

2.2.1 Disembodied Agents in Vehicles and Public Transport. In autonomous personal vehicles, disembodied voice assistants can fulfill the purpose of letting passengers control car systems and in-ride entertainment. Similarly, if they are connected to navigation, they can let drivers or passengers provide navigation goals [49].

Meck et al. have suggested that recent advances in conversational AI will mean that driving assistants will "soon" become more responsible for the activity of driving than the human driver [51], although it is unclear how "soon" it is realistic to expect this. Keller et al. have shown that convenience and system competence are important factors in leading drivers to accept using assistive systems that make decisions for them, while drivers reported not being worried about potential privacy issues [43].

Jestin et al. compared how participants perceived different voices in a driving assistant scenario. This experiment framed the agent as part of the vehicle, but not always in control—the fictional assistant could give control to the human if its systems malfunctioned, or take control away from the driver if it detected that an accident was about to happen [40]. This level of automation corresponds to Level 4 on the scale by the Society of Automotive Engineers [18]. The authors found that participants preferred a commanding or formal voice to an informal voice and that this effect was stronger in safety-critical scenarios than in less urgent scenarios. Participants also preferred a male voice to female and non-binary alternatives. The authors note that the gender effects may also be because of other features of the voices, unrelated to the implied gender of the driving assistant [40].

Luria et al. set up a scenario where a social agent who was responsible for driving an autonomous car also conversed with another social agent in the middle of the car ride. The fictional vehicle in this setup was fully autonomous, i.e., level 5 on the SAE scale [18]. Passengers found that it was unsafe for the driving agent to be *distracted* by the conversation with the other agent. In an example of anthropomorphism affecting perceived safety, they stated in interviews that they would prefer for the agent to "focus on driving" [50]. This illustrates that it may not be ideal to frame a vehicle-related agent as also responsible for driving.

2.2.2 Embodied Agents in Vehicles and Public Transport. An early example of using social robots in a public transport setting came from Svenstrup et al. [70]. The authors deployed a robot dressed like Santa Claus in a Danish mall and bus stop. The robot could not help humans with public transport situations. When participants were asked what the role of robots was in then-future transport settings, 40% of participants responded that they thought robots could have an assistive role in such spaces [70]. Given the highly gimmicked appearance of the robot and its limited functionality, it is unclear how applicable these results are to our study.

Kim et al. recently explored how in-car embodiments of co-driving assistants were perceived by drivers. The assistive agents were placed below the rear-view mirror in a car. The authors found that a trumpet-like agent was preferred over a lip-like agent to alert drivers when the car needed to hand over driving responsibilities to them, or alert them that the car would take over from the driver because of a traffic situation. Both the trumpet-like and the lip-like embodiment were chosen to have a visual association with the concept of alerting or warning the driver. The trumpet-shaped agent was preferred as it could also point (by moving the horn of the trumpet) in a vaguely anthropomorphic way [44].

When an AV drives in an environment with pedestrians, the interaction between the vehicle and those pedestrians becomes a **Human–Robot Interaction (HRI)**, with the robot's embodiment being the car. A recent field of interest in HRI research has been how such vehicles should communicate with humans around them. Colley et al. call the interaction modalities with the environment as an *eHMI*, for *external Human–Machine Interface* [17]. Connecting to *eHMI* research, Block et al. showed that pedestrians felt safer when crossing a (virtual) sidewalk when the vehicle was able to show that it was paying attention to them specifically, as opposed to a generic, static signal [12]—the HRI equivalent to eye contact while waving a pedestrian across the road.

Pelikan and Jung argued that sound design could control whether pedestrians perceived an autonomous bus as an ice cream truck-style vehicle (worthy of attention, but not critical) or an emergency responder-type vehicle (worthy of immediately turning one's attention to, as to avoid) [61]. Alternatively, Moore et al. argue that there may be reasons to design an AV to look as much like a human-driven car as possible, to minimize "griefing" such as brake-checking or pretending to cross a crosswalk, where humans in the vehicle's environment deliberately disturb the operation of the vehicles for fun or to test their capabilities [53].

2.2.3 Embodiment and Robot Authority in Unsafe Environments. An important aspect of deploying social robots in APT is whether they can have the *authority* of a human driver. Robot authority in HRI has been explored directly and indirectly in the past with mixed results, perhaps because authority means different things depending on the interaction and context.

Haring et al. showed that participants in an object search game were less likely to follow requests for compliance from a robotic coach than from a human coach. The authors also found that different robots with varying degrees of anthropomorphism were not different from each other in how compliant the humans were, and that there was generally no difference in compliance between the trainer being present virtually on a screen or physically present in the room [34]. Thus, the robots generally did not have authority regardless of how they were designed, and the human always had authority regardless of how they were co-present.

Recently, Boos et al. built an experiment to evaluate the effect of several factors on human acceptance of a robot's request. The study looked for a connection between the robot's embodiment, self-disclosure of personal information (an indirect measure of anthropomorphism), and its use of a smaller, more acceptable request before the larger request. No significant connection was found between these factors and participant acceptance of the request [13], suggesting that authoritative robots cannot be built merely through the use of these three design dimensions.

Fig. 1. A collage of participants in our experiment, as well as the driver who was legally mandated to be onboard (wearing a white shirt or yellow safety vest), as seen from the robot's 180-degree camera. The faces of the drivers have been blurred for privacy. The passengers in these images consented to being portrayed.

Another aspect of authority is to what extent humans trust an agent with their personal information. Fosch Villaronga et al. argued that trust in a social robot and privacy concerns overlap. One cannot fully trust a robot if one has concerns about how it is handling the personal data handed over to it [30].

3 Method

From April to June 2023, we conducted a field study in Barkarby, Sweden, with a total of 57 recruited participants. The purpose of this study was to examine how the physical embodiment of the agent (embodied or disembodied) influenced passenger perceptions and behavior (RQ1). Additionally, we sought to investigate passenger expectations and desires regarding the presence of an agent in an autonomous bus (RQ2 and RQ3). Since it was hard to know beforehand which situations might arise during the trip and what questions the passengers might have for the robot, a Wizard-of-Oz setup was used to control the agent. We analyze both the passengers' interactions with the agent during the ride, their answers to questionnaires given to them after the ride, as well as semi-structured focus group interviews.

The KTH Research Ethics support was consulted and they deemed that no ethical approval was needed for this type of study, according to Swedish law. All our data collection and processing follow the EU's **General Data Protection Regulation (GDPR)**. The participants who are visible in Figure 1 consented to being shown.

3.1 Autonomous Bus Site

Between 2018 and late 2023, a small fleet of autonomous EasyMile shuttle buses operated in Barkarby, a suburb of Järfälla municipality north of Stockholm, Sweden. The location of the site can be seen on the left of Figure 2. In early to mid 2023, the bus operator, Nobina AB, performed research and development of a short route where the autonomous buses could drive with a higher level of autonomy. In agreement with Nobina, we performed our experiment on this route. A map of the specific route can be seen on the right of Figure 2.

The EasyMile shuttle buses that were used are illustrated in Figure 3. Each shuttle can carry a maximum of six seated passengers and six standing passengers, but our experiment did not make

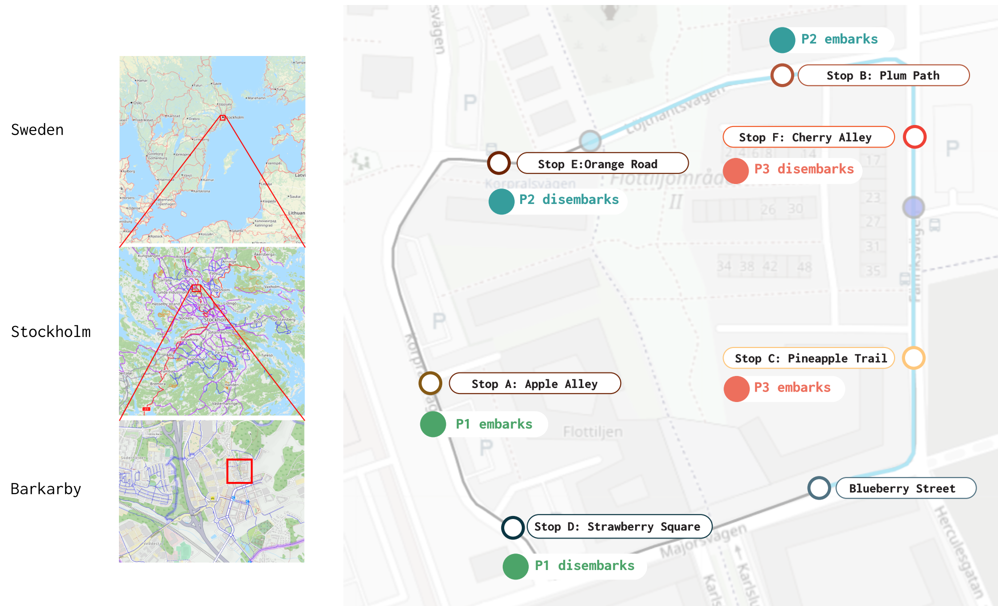

Fig. 2. *Left*: A progressively zoomed-in map showing the location of our experiment site. The top image shows the southern half of Sweden with the Baltic Sea on the right. The middle image shows Stockholm. The bottom image shows the Järfälla and Barkarby area. The red rectangle the bottom image corresponds to the area shown in the route map on the right. *Right*: The stop setup for one of our experimental runs. Stops are labeled A, B, C, D, E, and F in order. Passengers disembark one stop before a full lap around the circuit. The stop names have been translated from the original Swedish—the first run, shown here, used fictional stop names based on fruits, while the second run used different fictional stop names based on vegetables.

use of the full capacity—instead, we considered a maximum of three passengers at a time. For safety reasons, a trained onboard operator was also always present on the bus. The safety operator, who was always a trained bus driver, can be seen wearing a high-visibility vest in several of the photos shown in Figure 1. Although the operator was mostly seated backward, viewing the internal cameras and ride system on a laptop, the operator also occasionally had to stand up to monitor the environment.

3.2 Experimental Design

Our experiment was based on a within-subject design. The two conditions differed in which of the two embodiments of our agent the participants experienced first.

To explore the effect of embodiment, we compared the Furhat robot head [3] to a disembodied voice-only agent. The Furhat robot has an animated face back-projected on a semi-translucent mask, mounted on a mechanical neck with three degrees of freedom. This platform was chosen since the face is very expressive (in terms of facial expressions and lip movement), and the robot can direct its gaze using both head and eye movement, resulting in a very accurate gaze perception [82], which is important for multi-party interaction [2, 75]. The synthesized speech comes from a built-in speaker in the robot's base. The disembodied agent did not have any visual component, and its voice came out of a conference microphone-and-speaker unit attached to the ceiling of the bus. Figure 4 illustrates both embodiments.

Fig. 3. One of the EasyMile shuttle buses used in the experiment.

Fig. 4. *Left*: The Furhat agent during the experiment (marked as *Robot* in Figure 6). *Right*: A close-up of the conference microphone (marked as *Microphone and speaker* in Figure 6) in the ceiling of the bus that played the audio of the disembodied agent.

3.2.1 Agent Persona. Both the embodied and the disembodied agent were referred to as *Elin*. For speech synthesis, the female Swedish Amazon Polly voice "Astrid" was used. We also allowed the agent to switch to English, in which case the British English Amazon Polly voice "Amy" was used instead.

Table 1 shows the example lines from the script the robot used for recurring situations during the rides. The script was written in a professional and non-humorous personality. This more professional style aligns with preferences found in previous research by Jestin et al. [40]. Edwards et al. [26] have shown that an over-accommodating personality is preferable to an under-accommodating personality when designing the first interaction between a human and a robot. Despite this, the benefits of an over-accommodating personality—increased perceived warmth and personal

Table 1. Example Responses from the Script Used by the Robot

Situation	Script
Passenger embarks	*Hi! I am Elin, a bus assistant. Please sit down and put your seatbelt on! Could you say what your destination is?*
Passenger asks for out-of-domain response, e.g., "What kind of AI are you?"	*Sorry! I am not programmed to be able to respond to that.*
Passenger gives feedback on features of the bus or robot, e.g., "The bus should stop closer to the curb so people with a wheelchair can get on more easily"	*I will let my programmers know that.*
As the bus gets close to the passenger's stop	*Excuse me! We are getting close to your stop. Remember to take a look around so you don't forget anything.*

These are the English lines from the script, but the agent spoke Swedish during the experiment except if a participant asked it to speak English. The Swedish and English lines were written to be as equivalent as possible.

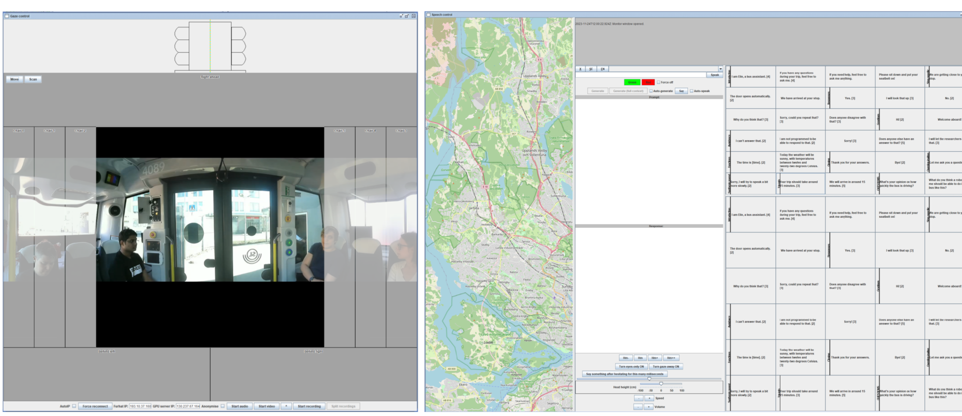

Fig. 5. The interface through which the Wizard teleoperated the agent. *Left*: The gaze control window. This window is made semi-transparent and then put on top of the Zoom feed of the bus from the wide-angle webcam. The embodied agent can be made to look at any one of the six seats in the bus by clicking on the semi-transparent gray areas surrounding the seats in the window. *Right*: The speech control window. Pre-written lines are sorted in a grid on the right, and the agent can be made to speak them if clicked. On the left side of the window, a map is shown which automatically zooms in on the GPS coordinates of the bus so the Wizard can see where the bus is at any given point.

closeness between the individual and the robot—are not necessarily factors we would want to maximize for our robot's use-case. Additionally, when pilot-testing the system, we found that a more personal style led to individual passengers monopolizing the system for one-on-one small talk. Such small talk put a large cognitive load on the Wizard to improvise responses and also prevented the system from giving timely transport information to other passengers than the one being currently spoken to.

The Wizard could control each sentence in a response separately, and some responses had multiple variations ("Hi", "Hello", etc.) to make the agent's speech vary a little between passengers and runs. The speech control panel is illustrated on the right of Figure 5.

3.3 Experimental Procedure

Our experiment consisted of two rides on the bus with a break in the middle to fill in questionnaires. We previously illustrated the setup and nature of the rides in a video publication [6].

Before the first ride, participants signed GDPR and data release forms. These forms also informed participants about their rights to opt out of the data collection at any time, including after the experiment. After the participants had consented to be part of the experiment, each participant was handed a sheet with the necessary information about the experimental procedure. These sheets told participants to avoid talking to the human bus operator if possible and to not use their phones while on the bus. The sheet also assigned each passenger to a stop on the circular route where they would embark, and a stop where they would disembark. This assignment was done by the name of the stop to hide the semicircle assignment of embarkations and disembarkations shown in Figure 2 from the participants.

As illustrated in Figure 2, the stops were given fictional names. The names were different across the two different rides. This ensured that participants could not know where they would disembark until they were told where it was by the agent. The locations of the embarkation and disembarkation stops were arranged such that the three passengers would embark one stop at a time. Passengers then disembarked in the same order they had embarked.

After disembarking from their first ride, passengers filled in Godspeed [7] questionnaires, as well as some questions about the quality of their ride. The questions are listed in Appendix A. While this was happening, we changed the embodiment of the agent by removing or installing the Furhat robot head out of view of the participants. In both the embodied and disembodied ride, participants interacted with a teleoperated agent, and they were led to believe that the agent was truly autonomous as part of the experimental setup. The two conditions differed in which embodiment the passengers experienced during their first run. The passengers then experienced the other embodiment during their second run.

The second ride took place after passengers finished filling in their first questionnaires. To set up for the second ride, passengers were handed a new sheet assigning them to a stop where they would embark and disembark. The passengers were moved around between the two rides so that the passenger who had embarked and disembarked last in the first ride now embarked and disembarked first, with the others shifting back one position in the order each.

When the passengers had all disembarked after their second ride, they again filled in questionnaires. These questionnaires are also listed in Appendix A. Then, we performed an extended semi-structured group interview with all of the passengers simultaneously, asking them questions about their experience of the ride, the robot, and the vehicle.

3.4 Implementation

The teleoperated agent was implemented using the Furhat SDK. The Furhat robot itself ran a basic *client* program. The *client* program could only receive remote-controlled commands from the Wizard and had no autonomous behavior. The commands from the Wizard could tell the agent to gaze at an arbitrary spot in 3D space, say an arbitrary string of text, or to swap its voice synthesis between the English and Swedish preset voices.

A separate remote control *server* program ran on the Wizard's computer. This application was designed to allow the Wizard to send speech and gaze commands to the agent as required. The Wizard could make the agent stop speaking in the middle of a sentence with a pedal if passengers spoke at the same time as the robot. The Wizard could also see the GPS position of the bus in real time and could view a real-time video of the bus, streaming from the wide-angle camera mounted to the window behind the robot. An image of this interface is shown in Figure 5.

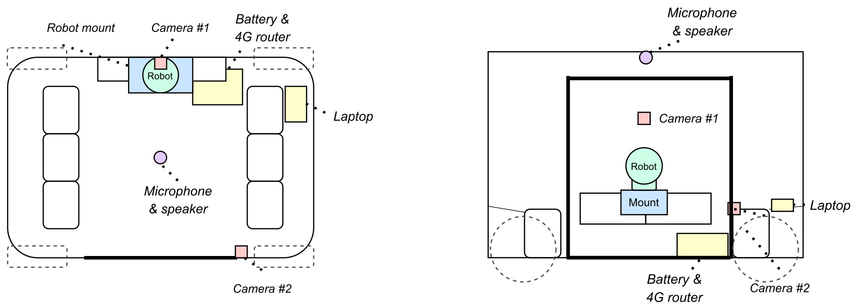

Fig. 6. A schematic of our experimental setup. *Left*: A top-down view. *Right*: A side view. The door of the bus is marked with a thicker line on the bottom-middle of the schematic, with the seats on the left and right side. The mount with the robot on top is visible in the middle, with the speaker and microphone in the ceiling of the bus. Miscellaneous equipment is on the floor and behind the passengers to the right.

In the *disembodied* condition, the client program ran locally on a virtual Furhat robot on the Wizard's computer, instead of on the robot in the bus. The virtual Furhat program was set up to output audio into the Zoom call used to record the bus, which meant that the agent's speech came out of the speaker and microphone unit attached to the bus's ceiling (Figure 4).

The setup in the bus is illustrated in a schematic in Figure 6. In the bus, two batteries powered the parts of our setup that needed electricity. A phone running a GPS data streaming app was placed near the batteries to give the GPS signals that the Wizard could view from the teleoperation interface. A wide-angle Jabra PanaCast 180-degree webcam was attached to the window behind the robot with a suction cup mount; it was angled in such a way that it would capture a view of all six seats in the bus (as shown in Figure 1). A GoPro camera was similarly attached to the right side of the door to record video for post-analysis from the other side of the bus.

The wide-angle webcam was plugged into a laptop, which in turn was connected to a Zoom call with video streaming enabled. The Wizard could see the interior of the bus with minimal latency by connecting to the same call. The conference microphone in the vehicle's ceiling was used as the laptop's microphone when the experiment was running, and also its speaker in the *disembodied* condition. The phone streaming GPS data, the laptop streaming the Zoom call, and the Furhat agent when it was present were all connected to a mobile 4G router giving all devices access to the Internet when required.

3.5 Recruitment of Participants and Demographics

Recruitment took place through a professional experiment recruitment platform. A total of 57 participants took part in the experiment. The demographic characteristics of the participants are shown in Table 2. The age distribution is relatively evenly spread among the groups, with 18–24, 25–34, 35–54 and 55+ representing a quarter of the respondents, respectively. The gender distribution is nearly equal, with males and females each comprising 49% of the respondents and non-binary individuals making up 2% of the respondents.

The majority of respondents had at least completed high school, with 95% having a high school education or higher. A significant portion (47%) had completed university education. The most common household size among respondents is two individuals, comprising 37% of respondents. 33% of respondents live alone. A slight majority (53%) of respondents have a driver's license. However, car ownership is less common, with only 23% of respondents owning a car. For commuting, the most preferred mode of transport is the metro (40%), followed by buses (35%). For leisure, walking (49%) is the most preferred mode, followed by the metro (46%). For errands, walking (49%) is again

Table 2. Demographic Characteristics of Participants

Variables	Responses	Percentage (%)	Variables	Responses	Percentage (%)
Age group			Preferred transport mode for commute		
18–24	14	25	Car	14	25
25–34	14	25	Bus	20	35
35–54	15	25	Commuter train	16	28
55 and above	14	25	Metro	23	40
Gender			Tram	9	16
Male	28	49	E-bike	2	4
Female	28	49	Bike	4	7
Non-Binary	1	2	Walking	14	25
Education			Other	1	2
Primary School	57	100	Preferred transport mode for leisure		
High School	54	95	Car	15	26
University	27	47	Bus	18	32
Household Size			Commuter train	20	35
1	19	33	Metro	26	46
2	21	37	Tram	14	25
3	7	12	E-bike	3	5
4	5	9	Bike	8	14
5+	4	7	Walking	28	49
No info	1	2	Preferred transport mode for errands		
Driver's License			Car	14	25
Yes	30	53	Bus	17	30
No	26	46	Commuter train	8	14
No info	1	2	Metro	24	42
Car Ownership			Tram	7	12
Yes	13	23	E-bike	2	4
No	43	75	Bike	7	12
No info	1	2	Walking	28	49

Table 3. The Number of Groups of Two and Three Participants, and How the Order of the Conditions Was Counterbalanced in Our Experiment

Condition	Groups of Three	Groups of Two	Total Groups	Total Participants
Embodied first	8	3	11	30
Disembodied first	7	3	10	27

the most preferred mode, followed by the metro (42%). E-bike usage is relatively low across all transport categories, with only 4–5% of respondents preferring it. Tram usage is lower compared to other modes for both leisure and errands. Biking is more popular for leisure (14%) compared to commuting (7%) and errands (12%).

The participants were sorted into 21 groups of two or three. Although the experiment was designed under the assumption that three passengers would ride the bus, groups of two were allowed in the instances where one participant failed to appear at their booked time slot. In total, 15 groups had three participants and 6 groups had two participants. Table 3 shows how many groups of two and three participants experienced our conditions.

3.6 Analysis of Spoken Interactions on the Bus

To analyze the interactions between the passengers on the bus and the agent Elin (embodied or not), all recorded dialogues were first manually transcribed, and manually annotated based on who was the speaker and who was the addressee(s) for each utterance. Thus, we can distinguish between Human–Agent, Human–Human, and Agent–Human utterances. We also manually annotated how much time each participant spent on the bus (from getting on to getting off) so that we could calculate how many utterances they were speaking (or receiving) per hour and thus get comparable numbers across conditions. One of the rides had to be excluded from this analysis since the video from that ride was missing.

To further analyze the interactions, we used GPT-4 (gpt-4-1106-preview) from OpenAI [58] to automatically classify utterances into dialogue acts. GPT-4 was chosen over human annotators to enhance annotation efficiency, especially given the size of the dataset. Previous work have demonstrated the feasibility of using LLMs for zero-shot dialogue act classification [68]. While automated classification risks missing out on crucial insights into the data, we mitigated this risk by basing our annotation scheme on events we had already seen in real time during the experiment. Rather than using an existing dialogue act scheme (such as DAMSL [19]), we thus designed a more specific scheme to fit the scenario in the bus. Using the set of dialogue acts, GPT-4 was then prompted to incrementally assign one of the dialogue acts to each utterance while having access to the preceding dialogue (up to 10 turns). The prompt used for this is included in Appendix C. The scheme was applied and refined iteratively on parts of the data until we deemed the resulting scheme to appropriately cover the various types of utterances occurring throughout the dialogues. The final dialogue act scheme is presented in Section 4.1.1.

To verify the reliability of the coding scheme, we let one human annotator annotate two of the sessions and compared them with the GPT-4 annotations. This yielded a Cohen's kappa inter-annotator agreement score of 0.687 (a substantial agreement), which was deemed to be good enough for using the GPT-4 annotations for further analyses.

The dialogue act REQUEST-INFORMATION from the passengers toward the agent is especially interesting to analyze further, as those reflect the kind of information that an agent acting as a host on a bus should be able to answer. In total, there were 311 such requests in all analyzed dialogues. Again, a classification scheme was defined and refined iteratively, using GPT-4, in a similar way as with the dialogue acts. However, in this case, each utterance was classified without its context. This scheme is presented in Section 4.1.2.

3.7 Qualitative Data Collection and Analysis

After each group had taken its two rides, a semi-structured focus group interview was conducted to understand participants' needs, travel habits, and reflections on their rides with the autonomous buses. This interview was conducted with all participants in each group (two to three participants) present simultaneously, following a semi-structured interview guide to ensure consistency in all the interviews while allowing for open-ended discussions. The interview questions are listed in Appendix B.

We refer to these sessions as "semi-structured focus group interviews," as they combined a clear interview structure with space for participants to respond to each other. While the format differed from traditional focus groups regarding group size and moderation style, it followed the same idea of using group discussion to bring out different perspectives [1]. These formats are standard in applied and design-oriented research, especially when participants have shared an experience [35]. In our case, the group setting facilitated dynamic interactions, enabling participants to build on each other's responses in ways that might not have appeared in individual interviews. This provided stronger insights into shared experiences that simulated an actual bus ride environment for the participants.

However, we acknowledge the drawbacks to the semi-structured focus group interviews we conducted. As mentioned in [1], some of the drawbacks to semi-structured focus group interviews include dominant participants influencing the discussion; participants feeling inhibited from sharing their perspectives, and a risk of participants providing responses that they believe are expected, rather than sharing their actual opinions. Therefore, to mitigate these risks, the interviewers encouraged balanced participation, invited quieter participants to contribute, and maintained a neutral facilitation style throughout the session. The small group size contributed to more balanced discussions and made the sessions easier to moderate [45].

In total, 21 groups of participants, each consisting of 2–3 individuals who had experienced both rides (with the embodied robot and the voice assistant), participated in the interviews.

The interview data resulting from the semi-structured focus group interviews were analyzed using qualitative content analysis, “a research method for the subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns” [37, p. 1278]. This analysis method helps identify core aspects of the data that relate to the overarching purpose of the research. It also uses inductive reasoning, by which themes and categories were systematically identified in the data through the researcher’s careful examination and constant comparison, emphasizing that these patterns are constructed through interpretation rather than passively emerging from the data. The coding process was carried out through the MaxQDA software. Two researchers involved in the experimentation stages conducted the transcription and coding process per the eight steps suggested by Creswell and Creswell [20] and Tesch [71]. To validate the codes to maintain consistency and clarity, two researchers not involved in the experimentation were consulted under the inter-coder agreement [83]. Two rounds of coding were performed to finalize the code structure used to analyze the interviews. The categories and descriptions of the codes are shown in Table 4.

4 Results

The results are presented in an order corresponding to when the data were collected during the experiment. Section 4.1 presents an analysis of the spoken interaction between the passengers and the agent on the bus. Section 4.2 analyses responses to the post-questionnaires. Finally, Section 4.3 presents an analysis of the semi-structured focus group interviews we ran after the experiment.

4.1 Spoken Interactions on the Bus

Based on the analysis of the spoken interactions on the bus, as described in Section 3.6, the overall speech activity between the agent and the passengers in the two conditions is shown in Figure 7.

As can be seen, the most common speaker/addressee pair is Agent–Human, followed by Human–Agent and Human–Human. Thus, most interactions were between the agent and the passengers and not between the passengers. There is also an effect of the condition: There are more utterances from the passengers toward the agent when it was embodied (Wilcoxon signed-rank test; $W = 1023.50,\ z = 3.046,\ \mathrm{p} = 0.002$). Similarly, there were also somewhat more utterances from the agent toward passengers when it was embodied (Wilcoxon signed-rank test; $W = 1012.00,\ z = 2.320,\ \mathrm{p} = 0.021$). The difference in utterances between passengers was not significant (Wilcoxon signed-rank test; $W = 243.00,\ z = -0.670,\ \mathrm{p} = 0.508$).

4.1.1 *Dialogue Acts.*

The final dialogue act scheme that we derived at after iterative development with GPT-4, as described in Section 3.6, is shown in Table 5.

Table 4. Categories of Codes Used for the Thematic Analysis of the Semi-Structured Interview Data

Sub-Codes	**Descriptions**
Participants' Expectations (What Were the Participants Expecting from the Experiment)	
Previous experience	Participant awareness of or experience with the autonomous bus, the Furhat, or both
Feelings before participating	Curiosity, surprise, or open-mindedness before participating in the experiment
Preconceived notions/opinions	Preconceived notions about the technology, opinion on the purpose of the technology
Perception of the technology	Perception of the setup as something out of "science fiction," perception of the Furhat as a "toy"
Cultural expectations	Cultural background influencing the expectations
Interactions on the Bus (How Did the Participants Experience the Furhat and the Voice Assistant)	
Familiarity with voice assistants	Familiarity with voice assistants like "Siri" or "Alexa" influencing the experience
Language knowledge	Furhat's ability to acknowledge "English" or "Swedish" or another language influencing the experience
Forced interaction	The prominent presence of the Furhat leading to forced interactions
Interaction preferences	Participants' thoughts on the Furhat/voice-initiated conversations
Comfort level	The level of comfort or discomfort experienced during Furhat/voice interactions
Appearance and movement	Participants' reactions to the Furhat's physical attributes and actions
Humans vs. robots	Preferences regarding the presence of a human driver/host vs. a Furhat host
Voice assistant vs. physical robot	Preferences regarding the Furhat's vs. the voice assistant's communication style
Fun vs. utility	Perception of the balance between entertainment and practicality
Cultural comfort levels	Reactions to the Furhat's appearance and behavior within a cultural context
Bus Ride Experience (How Did the Participants Experience the Bus Ride)	
Capacity concerns/bus size	Considerations about the number of passengers, potential congestion, and impact on comfort
Ride-sharing	Willingness to share a ride with strangers in the smaller bus format
Seatbelt	Concerns and uncertainties regarding seatbelt compliance
Ordering the ride/operating the app	Concerns or positives of ordering the ride via the app
Accessibility features	Reflections on features for individuals with disabilities
Safety Perception (How Did the Participants Perceive the Safety on the Bus during Both Rides)	
Overall safety perception	General feelings of safety during the experience
Bus safety	Discussions regarding safety measures inside the bus
Ride-sharing	Concerns regarding sharing a ride with strangers who might pose a threat
Operator and control rooms	Reactions to the presence of an operator and control rooms in ensuring safety
Speed and control	Concerns or thoughts about the vehicle's speed and issues regarding loss of control
Surveillance	Concerns related to how the bus is being surveilled/recorded to be used in case of an incident
Inclusivity concerns	Concerns regarding inclusivity of the voice assistant and the Furhat for diverse user groups
Perceived threat of violence	Concerns regarding violent actions within or outside the bus
Authority figure	Reflections on whether the Furhat acts as an authority figure when rules are broken
Emergency services	Reflections on how emergency services could be engaged in case of an incident
Safety perception at night	Feelings of safety during night rides

Main codes are presented as full-width headlines, with the sub-codes and their descriptions contained below each headline.

Fig. 7. Number of utterances per hour and participant, broken down on speaker/addressee as well as experimental conditions.

Table 5. The Dialogue Act Scheme Used

Dialogue Act	Description
GREETING	Initiating a conversation or acknowledging each other's presence, like "Hej" (Hello), "Goddag" (Good day).
GIVE-INFORMATION	Providing details or answers, such as Elin introducing herself as a bus assistant, providing information about bus stops, or informing about the next stop.
REQUEST-INFORMATION	Seeking information from others. This includes Elin asking passengers about their destinations.
CONFIRMATION	Short responses that acknowledge or confirm what someone else has said, for example, "Okej" (Okay), "Ja" (Yes), "Visst" (Sure).
THANKS	Expressions of gratitude, such as "Tack så mycket" (Thank you very much).
INSTRUCTION	When Elin gives directions or instructions to the passengers, like advising them to sit down and buckle up.
FEEDBACK	Elin asking passengers about what they think a robot like her should be able to do in a bus, or passengers giving such feedback.
FAREWELL	Ending conversations, such as "Hejdå" (Goodbye).
SMALL-TALK	Casual or off-topic conversations, such as the discussion between passengers about the bus feeling like a car or their plans.
APOLOGY	Expressions of regret for an inconvenience or mistake.
CLARIFICATION	Seeking to clear up confusion or misunderstandings.
OTHER	Something that does not fit any of the above. These utterances are characterized by non-lexical vocalizations, laughter, unclear or indistinct speech, minimal or fragmented responses, hesitation markers, interjections, and repetitive discourse fillers.

Figure 8 shows the breakdown of dialogue acts in utterances directed from humans toward the agent. We can see that the biggest differences between experimental conditions are in REQUEST-INFORMATION, THANKS, and CONFIRMATION, which are more frequent in the embodied condition. The latter two typically serve as acknowledgments after having requested and received information, but the differences are larger, so it might also be that the users produce more thanks and confirmations in general to the robot (i.e., there is generally a stronger feeling that you need to acknowledge what the robot says than what the speaker says). There are also more FAREWELLs to

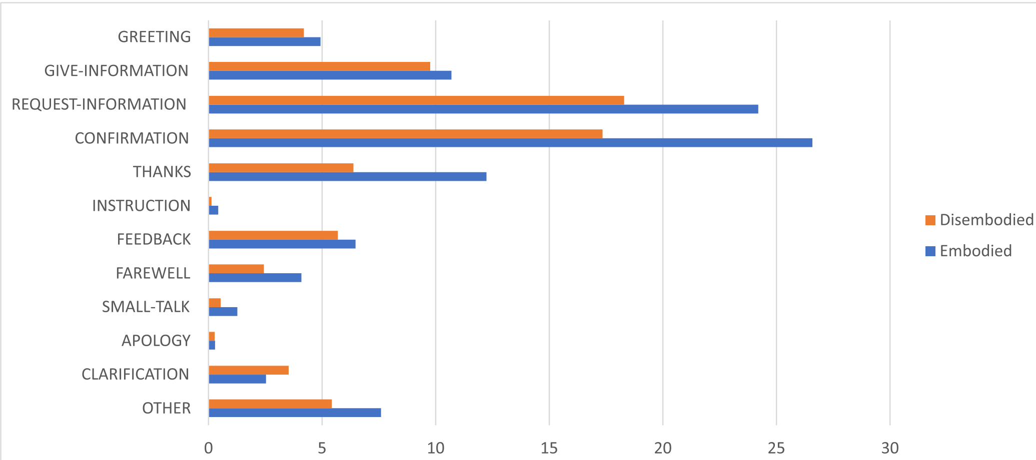

Fig. 8. Number of utterances *from passengers toward the agent*, per hour and participant, broken down on dialogue acts as well as experimental conditions.

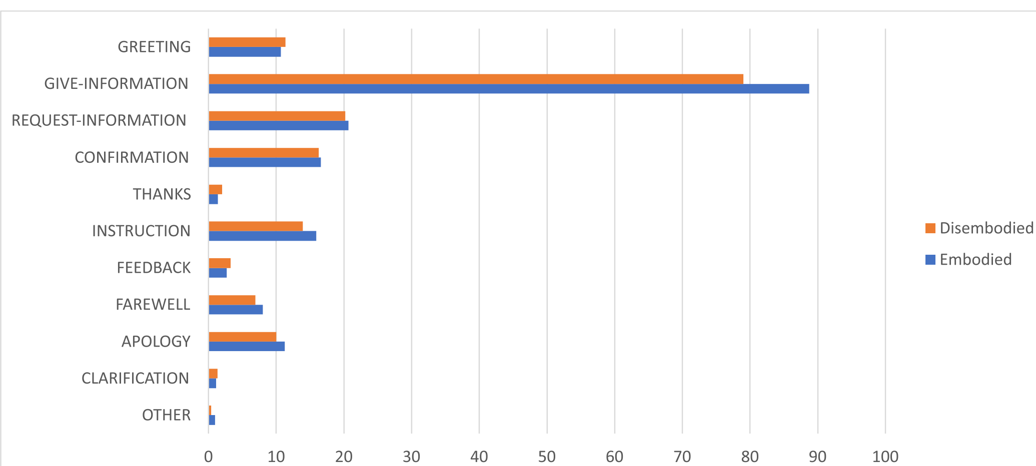

Fig. 9. Number of utterances *from the agent toward passengers*, per hour and participant, broken down on dialogue acts as well as experimental conditions.

the robot, indicating that they felt a bigger need to say farewell compared to the speaker. These observations indicate a higher level of anthropomorphization.

Figure 9 shows the breakdown of dialogue acts in utterances directed from the agent toward humans. There are no clear differences between the two conditions, which is reassuring, as it indicates that the Wizard behaved similarly in the two conditions. One exception is GIVE-INFORMATION, which is likely because the users also requested more information (as discussed above).

Figure 10 shows the breakdown of dialogue acts in utterances between the participants. There are no clear patterns here, the main trend is mostly that there is more talking in general between the humans in the disembodied condition, as already noted above.

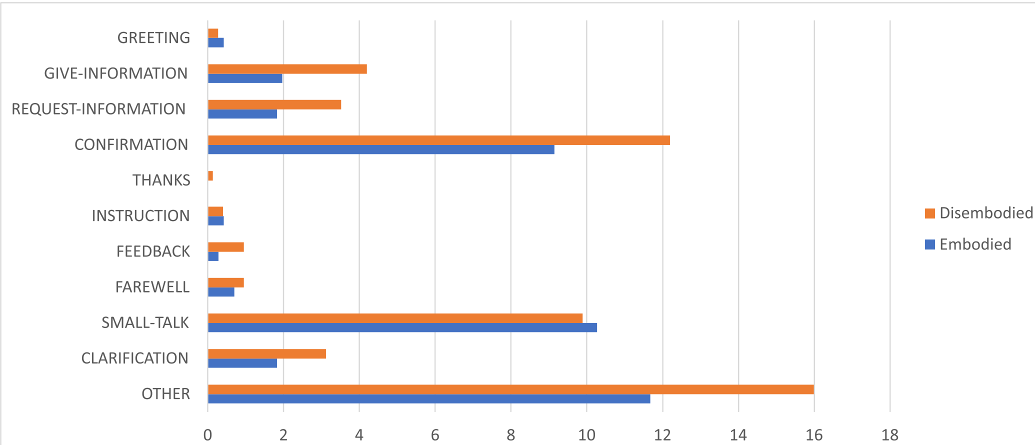

Fig. 10. Number of utterances per hour *between passengers*, broken down on dialogue acts as well as experimental conditions.

4.1.2 Information Requests to the Agent. As described in Section 3.6, the dialogue act REQUEST-INFORMATION, from the passengers toward the agent, was further categorized using GPT-4. The resulting list of categories is shown in Table 6, as well as their frequencies in the two conditions. As could perhaps be expected, Personal Questions are much more frequent toward the robot than the speaker, but there are also more questions in other categories as well, such as Clarify Situation, suggesting that the embodied condition makes people more prone to ask more task-oriented questions.

4.2 Questionnaires

After the first of their two runs, participants filled in a questionnaire that started by asking *if something special happened* during the run, followed by space for participants to write down their memories of surprising events that had happened during the run. Participants could also choose to not write anything if they believed that the run had finished without any unexpected events. All questionnaires were presented to the participants in Swedish. The questions presented here have been translated back to English.

Our primary evaluation of the participant experience of the agent was done via the Godspeed [7] questionnaires. Before the questions taken from Godspeed, we asked six Likert-scale questions evaluating aspects of the ride that we were expecting beforehand to be complicated with the embodiments we had chosen to evaluate in the bus environment. While these questions did not come from a validated questionnaire, passenger opinions on the concepts measured by them were further explored in the semi-structured focus group interviews. These six questions were phrased as statements that the participants could rank from *fully disagree* to *fully agree*. These performance-related questions were:

- Elin helped me understand why the bus had stopped when it stopped for technical reasons.
- When Elin talked to a passenger, the turn-taking (who started and stopped talking at what time) worked well.
- It was easy to understand what Elin said.
- Elin understood us passengers when we spoke.

Table 6. Breakdown of the Dialogue Act REQUEST-INFORMATION from Passengers toward the Agent, in Terms of Number of Requests per Hour per Passenger, as well as Distributions

Category	Examples	Embodied	Disemb.
Location and Direction Queries	Asking about current, next, or specific stops (e.g., "Where are we now?" "Where is the next stop?"). Inquiring about distances or travel time to specific destinations (e.g., "How far is it to Apple Alley?" "How long time until we reach Radish Road?"). Questions about nearby amenities (e.g., "Are there any good restaurants close to Plum Path?" "Where is the closest café?").	8.5 (35%)	7.0 (38.1%)
Capabilities and Functions of the Agent	Questions about the agent's capabilities, including language (e.g., "What are you programmed to answer?" "Can you speak English?"). Asking about the agent's awareness or perception abilities (e.g., "Can you see us?" "Can you hear us?").	2.6 (10.7%)	1.8 (9.7%)
Requests for Action	Requests for specific actions from the agent or the bus (e.g., "Can you lower the bus at the next stop?" "Can you inform me when we are there?").	1.4 (5.6%)	1 (5.2%)
Operational and Technical Questions	Inquiries about the operational aspects of the bus (e.g., "What is the top speed for this bus?" "Can strollers go on this bus?" "Is the bus electric?").	2.5 (10.2%)	1.6 (9%)
Clarify Situation	Questions about why something is happening (e.g., "Why doesn't the bus move faster?" "Why did we stop?").	1.8 (7.3%)	0.5 (3%)
Safety-Related Questions	Questions about safety and emergency procedures (e.g., "What happens if there is an emergency situation?" "Can you guarantee the safety of the passengers?").	0.4 (1.7%)	0.3 (1.5%)
Personal Questions	Questions seeking personal information about the agent (e.g., "How old are you, Elin?" "What kind of AI are you?"). Queries probing the agent's "thoughts" or "feelings" (e.g., "What are you thinking of?" "Do you like your job?").	2.2 (9%)	0.7 (3.7%)
General Information	Questions about information not directly related to the trip (e.g., "What is the weather tomorrow?" "Do you know what day it is today?").	0.7 (2.8%)	0.1 (0.7%)
Other	Not matching any of the above, including truncated utterances or ellipses.	4.2 (17.5%)	5.3 (29.1%)

Examples translated from Swedish.

— Elin gave reasonable answers to questions.
— Elin talked too much.

These six questions were followed by translated versions of the Godspeed forms by Bartneck et al. [7], accounting for 22 questions, to assess the passengers' perception of the agent. One question from Godspeed, *moving rigidly/moving elegantly*, was not asked as it did not apply to the disembodied agent (which, by its nature, did not have any moving parts).

After their second run, participants answered the same questions about unusual events during the ride. They also answered the first six questions on the agent's behavior once more, phrased identically to the first run. This was followed up by a variant of the Godspeed form expressed as a comparison to the first ride; instead of starting with a comparison between *Elin was fake* and *Elin was natural* as in the first run, the questions were rephrased as *Elin was more fake than in the*

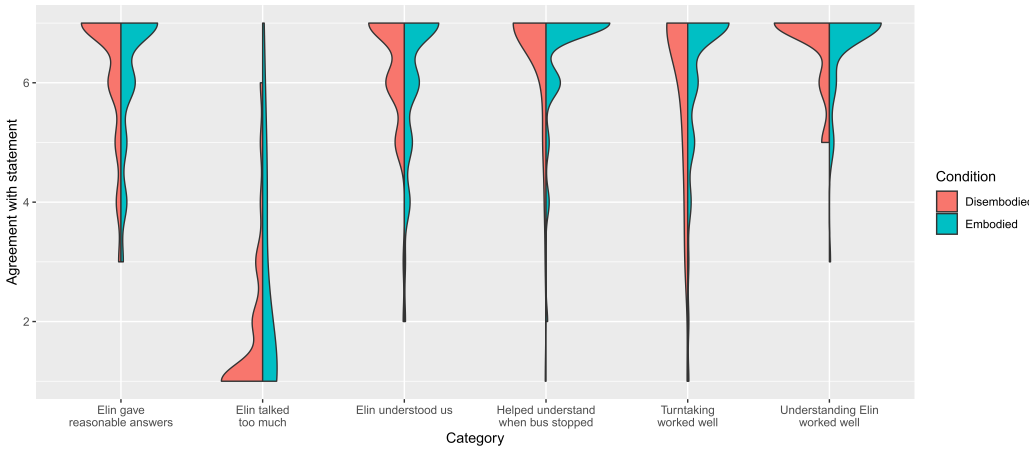

Fig. 11. The distribution of answers given to the first six questions, related to the agent's performance, asked after both runs. A one (bottom) expressed maximal disagreement with the statement, and a seven (top) expressed maximal agreement. The disembodied condition is displayed in red on the left side of each violin plot, with the embodied condition in teal on the right side.

first run compared to *Elin was more natural than in the first run*. This was done for all 22 Godspeed questions.

4.2.1 First Six Questions. Figure 11 shows how participants respond to the performance-related questions after each run; these questions are the same after both the first and second run. In Figure 11, the answers given in response to the *disembodied* condition (both for those participants where it is the first and the second run) are shown in red on the left side of each bar, while those given to the *embodied* condition are shown in teal color on the right side of each bar. Participants generally gave a high rating in both conditions on all questions except *Elin talked too much*; on that question, both conditions are given a very low rating. All questions appear to show a ceiling effect where both the embodied and the disembodied system are generally evaluated as positively as possible on each question.

4.2.2 Godspeed Questions. Rather than consider the answers to each Godspeed question separately (which would lead to a large number of comparisons), we average the questions by their categories as defined by Bartneck et al. [7]. For *animacy*, we cannot include the *moving elegantly* question in the average as it was not asked. For *perceived safety*, the responses to *calm/agitated* and *still/surprised* were inverted compared to *anxious/relaxed*; this was compensated for by adding the inverse ($8 - x$ where x is between 1 and 7) of the first two to the mean rather than the direct answer.

For the second run, we do not have absolute numbers for the participants' ratings as we asked them to rate the difference between the first run and the second. An approximation of what participants would have rated the second system can thus be created by adding the relative numbers from the second run to the absolute numbers from the first run. To make these extrapolated numbers comparable to numbers from the first run, we capped them between 1 and 7. A visualization of the extrapolated numbers from the second run plotted alongside the numbers from the first run is shown in Figure 12. Both *perceived safety* and *animacy* appear visually different between the conditions.

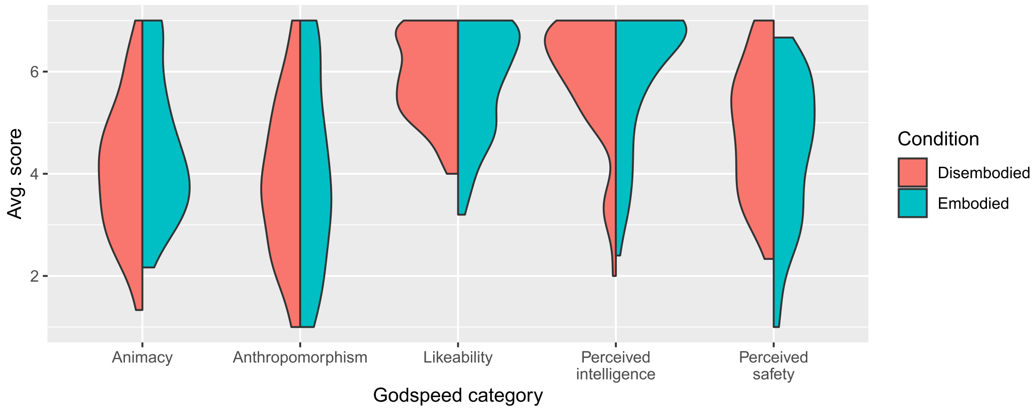

Fig. 12. The distribution of responses to each category of Godspeed questions. Questions have been averaged by their Godspeed questionnaire (animacy, anthropomorphism, etc.) as defined by Bartneck et al. [7]. For each category, the left half of the violin plot shows answers given to the *disembodied* condition, and the right half shows answers given to the *embodied* condition. See Section 4.2.2 for notes on how answers from the second run were considered.

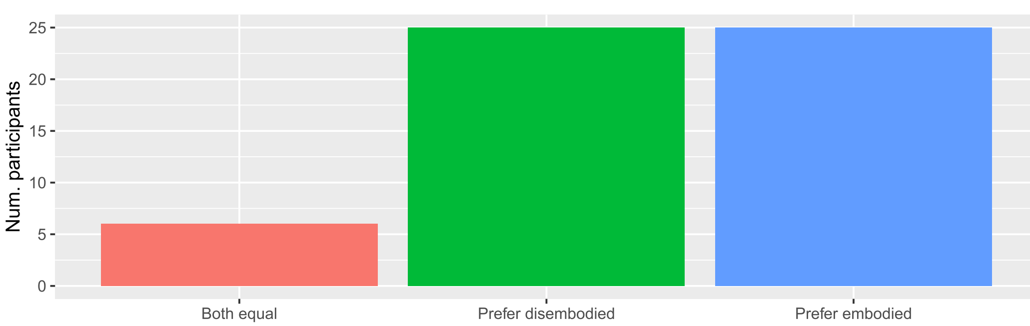

Fig. 13. The counts of positive, negative, and neutral scores given to the embodied condition over the disembodied condition for the Godspeed measure *like/dislike*.

4.2.3 Nonparametric Comparisons between the Embodied and Disembodied Runs. The Wilcoxon signed-rank test was applied to the questions from our questionnaire. Following the method proposed by Pratt [63], zero values in the dataset (indicating that the participant did not consider one ride to be better than the other) are resolved by reducing the sample size by one, while still shifting the rank sums up by one. After applying this correction, no comparisons were significantly in favor of the embodied or disembodied agent. The test statistics and critical values for each of these comparisons is reported in Appendix D.

4.2.4 Preference between Embodied vs. Disembodied Agent. The counts of positive, negative, and neutral scores toward the comparison *I liked embodied Elin more than disembodied Elin* are visualized in Figure 13. Twenty-five participants expressed a preference for the embodied version of Elin, 25 expressed a preference for the disembodied version, and 6 expressed no preference. Thus, there is no clear winner and preferences appear to be very individual.

4.3 Interviews

In the following sections, we present our findings around the four key themes that were identified in the analysis in Table 4.

4.3.1 Participant Expectations from the Elin Agent. The group interviews revealed diverse reactions among participants regarding their expectations for the experiment involving a robot inside an autonomous bus. While 41 participants approached the experience with curiosity and an open mind regarding the Elin agent, as introduced in the participant priming, 31 participants expressed surprise and even apprehension upon encountering the physical robot onboard. Initial reactions ranged from excitement and enjoyment (20 out of 57) to skepticism and mild discomfort (32 out of 57).

> *"I was very curious because I had never done something like this before. It was a really good experience for me. I didn't think Elin would be a robot, it wasn't in my mind originally. I am impressed."*

Although the addition of the robot was unexpected for many (14 out of 57), this was particularly surprising when the robot made remarks about what is acceptable behavior on the bus, such as drinking coffee (seven out of 57).

> *"I did not expect a robot. So, it came as a bit of a shock. A face moving and looking at you. I thought it would be just a voice, I didn't think it would interact so much with us. She told us it was okay to drink coffee, and we were so surprised."*

Fourteen participants found the Furhat robot's embodiment, along with its head movements and facial expressions, surprising and even frightening.

> *"I was looking at the mix of a robot and a human. I get scared of that, and I wasn't expecting it to move around, looking at everybody, trying to smile at some point. That was a little scary in the beginning. And it was quite big; I was expecting something more like Alexa."*

4.3.2 Interactions on the Bus. The survey data presented in Section 4.2.4 showed a division in passenger preferences; some favored the robot, while others preferred the voice assistant. We found that this division was also evident in the interviews. Thirteen participants explicitly felt comfortable with Elin the robot and often described the experience as *"fun."* However, 18 participants explicitly preferred the voice assistant, considering it more appropriate for a less intrusive experience during their daily travel. They found Elin's physical appearance and movements slightly unsettling, with 11 participants describing the robot as *"uncanny valley-like."* Despite this, eight participants acknowledged that spending more time with Elin in its embodied form might help them become more familiar with her movements and mannerisms.

> *"Fun! Like the first trip with actual interaction with the face was much more fun. The second [non-robotic] one felt more like any other ordinary bus ride [compared to the usual local bus ride] bit less exciting and new."*

> *"I certainly did not like Elin! (Interviewer: As a robot?) Because you are used to the subways and all. You hear the calls, this is what you are used to and this [the robot experience] was something else. It was not scary, but I really don't like her. This reminds me of the pop music videos with the dancers with faces [masks?] like this or space movies or something. It is very unnatural. It is difficult to make a robot look human. So, no I did not like her. Perhaps if you get used to it, but today I did not like her."*

We also found that nine participants explicitly perceived Elin in her robot form to be *"smart"* in comparison to Elin in her voice assistant form.

> *"I thought she was dumber in the beginning [voice assistant first ride], but she sounded smarter in the second ride [with the robot] just because you can see her you know. Otherwise [voice assistant form], it was just like talking to a bus and that's a bit stupid. Like "Oh the bus can't answer this question" but it felt more sense when she [the robotic Elin] was answering the questions in the second ride."*

Suggestions for improvement made by our participants included incorporating visual feedback to enhance the user experience (six participants), providing more detailed information about travel times (five participants), and addressing language barriers to ensure inclusiveness for non-Swedish or English-speaking individuals (seven participants).

> *"I would say you get used to it, the Swedish, I only need to remember the name of the stop. but sometimes abnormal things happen, someone will say something very fast in Swedish, or in English, and there is a higher chance I will miss it. And sometimes I will ask others for help and sometimes I just accept the result."*

The need for communication through a screen/app was mentioned in the context of the robot's continuous movement, which was found to be distracting.

> *"Thinking about Elin as the head, it was twitchy, continuously moving and looking around, I don't know if that's something you can tone down or what, but it was something you reacted a little bit to. Maybe you get used to it over time, but it is moving all the time looking around, smiling, showing it's still here, not just sitting passively, showing it hasn't turned off. For me, it took a little bit of adjusting to that, a bit of an internal smile that it is working a little too hard to show me that it's here. But again, that's the first time I've ever seen one maybe over time [unclear] I might think differently about that."*

Five participants also discussed the possible functions the robot should have such as communications with the bus stop and the ride booking app to minimize unnecessary stops and delays.

> *"Can Elin interact with outside as well as inside? For example: Can Elin interact with the bus stop and the app and see if no one is there [in the case someone booked the ride but did not show up] and inform the bus to not stop and open the doors and wait for this person to minimize the time frame"*

4.3.3 Bus Ride Experience. Five participants explicitly appreciated the compact size of the bus, likening it to a taxi and finding it exclusive and cost-effective.

> *"Yes, this bus is small, almost like a taxi. I think it is nice. It makes the ride cheaper and it makes it exclusive."*

Four participants explicitly expressed concerns about traveling in smaller spaces, particularly in light of the pandemic.

> *"It has been even worse after the pandemic. It is terrible sometimes! People are so scared of traveling together in smaller spaces."*

Mixed opinions were expressed regarding sharing the ride with others, with seven participants stating feeling comfortable in intimate settings that fostered increased interaction among passengers, while 11 participants voiced concerns about potential discomfort, especially in crowded conditions. Ways to exit the space were considered, to leave the crowded situation. Having only one exit, as this bus does, was seen as a downside. More seats would also allow a passenger who wants to avoid crowds to move to less busy areas.

> *"Interviewer: Imagine if the bus was full and two people were also standing here. P3: aw, too much, no! P2: yeah, yeah, no it's so small, you'd get claustrophobic. If there were more exits or more seats it would have been a little bit safer. then at least you have options."*

Questions were raised about the bus capacity and speed, highlighting considerations about comfort, efficiency, and utility in their daily life. The bus seems to find no place among the different transportation modalities available to some participants (6 out of 57). The trip is too short and can be easily done on foot or biking.

> *"Compared to the tunnelbana [Stockholm Metro], I would say, this bus is not for me right now. If it would be out operating now, I would not take the bus. But I can see the functionality, people having like, functionality issues taking it and I love it. I can see the future of people who don't have a car but want to go to the shopping center and carry heavy stuff, taking the car. But it's not for me, it's not for my everyday life. It is too short of a ride, I would rather bike. I am a fast person, I would never take this bus."*

One participant questioned the efficiency of the autonomous bus system, as they perceived it as a potentially inefficient method of transport when unexpected issues arose.

> *"For me on the first ride, the bus stopped, and the doors opened but no one got in or out. So, it said [Elin the voice] that you need the driver [the onboard operator] to see if someone is standing in front of the bus. She explained that someone must have ordered the bus but did not show up. Then I thought maybe this is a very inefficient way to ride the bus."*

In terms of the voice assistant and the robot, three participants suggested equipping the robot with "hands" for passengers to hold. They also reflected on the placement of the embodied robot for better interaction.

> *"Maybe you can give her some hands so people can hold it and not be scared. The only thing I was thinking of was the placement of the face. I don't know where the face would be in the bus but if it was somewhere on the top, somewhere where the driver would be but facing the passengers maybe it would be better."*

4.3.4 Safety Perception. Safety concerns were the focal point of our interviews, with participants initially expressing varying levels of concern about the overall safety perception of the autonomous bus. However, since the bus speed was around 10 km/h, the passengers felt quite safe. Sixteen participants explicitly highlighted that Elin giving them information when an incident occurred made them feel reassured that they were safe. The interviews also suggest that 12 participants explicitly expressed feeling slightly more comfortable by the second ride, which suggested that those participants needed time to get familiar with both the information assistance systems as well the autonomous bus. These preferences for the first or second ride were not statistically significantly visible in the questionnaires analyzed in Section 4.2, and we can thus only treat it as an explanation for why some passengers may have preferred the ride that they preferred.

> *"Just it saying if it stopped because a car got too close if it hadn't said anything you would be wondering, but when it says 'we stopped because of this reason,' it felt a lot safer because it didn't feel like it was breaking down and I had no idea."*

The onboard operator who needed to be on the bus because of the regulations in Sweden also contributed to the feeling of safety among the participants (even though the participants were asked not to interact with the operator). Sometimes the on-board operator needed to drive the bus manually with a joystick to get around parked cars or roadworks; this increased the feeling of safety. Nineteen participants stated that the presence of an actual human operator was a lot more comforting.

> *"The actual person who was there made me feel reassured. When there were some traffic stops the manual driver [on board operator] took over and drove the bus, I don't know what would have happened without him, at that kind of traffic stop. maybe we would have just been left standing there. So I guess that person made me feel a lot more relaxed."*

Safety, as introduced to participants in the interviews, was perceived in different ways. Seventeen participants thought about safety protection in case of collisions and compared familiar safety features of cars to those of the bus.

The trained operator on the bus normally had the job of ensuring that each passenger put on a seatbelt. We integrated this into the script our agent followed as passengers got on the bus, and six participants specifically mentioned following the instruction to put on a seatbelt, while others ignored it, with three participants stating that it was perhaps not warranted with the low speeds, or because they do not usually wear seatbelts in buses. Five participants suggested that the robot should be more assertive in ensuring passengers comply with safety protocols, including putting on their seatbelts.

> *"Maybe forcing/ engaging people to put on their seatbelts. I am a safety guy even though I don't follow this rule myself."*

Nine participants specifically perceived the notion of safety concerning dangers that could come from other passengers. This included gender-related safety, for example, women traveling at night, or women traveling with a group of men.

> *"As long as they're not getting aggressive, I think it's [more of a concern for] women. Nighttime, I think the problem will occur at night, you sit here alone and an aggressive person. Then it's good if there's that communication we talked about."*

> *"If I was a woman alone and a group of guys were getting on the bus and sitting close to me, maybe then it's not so nice?"*

Another way to conceptualize safety for our participants was related to information security and privacy. Five passengers specifically discussed how communicating with the robot out loud in front of other passengers (essentially, strangers) could be an issue in certain situations, matching concerns mentioned by Fosch Villaronga et al. [30].

> *"The fact that you are telling [out loud] the bus where you should get off when you enter it, it can sometimes be uncomfortable and scary."*

Six of our participants also questioned how the bus is being monitored and suggested alternatives such as emergency buttons under the seats, private communication options such as over an app or a screen and speakers at each seat, or being able to change their trip during the ride in case of threatening situations.

> *"Instead of talking to the AI [meaning the voice or the robot] I would prefer to write my questions to the AI? In an app or a small icon. When a lot of people talk to it, it might not be able to focus on one person alone."*

"I think it also depends on if someone is watching this camera footage or is it being recorded as evidence. Because that makes a difference. Also, it would help to have emergency buttons under the seat if someone is threatening you."

"It would be nice to have a feature [...] which can be used at nighttime when there are idiots on the bus so I can change the destination and stop at the very next stop coming up."

Passengers largely questioned whether Elin the robot could act as an authority figure and often began to look around the cabin for cameras, bringing up surveillance-related concerns when asked about the feeling of safety. Six participants raised these concerns. While three participants stated that the presence of the physical robot could bring about a feeling of authority, three participants still believed that they needed a human presence. Three participants also specifically raised concerns regarding vandalism against the robot.

"I guess if something is happening to me, I am locked in this bus with someone else. I guess that is not safe. But, I don't know if I can tell Elin to call the police."

Our participants started realizing the non-driving roles of the bus driver, especially since there was no traditional driver on the bus, and only a (mostly passive) onboard operator was present due to legal restrictions. Five participants reflected on this. Four participants reflected that the human bus driver/operator was more than just a driver but also someone who could exercise authority when needed, for example, when passengers intrude on one's personal space or become violent with each other.

"It can't really help you if someone is being violent or physical. If there is a person driver then it is a bit better."

"If I am on the usual bus, the bus driver can help you if someone starts to fight, and here you are alone. If something happens it's like 'how....'"

Besides ensuring passenger safety, the role of the human bus driver/operator becomes more evident regarding passenger behavior rules, such as smoking interdiction. Three participants specifically mentioned this.

"The problem is, even if there are rules that everyone must follow, if there is no driver, you don't know if they [passengers] follow the rules. If someone wants to smoke, they will smoke."

Eight participants started to reflect on how surveillance technologies can potentially be used to ensure compliance. The single camera on the bus ceiling was regarded as a safety weakness because it could be easily disabled through physical covering, rendering the bus completely unsafe.

"Is it only this camera in the ceiling? if it's only one, people who want to do bad stuff can cover the camera and do whatever they want. I would want to know that there were several cameras. So they don't just silver-tape the camera. One guy blocking the camera and the other can do what they want. I would want to know that there are some hidden cameras."

Additionally, one of the participants also shared instances where the robot provided explanations for the bus stopping, despite which the participant preferred a normal bus over an autonomous one but still acknowledged that their preference might change as the technology evolves.

"There are always some questions about safety that I would ask [to the robot]. For example, when the car was driving too close to us, the robot told us that the car had come quite close to us and that's why the bus stopped. There was also another situation where the robot told us the bus stopped due to technical issues. So, if I had to choose this bus over a normal bus, I would choose

the normal bus. But I am aware that this technology is still evolving so maybe in the future I would choose this bus."

Three participants also expressed specific concerns about the absence of airbags and the need for more clearly differentiated seat belts for improved safety.

"We don't have any airbags in here, but well most buses don't have them."

"I liked that Elin reminded me to put on my seat belt. I would prefer them in another color, so she knows I have put them on, so we are not driving before we have them on. I don't want anyone standing in the bus. We need to be safe, because I would feel more safe with the belts on."

Reflecting on the capacity of the bus (8–12 passengers allowed, 6 seated and the rest standing), one participant stated:

"I work with security systems. I don't see the problem for the people sitting but I wouldn't prefer if people are standing in this bus from a safety perspective."

Five participants stated that the bus speed was considered safe and adequate for short distances (covered during the experiment) but too slow for longer journeys. Three participants explicitly stated that if the bus ran at higher speeds, they would consider it for everyday travel.

"I also thought maybe a little bit slow but considering everything, the area, the circumstances, no problem, but I thought for longer distances would maybe feel a bit too slow."

Overall, the interviews suggest that while there were some mixed feelings regarding the safety and practicality of such combinations of technologies for daily use, the participants expressed that such integrated technologies are likely to be part of our daily lives and suggested improvements that would encourage them to use such services regularly. However, some of them stated that not everyone might need such services and identified key user groups such as the elderly, tourists, and people with limited mobility that might make the most out of these combined technologies.

5 Discussion and Design Implications

To address the research questions we posed in Section 1, we consolidated the findings from our results into design implications for how to implement robotic assistants in the future of APT. The design implications are divided into three parts. Section 5.1 is about concerns about embodiment, answering Research Question 1. Section 5.2 is about the capabilities of the agent, answering RQ2. Finally, Section 5.3 is about safety concerns in the bus, addressing RQ3.

5.1 Embodiment

Our findings in this subsection address *RQ1: What are the differences between a disembodied and an embodied agent serving as a host on a self-driving bus, in terms of how passengers perceive it and interact with it?*

The analysis of the passenger interaction with the agent in Section 4.1 revealed that they initiated interactions with the robot more often than with the speaker, matching earlier results seen by Fischer et al. [28]. One possible explanation for this is that the physical embodiment offers a clearer affordance for interaction. Another potential explanation is that they felt more obliged to talk to the robot than the speaker, given its physical presence. The analysis also showed that they were much likelier to follow up the robot's response with an acknowledgment, such as an expression of gratitude ("thank you") or confirmation ("okay," "yes," "sure"), compared to the disembodied agent. These can be seen as indications of anthropomorphization [84], as they are not strictly needed from an information exchange perspective.

One theoretical basis for analyzing the differences we see in Figure 8 is to consider the exchanges with the disembodied robot as isolated question–answer exchanges, while the exchanges toward the embodied robot tends more toward an environment of *incipient talk*, as informally defined by Schegloff and Sacks [65]. In a continuing state of incipient talk, individuals who exist in a shared space do not feel the need to reintroduce or reopen the conversation even after a longer adjournment of speech activity [9, 65]. This describes the communicative environment that occasionally formed between our *passengers*. We believe that the embodiment of the agent affected whether passengers were likely to consider the robot part of that continuing state of incipient talk, but we also see an increase in task-based dialogue from passengers toward the agent (the GIVE-INFORMATION and REQUEST-INFORMATION categories in Figure 8). To fully explore how an embodied robot becomes part of a continuing state of incipient talk, care must be taken to control for the social environment between the humans in the robot's environment. Gilmartin et al. [31] also point to the importance of active listening—laughing and reacting—in casual talk [31]. Our agent by design did not laugh, and it could not provide communicative backchannels beyond gaze to confirm attention. It was thus not able to participate in the continuing state of incipient talk as effectively as the human passengers could—whether it was embodied or disembodied.

A bit surprisingly, the Godspeed [7] post-questionnaire (Section 4.2.2) did not reveal any consistent differences in the perception of the embodied and disembodied agents. However, the interviews with the passengers (Section 4.3.2) revealed that they perceived the embodied agent as *someone* giving information, whereas the disembodied voice was perceived as coming from the bus itself. This could perhaps help to explain why passengers behaved differently toward the agent in the two conditions. Additionally, some passengers found the embodied agent *uncanny valley-like* due to its movements and facial expressions, which could be one of the reasons why some passengers preferred the disembodied agent.

The questionnaire related to the agent performance (its ability to understand, give reasonable answers, and take turns, as analyzed in Section 4.2), also did not reveal any differences. However, in those questions, there were clear ceiling effects, as shown in Figure 11. Both agents were given extremely positive assessments. It is unclear whether this reflects the passengers' true perception of the agents: while the Wizard-of-Oz setup made sure that the agent would give reasonable answers, the slow response times should also have impacted the perception of turn-taking. We will choose to interpret the generally positive feedback toward both agents seen in these questions as indicating that any feedback the passengers later gave in the interviews is not because of fundamental flaws in the quality of the interaction with the agent.

Given the multi-party setting of the bus, we would have expected the embodied Furhat robot to allow for better turn-taking than the disembodied agent, as previous studies have shown [67]. In multi-party settings, it can be confusing to know who is addressing whom, for example, whether the agent is addressing a specific passenger or all of them. A physical robot with an anthropomorphic gaze could potentially help to indicate the addressee. Due to the ceiling effect described above, we were not able to see any such differences between the embodied and disembodied conditions. It also did not come up during the semi-structured focus group interviews. However, as shown in Figure 8, passengers were in general more inclined to ask questions to the robot than to the speaker.

In the questionnaire (Section 4.2.4), the passengers were evenly divided on their preference for an embodied or disembodied agent. In the semi-structured focus group interviews, some passengers said that they perceived the embodied agent as more of a novelty and that they would likely prefer a disembodied agent for their daily travel needs. However, others expressed that familiarity with the robot could potentially come with more experience with it.

Embodiment is a difficult design choice to make in a shared environment like the bus. All passenger preferences cannot be satisfied at the same time with a single embodiment. Other options, which we have not explored in this study, would be a non-speech-based interface, like an information kiosk, or a smartphone app. For the latter alternative, it is not certain that all passengers would want to download an app to their smartphone to get answers to questions they might have during the ride. Thus, a mix of options might be preferable. Luria et al. showed that mixed embodiment and re-embodiment led to experiment participants feeling uncomfortable [50]. Alternatively, each bus could have its agent personality, to prevent passengers from treating disconnected agents on separate vehicles as being representations of the same agent identity—this would partially address the concerns seen by Luria et al. [50].

For future work, we would like to explore if passenger preference for an embodied or disembodied agent changes if the bus environment feels less safe or less secure. This is further explored in Section 5.3. Although it was out of scope for this study, future work could also integrate the embodiment of the agent with the AV itself, creating non-humanoid agents more like those explored by Kim et al. [44]. Such embodiments could get the best bits of both of our proposed agents. Passengers could interact with the agent as they would with the familiar disembodied announcement voice in a bus, while we could retain the positive effects of having the agent be co-present in the bus environment with the passengers [52].

5.2 System Capabilities and Offered Services

Our findings in this subsection address *RQ2: What services could an agent (embodied or not) provide in APT and what would passengers want to use it for?*

The analysis of the interactions on the bus presented in Section 4.1.2 showed that the single most common type of requests from the passengers toward the agent were related to location and directions, which is not very surprising. Other fairly frequent requests related to the ride include clarification of the situation (e.g., “why did we stop?”), requests for action (e.g., “Can you lower the bus [for wheelchair/stroller accessibility] at the next stop?”), and technical questions (e.g., “Can strollers go on this bus?”). In the interviews (Section 4.3), some passengers said that they mostly expected the agent to give them information about their travel experience. However, they expressed being impressed by the agent’s responses to requests for situational clarifications and safety precautions, such as reasons for the bus stopping or whether beverages could be consumed inside the bus. The agent was thus largely able to address the commonly seen feeling of lack of control in APT [32, 36].

If there is no agent present to answer these questions in a self-driving bus, there should probably be some other way for the passengers to get answers to them, for example, through a smartphone app or some kind of information kiosk onboard the bus. The types of information passengers expressed a preference for getting from our agent was specific to the exact vehicle they were traveling in, and not generic information about the transit system as a whole.

Many passengers expressed a preference for getting information on their phones but said that an interactive agent on the bus could be interesting for disabled or elderly passengers, as well as for tourists who need extra guidance.

Connecting to the next section on safety and privacy, passengers expressed a need for the ability to call emergency services through the agent. While alert buttons to call such services generally exist in Stockholm’s public transit, the conversational interface may give incapacitated passengers a way to call for help they would otherwise not be able to get.

5.3 Passenger Safety and Privacy

Our findings in this subsection address *RQ3: Does the presence of an agent in APT address passengers' feelings of lack of safety or insecurity prompted by the absence of an authoritative human or smaller vehicles?*

As highlighted in Section 2.1, one of the key concerns passengers of APT generally have is that of lack of personal safety [25, 48]. When talking about safety, participants in our semi-structured focus group interviews tended to focus on concerns about both the safety of the APT vehicle in traffic and interpersonal safety concerning other passengers. Their concerns regarding traffic-related safety came from frequent stops for unclear traffic-related reasons. Even though the passengers had no reason to assume that the other passengers they were traveling with would be dangerous, when asked during the interviews how they imagined taking an autonomous bus in other situations with unknown co-passengers at night, some passengers expressed that they would not want to take the bus at all in such a situation. Others expressed that they did not generally feel unsafe in public transit and that they would have no problem riding APT in any such situations.

Additionally, as highlighted in Section 2.1, one of the factors influencing passenger willingness to use APT vehicles is the lack of authority or supervision within the buses [24, 54]. Even with the presence of the embodied robot, passengers often found more comfort in the fact there was a human onboard operator always present (even though they were not allowed to interact with them). The human onboard operator was expected to resolve any passenger-related issues as well as stop a traffic incident. The presence or absence of an agent did not play into these judgments.

Even in terms of traffic-related safety, as with other APT trials in Europe [8], the shuttle bus part of our experiment was labeled *too slow*, showcasing the need for more mature technology [15, 54]. Due to these lower speeds, the passengers in our study had little reason to feel unsafe. Additionally, the human onboard operator often intervened if the bus got stuck behind parked cars or in roadworks as per Swedish laws.

In line with Eden et al. [25], our passengers suggested that the safety concerns will likely change if the vehicle is larger than a shuttle bus, operating at higher speeds on major roads and highways without an onboard operator. Therefore, for future work, we believe that there is a need to measure the effect of an agent on passenger perception of an unsafe environment, of an unsafe vehicle, or of unsafe co-passengers.

As reported in Section 4.3.4, passengers also suggested in the interviews that one of the key services that an agent could provide would be to call emergency services in unexpected or threatening conditions, especially in the absence of an onboard operator. No participants suggested that the agent could physically stop such problems on board, as our Furhat agent did not have arms or appendages to intervene physically. Some passengers also suggested that the embodied agent could be used for surveillance and calling the police. For the disembodied agent, passengers instead imagined safety functions in the bus itself, like emergency buttons.

Another safety concern identified in Eden et al. [25] was the need for strict usage of seat belts in APT shuttles due to frequent sudden stops when detecting any obstacles in the way. In line with this, our passengers also suggested that the agent should detect seat belt usage, potentially with colored seat belts for better visibility, as their experience in autonomous shuttles differed from regular buses where passengers often ignore seat belt signage.

As described in Section 4.3.2 and connecting strongly to RQ1, passenger suggestions often focused on the anthropomorphized [57, 76, 84] idea that the robot could *see* the bus—or the outside of the bus—and control the bus or environment in the bus based on what it could perceive. While it may not have been clear to all passengers, there were already surveillance cameras on the bus, and remote operators could see and hear the interior of the bus in the same way that our agent and teleoperator could. All the suggestions relating to surveillance and watching the bus are thus equally possible to implement for a disembodied agent as with an embodied agent—however, it seems quite clear that passengers do not necessarily *feel* surveilled without a physical agent in the bus, which could be an argument in favor of embodiment for addressing safety concerns.

Connecting to these concerns of surveillance, another crucial safety aspect highlighted by passengers was privacy. During the experiments, both the embodied and the disembodied agents asked each passenger for their destination. Everyone already on board could thus hear where the new passenger was going. This was seen as a potential invasion of privacy that could potentially lead to unwanted attention from fellow passengers. Therefore, it was expressed that both the embodied and disembodied agents should avoid sharing personal information out loud unless explicitly requested by the passenger. One potential solution to this would be to equip the agent with a screen for manual input when the information is private. In practice, an APT system would typically know where a passenger is going before the bus drives to pick them up, so passengers would not necessarily have to tell the agent where they are going. However, this concern extends to other topics an agent may want to talk to a passenger about.

We saw in Section 4.1.1 that passengers were more willing to talk to the embodied agent as if it were human. If the agent is capable of small talk and open conversation with the passengers, then a balance must be found between safety and privacy. If the agent can remember the names of its passengers, then referring to a passenger by name when they embark may be a privacy breach—but doing so may also discourage a passenger from harassing or bothering other passengers, since they would know that someone who knows who they are is onboard. Fosch Villaronga et al. have pointed to similar concerns as social robots are integrated in healthcare [30]. APT and healthcare share the property that a patient or passenger may not have chosen to be in the environment with the social robot—patients may not be able to choose where they are admitted, hospital staff may have to work in certain rooms where robots are active, and in our case, APT passengers may be forced to take a specific vehicle to work as they have no other way of commuting there on time.

Designers may not always have full control over the embodiment of a selected robotic agent. We used the Furhat robot, and thus choices like the presence of a projected face and a moving head, and the lack of arms or limbs, were made before we started showing our embodied agent to the users. On one hand, it can be useful to consider the embodiment of the agent to be locked and design the capabilities of the agent around user expectations from that embodiment—passengers can only provide realistic input about the uses they see for an agent if they are using that agent with concrete choices already made for embodiment. On the other hand, future designers of APT systems may have the freedom to design embodied agents as part of the vehicle or transit itself. It is therefore not a given if safety-based thinking, embodiment-based thinking or capability-based thinking should come first, although all three must be part of the design process.

5.4 Limitations

The original plan for the study was to put the robot on the bus on ordinary routes with real passengers. However, getting consent from all passengers for being filmed (for GDPR purposes) turned out to be impossible, especially since the ride was quite short. There was also a risk that people outside the bus would be filmed. In addition, it would have been hard to get them to fill out a questionnaire after the ride. Thus, we finally decided to use recruited participants, at the expense of ecological validity.

We had originally set up the experiment to have an unexplained event happen in the middle of the ride. Two events were balanced so that half of our participants experienced each event in their first run and the other event in their second run. The first event was labeled *no-show* and involved the bus stopping as if to pick up a fourth passenger, who would not show up. The second event was labeled *sudden stop* and was similar, except the doors would not open and Elin would remark that the bus had stopped for an unknown reason and would get going soon. Both events were intended to create an opportunity for passengers to reflect on the quality of the ride.

When running the experiment in practice, the conditions did not play out consistently. Sudden stops and no-shows happened naturally as part of the experimental environment, in ways that we could not control. Several groups of three passengers were run with two passengers instead when one participant could not show up, which led to genuine *no-show* events every time the bus stopped to pick up the missing passenger. The bus route software did not allow up to skip the stop for the participants who were not part of the experiment. Similarly, the bus often stopped suddenly for pedestrians, bikers, or other vehicles, which lessened the impact of our *sudden stop* event since the bus could stop suddenly at any point. Thus, we did not break down our analyses based on these conditions.

In future work, it is possible to plan for more extreme events, like actors pretending to be drunk or confederates running out in front of the AV to cause it to stop suddenly. However, it is important to consider the ethical implications of such interventions.

Similarly to the concern above, we could not try our experiment at night because the buses did not run at night. This meant that we could not explore how passenger safety was affected by a darker environment. The experiment also ran in early summer, when the sun sets late at night in Sweden, which would have complicated running the experiment in the dark. This could be compensated for by running the experiment in the winter, where the opposite holds.

Another limitation on our ability to explore feelings of lack of safety or insecurity was how slowly the buses drove. While the buses were on SAE automation level 4 [18], safety regulations meant that they had to drive slowly and had a human operator on board. The presence of the operator meant that passengers could trust that a human was on site if the bus was in an accident or major fault, which also limited how unsafe they could feel.

Some passengers asked Elin for instructions on how to get from one of our fictional stops to some other connection for public transit, like a nearby bus hub or the nearest commuter train station. We had not prepared answers for this kind of question in the script available to the Wizard, and the time needed to research the needed information and type out a travel plan manually was much longer than passengers were comfortable waiting. Elin thus was generally not capable of answering questions about connections. This could have been resolved by adding semi-generic answers to the script ("You can go to the nearest square and get a bus from there"), or by adding support for automated travel planning to the teleoperation interface. A limiting factor in automated queries was that our stops were fictional and did not appear on any transit maps, but this could be addressed in a future experiment by either running the experiment on real stops or with real passengers in actual public transit.

6 Conclusion

In this study, we examined the role of social robots (embodied agents) and voice assistants (disembodied agents) in APT at a test site in Barkarby, Stockholm. Data from passenger-agent interactions, post-questionnaires, and semi-structured focus group interviews were analyzed to answer the research questions we posed in Section 1:

RQ1: *What are the differences between a disembodied and an embodied agent serving as a host on a self-driving bus, in terms of how passengers perceive it and interact with it?* Our findings suggest that passengers interacted more frequently with the embodied agent due to its physical presence, viewing it as a person, while the disembodied voice was seen as part of the bus. The preferences of our passengers were split with some favoring the embodied agent, while others favoring the familiar disembodied agent for daily use. This suggests that the future designs of robotic assistants in APT should consider the embodiment to enhance interaction and perceived presence.

RQ2: *What services could an agent (embodied or not) provide in APT and what would passengers want to use it for?* Our findings suggest that passengers frequently requested information on locations, directions, ride clarifications, and technical questions, valuing detailed travel information and traffic explanations. In the absence of an agent, alternatives like smartphone apps or onboard kiosks could be explored to offer similar services.

RQ3: *Does the presence of an agent in APT address passengers' feelings of lack of safety or insecurity prompted by the absence of an authoritative human or smaller vehicles?* Our findings indicate the passengers felt generally safe but had concerns about unknown co-passengers at night, hence recognizing the need for the agent to contact emergency services without substituting physical interventions. We found that while agents can relieve some feelings of not feeling safe, these effects were hard to measure in our controlled environment, hence highlighting the need to study various scenarios that could pose a threat to passenger safety during the trip.

To conclude, integrating social robots and voice assistants into APT could improve passenger experience and safety. However, further research on passenger-agent interactions in varied scenarios is key to developing more effective and widely accepted solutions.

Acknowledgment

The authors thank Kjell Wanér for being the in-bus driver for most of the experimental runs, and also for helping solve many issues on the ground with the vehicles and depot.

References

[1] Ivana Acocella. 2012. The focus groups in social research: Advantages and disadvantages. *Quality and Quantity* 46 (2012), 1125–1136.

[2] Henny Admoni and Brian Scassellati. 2017. Social eye gaze in human-robot interaction: A review. *Journal of Human-Robot Interaction* 6, 1 (May 2017), 25–63. DOI: https://doi.org/10.5898/JHRI.6.1.Admoni

[3] Samer Al Moubayed, Jonas Beskow, Gabriel Skantze, and Björn Granström. 2012. Furhat: A back-projected human-like robot head for multiparty human-machine interaction. In *Cognitive Behavioural Systems*. Anna Esposito, Antonietta M. Esposito, Alessandro Vinciarelli, Rüdiger Hoffmann, and Vincent C. Müller (Eds.), Springer Berlin, Berlin, 114–130.

[4] Erik Almlöf. 2022. *Exploring Societal Impacts of Self-Driving Public Transport Using Four-Step Transport Models*. Ph. D. Dissertation. KTH Royal Institute of Technology.

[5] Shunsuke Aoki, Chung-Wei Lin, and Ragunathan Rajkumar. 2021. Human-Robot cooperation for autonomous vehicles and human drivers: Challenges and solutions. *IEEE Communications Magazine* 59, 8 (2021), 35–41. DOI: https://doi.org/10.1109/MCOM.001.2001241

[6] Agnes Axelsson, Bhavana Vaddadi, Cristian Bogdan, and Gabriel Skantze. 2024. Robots in autonomous buses: Who hosts when no human is there?. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24)*. ACM, New York, NY, 1278–1280. DOI: https://doi.org/10.1145/3610978.3641115

[7] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1, 1 (Jan. 2009), 71–81. DOI: https://doi.org/10.1007/s12369-008-0001-3

[8] Mauro Bellone, Azat Ismailogullari, Tommi Kantala, Sami Mäkinen, Ralf-Martin Soe, and Milla Åman Kyyrö. 2021. A cross-country comparison of user experience of public autonomous transport. *European Transport Research Review* 13, 1 (2021), 19.

[9] Israel Berger, Rowena Viney, and John P. Rae. 2016. Do continuing states of incipient talk exist? *Journal of Pragmatics* 91 (2016), 29–44. DOI: https://doi.org/10.1016/j.pragma.2015.10.009

[10] Oscar Bergqvist and Mikaela Åstrand. 2017. Bus Line Optimisation Using Autonomous Minibuses.

[11] Kari Blindheim, Mads Solberg, Ibrahim A. Hameed, and Rigmor Einang Alnes. 2023. Promoting activity in Long-Term care facilities with the social robot pepper: A pilot study. *Informatics for Health and Social Care* 48, 2 (2023), 181–195. DOI: https://doi.org/10.1080/17538157.2022.2086465

[12] Avram Block, Seonghee Lee, Aryaman Pandya, and Paul Schmitt. 2023. I see you! design factors for supporting pedestrian-AV interaction at crosswalks. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23)*. ACM, New York, NY, 364–368. DOI: https://doi.org/10.1145/3568294.3580107

[13] Annika Boos, Tatjana Schauer, Elisabeth Arndt, and Klaus Bengler. 2024. (Un-)persuasive robots: Exploring the effect of anthropomorphic cues on the foot-in-the-door effect across three experimental studies. *Computers in Human Behavior: Artificial Humans* 2, 1 (2024), 100061. DOI: https://doi.org/10.1016/j.chbah.2024.100061

[14] Glen Bramley and Karryn Kirk. 2005. Does planning make a difference to urban form? Recent evidence from Central Scotland. *Environment and Planning A: Economy and Space* 37, 2 (2005), 355–378.

[15] Jong Kyu Choi and Yong Gu Ji. 2015. Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human-Computer Interaction* 31, 10 (2015), 692–702.

[16] Herbert H. Clark and Kerstin Fischer. 2023. Social robots as depictions of social agents. *The Behavioral and Brain Sciences* 46 (2023), e21. DOI: https://doi.org/10.1017/S0140525X22000668

[17] Mark Colley, Julian Czymmeck, Mustafa Kücükkocak, Pascal Jansen, and Enrico Rukzio. 2024. PedSUMO: Simulacra of automated vehicle-pedestrian interaction using SUMO to study large-scale effects. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24)*. ACM, New York, NY, 890–895. DOI: https://doi.org/10.1145/3610977.3637478

[18] On-Road Automated Driving (ORAD) Committee. 2021. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. Standard. SAE International, Warrendale, PA. DOI: https://doi.org/10.4271/J3016_202104

[19] M. Core and J. Allen. 1997. Coding dialogs with the DAMSL annotation scheme. In *In Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines*. Association for the Advancement of Artificial Intelligence, Cambridge, 28–35.

[20] John W. Creswell and David J. Creswell. 2018. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed.). Sage Publications, Inc, Thousand Oaks, CA, 1–260.

[21] Eric Deng, Bilge Mutlu, and Maja J. Mataric. 2019. Embodiment in socially interactive robots. *Foundations and Trends® in Robotics* 7, 4 (2019), 251–356. DOI: https://doi.org/10.1561/2300000056

[22] Muhammet Deveci, Dragan Pamucar, Ilgin Gokasar, Bilal Bahaa Zaidan, Luis Martinez, and Witold Pedrycz. 2023. Assessing alternatives of including social robots in urban transport using fuzzy trigonometric operators based decision-making model. *Technological Forecasting and Social Change* 194 (2023), 122743.

[23] Claudia Di Napoli, Giovanni Ercolano, and Silvia Rossi. 2023. Personalized home-care support for the elderly: A field experience with a social robot at home. *User Modeling and User-Adapted Interaction* 33, 2 (Apr. 2023), 405–440. DOI: https://doi.org/10.1007/s11257-022-09333-y

[24] Xiaoxia Dong, Matthew DiScenna, and Erick Guerra. 2019. Transit user perceptions of driverless buses. *Transportation* 46 (2019), 35–50.

[25] Grace Eden, Benjamin Nanchen, Randolf Ramseyer, and Florian Evéquoz. 2017. Expectation and experience: Passenger acceptance of autonomous public transportation vehicles. In *Human-Computer Interaction (INTERACT '17)*. Regina Bernhaupt, Girish Dalvi, Anirudha Joshi, Devanuj K. Balkrishan, Jacki O'Neill, and Marco Winckler (Eds.), Springer International Publishing, Cham, 360–363.

[26] Chad Edwards, Autumn Edwards, and Varun Rijhwani. 2023. When in doubt, lay it out: Over vs. under-accommodation in human-robot interaction. *Language Sciences* 99 (2023), 101561. DOI: https://doi.org/10.1016/j.langsci.2023.101561

[27] Daniel J. Fagnant and Kara Kockelman. 2015. Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice* 77 (2015), 167–181.

[28] Kerstin Fischer, Katrin S. Lohan, and Kilian Foth. 2012. Levels of embodiment: Linguistic analyses of factors influencing HRI. In *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '12)*. ACM, New York, NY, 463–470. DOI: https://doi.org/10.1145/2157689.2157839

[29] Leopoldina Fortunati, Filippo Cavallo, and Mauro Sarrica. 2020. Multiple communication roles in human–robot interactions in public space. *International Journal of Social Robotics* 12, 4 (Aug. 2020), 931–944. DOI: https://doi.org/10.1007/s12369-018-0509-0

[30] Eduard Fosch Villaronga, Heike Felzmann, Robin Pierce, Silvia de Conca, Aviva de Groot, Scott Robins, and Aida Ponce Del Castillo. 2018. Nothing comes between my robot and me: Privacy and human-robot interaction in robotised healthcare. In *Data Protection and Privacy (1 ed.) (Computers, Privacy and Data Protection)*. Ronald Leenes, Rosamundevan Brakel, Serge Gutwirth, and Paul de Hert (Eds.), Hart Publishing, United Kingdom, 135–170.

[31] Emer Gilmartin, Benjamin R. Cowan, Carl Vogel, and Nick Campbell. 2018. Explorations in multiparty casual social talk and its relevance for social human machine dialogue. *Journal on Multimodal User Interfaces* 12, 4 (Dec. 2018), 297–308. DOI: https://doi.org/10.1007/s12193-018-0274-2

[32] Carina Goldbach, Jörn Sickmann, Thomas Pitz, and Tatjana Zimasa. 2022. Towards autonomous public transportation: Attitudes and intentions of the local population. *Transportation Research Interdisciplinary Perspectives* 13 (2022), 100504.

[33] O. Can Görür, Benjamin Rosman, Fikret Sivrikaya, and Sahin Albayrak. 2018. Social cobots: Anticipatory decision-making for collaborative robots incorporating unexpected human behaviors. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. ACM, New York, NY, 398–406. DOI: https://doi.org/10.1145/3171221.3171256

[34] Kerstin S. Haring, Kelly M. Satterfield, Chad C. Tossell, Ewart J. de Visser, Joseph R. Lyons, Vincent F. Mancuso, Victor S. Finomore, and Gregory J. Funke. 2021. Robot authority in human-robot teaming: Effects of human-likeness and physical embodiment on compliance. *Frontiers in Psychology* 12 (2021), 1–14. DOI: https://doi.org/10.3389/fpsyg.2021.625713

[35] Monique M. Hennink. 2007. *International Focus Group Research: A Handbook for the Health and Social Sciences*. Cambridge University Press.

[36] Bernd Herrenkind, Ilja Nastjuk, Alfred Benedikt Brendel, Simon Trang, and Lutz M. Kolbe. 2019. Young people's travel behavior–using the life-oriented approach to understand the acceptance of autonomous driving. *Transportation Research Part D: Transport and Environment* 74 (2019), 214–233.

[37] Hsiu-Fang Hsieh and Sarah E. Shannon. 2005. Three approaches to qualitative content analysis. *Qualitative Health Research* 15, 9 (2005), 1277–1288. DOI: https://doi.org/10.1177/1049732305276687

[38] Calin Iclodean, Nicolae Cordos, and Bogdan Ovidiu Varga. 2020. Autonomous shuttle bus for public transportation: A review. *Energies* 13, 11 (2020), 2917.

[39] Ryan Blake Jackson and Tom Williams. 2021. A theory of social agency for human-robot interaction. *Frontiers in Robotics and AI* 8 (2021), 1–13. DOI: https://doi.org/10.3389/frobt.2021.687726

[40] Iris Jestin, Joel Fischer, Maria Jose Galvez Trigo, David Large, and Gary Burnett. 2022. Effects of wording and gendered voices on acceptability of voice assistants in future autonomous vehicles. In *Proceedings of the 4th Conference on Conversational User Interfaces (CUI '22)*. ACM, New York, NY, Article 24, 1–11. DOI: https://doi.org/10.1145/3543829.3543836

[41] Eva Kassens-Noor, Zeenat Kotval-Karamchandani, and Meng Cai. 2020. Willingness to ride and perceptions of autonomous public transit. *Transportation Research Part A: Policy and Practice* 138 (2020), 92–104.

[42] Daniel Kaszas and Adam Charles Roberts. 2023. Comfort with varying levels of human supervision in self-driving cars: Determining factors in Europe. *International Journal of Transportation Science and Technology* 12, 3 (2023), 809–821.

[43] Katharina Keller, Kim Valerie Carl, Hendrik Jöntgen, Benjamin M. Abdel-Karim, Max Mühlhäuser, and Oliver Hinz. 2019. "K.I.T.T, where are you?": Why smart assistance systems in cars enrich people's lives. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct)*. ACM, New York, NY, 1120–1132. DOI: https://doi.org/10.1145/3341162.3350844

[44] Hyunjung Kim, Max Fischer, Jongik Jeon, Seunghwa Pyo, Kentaro Honma, Miles Pennington, Hailong Liu, Shimono Keisuke, and Hiraoka Toshihiro. 2024. Design exploration of robotic in-car accessories for semi-autonomous vehicles. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24)*. ACM, New York, NY, 603–607. DOI: https://doi.org/10.1145/3610978.3640596

[45] Richard A. Krueger. 2014. *Focus Groups: A Practical Guide for Applied Research*. Sage Publications.

[46] Divesh Lala, Koji Inoue, and Tatsuya Kawahara. 2019. Smooth turn-taking by a robot using an online continuous model to generate turn-taking cues. In *2019 International Conference on Multimodal Interaction (ICMI '19)*. ACM, New York, NY, 226–234. DOI: https://doi.org/10.1145/3340555.3353727

[47] Hui Hui Lim, Oliver F. Shyr, and Tzu-Yuan Stessa Chao. 2017. Ageing and mobility: Towards age-friendly public transport in Taiwan. In *Planning for Greying Cities*. Routledge, England, UK, 89–109.

[48] M. Eugenia López-Lambas and Andrea Alonso. 2019. The driverless bus: An analysis of public perceptions and acceptability. *Sustainability* 11, 18 (2019), 4986.

[49] Giuseppe Lugano. 2017. Virtual assistants and self-driving cars. In *2017 15th International Conference on ITS Telecommunications (ITST)*. IEEE, New York, 1–5. DOI: https://doi.org/10.1109/ITST.2017.7972192

[50] Michal Luria, Samantha Reig, Xiang Zhi Tan, Aaron Steinfeld, Jodi Forlizzi, and John Zimmerman. 2019. Re-embodiment and co-embodiment: Exploration of social presence for robots and conversational agents. In *Proceedings of the 2019 on Designing Interactive Systems Conference (DIS '19)*. ACM, New York, NY, 633–644. DOI: https://doi.org/10.1145/3322276.3322340

[51] Anna-Maria Meck, Marion Sardone, and Jacqueline Cullmann. 2023. Will the assistant become the driver, and the driver become the assistant? In *Proceedings of the 5th International Conference on Conversational User Interfaces (CUI '23)*. ACM, New York, NY, Article 30, 1–5. DOI: https://doi.org/10.1145/3571884.3603753

[52] Blanca Miller and David Feil-Seifer. 2016. Embodiment, situatedness, and morphology for humanoid robots interacting with people. In *Humanoid Robotics: A Reference*. Ambarish Goswami and Prahlad Vadakkepat (Eds.), Springer Netherlands, Dordrecht, 1–23. DOI: https://doi.org/10.1007/978-94-007-7194-9_130-1

[53] Dylan Moore, Rebecca Currano, Michael Shanks, and David Sirkin. 2020. Defense against the dark cars: Design principles for griefing of autonomous vehicles. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20)*. ACM, New York, NY, 201–209. DOI: https://doi.org/10.1145/3319502.3374796

[54] Kostas Mouratidis and Victoria Cobeña Serrano. 2021. Autonomous buses: Intentions to use, passenger experiences, and suggestions for improvement. *Transportation Research Part F: Traffic Psychology and Behaviour* 76 (2021), 321–335.

[55] Sina Nordhoff, Joost De Winter, Miltos Kyriakidis, Bart Van Arem, and Riender Happee. 2018. Acceptance of driverless vehicles: Results from a large cross-national questionnaire study. *Journal of Advanced Transportation* 2018 (2018), 1–22.

[56] Marco Nørskov. 2017. *Social Robots: Boundaries, Potential, Challenges*. Taylor and Francis, New York.

[57] Linda Onnasch and Clara Laudine Hildebrandt. 2021. Impact of anthropomorphic robot design on trust and attention in industrial human-robot interaction. *Journal of Human-Robot Interaction* 11, 1 (Oct. 2021), Article 2, 1–24. DOI: https://doi.org/10.1145/3472224

[58] OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, et al. 2024. GPT-4 technical report. arXiv:2303.08774. Retrieved from https://arxiv.org/abs/2303.08774

[59] Christina Pakusch and Paul Bossauer. 2017. User acceptance of fully autonomous public transport. In *ICE-B*. Scitepress, Portugal, 52–60.

[60] Marco Pavanini. 2024. Postphenomenology and human constitutive technicity: How advances in AI challenge our self-understanding. *Journal of Human-Technology Relations* 2, 1 (Dec. 2024), 1–25. DOI: https://doi.org/10.59490/jhtr.2024.2.7377

[61] Hannah R. M. Pelikan and Malte F. Jung. 2023. Designing robot sound-in-interaction: The case of autonomous public transport shuttle buses. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23)*. ACM, New York, NY, 172–182. DOI: https://doi.org/10.1145/3568162.3576979

[62] Elena Popa. 2021. Human goals are constitutive of agency in artificial intelligence (AI). *Philosophy and Technology* 34, 4 (Dec. 2021), 1731–1750. DOI: https://doi.org/10.1007/s13347-021-00483-2

[63] John W. Pratt. 1959. Remarks on zeros and ties in the Wilcoxon signed rank procedures. *Journal of the American Statistical Association* 54, 287 (1959), 655–667. Retrieved from http://www.jstor.org/stable/2282543

[64] Eileen Roesler, Dietrich Manzey, and Linda Onnasch. 2023. Embodiment matters in social HRI research: Effectiveness of anthropomorphism on subjective and objective outcomes. *Journal of Human-Robot Interaction* 12, 1 (Feb. 2023), Article 7, 1–9. DOI: https://doi.org/10.1145/3555812

[65] Emanuel A. Schegloff and Harvey Sacks. 1973. Opening up closings. *Semiotica* 8, 4 (1973), 289–327. DOI: https://doi.org/10.1515/semi.1973.8.4.289

[66] Andreas Schrank, Carmen Kettwich, and Michael Oehl. 2024. Aiding automated shuttles with their driving tasks as an on-board operator: A case study on different automated driving systems in three living labs. *Applied Sciences* 14, 8 (2024), 3336.

[67] Gabriel Skantze. 2021. Turn-taking in conversational systems and human-robot interaction: A review. *Computer Speech and Language* 67 (2021), 101178. DOI: https://doi.org/10.1016/j.csl.2020.101178

[68] Randall Spain, Wookhee Min, Vikram Kumaran, Jay Pande, Jason Saville, and James Lester. 2025. Applying large language models to enhance dialogue and communication analysis for adaptive team training. *International Journal of Artificial Intelligence in Education* (May 2025). DOI: https://doi.org/10.1007/s40593-025-00479-5

[69] Carolin Straßmann, Alina Grewe, Christopher Kowalczyk, Alexander Arntz, and Sabrina C. Eimler. 2020. Moral robots? How uncertainty and presence affect humans' moral decision making. In *HCI International 2020—Posters*. Constantine Stephanidis and Margherita Antona (Eds.), Springer International Publishing, Cham, 488–495.

[70] Mikael Svenstrup, Thomas Bak, Ouri Maler, Hans Jørgen Andersen, and Ole B. Jensen. 2009. Pilot study of person robot interaction in a public transit space. In *Research and Education in Robotics—EUROBOT 2008*. Achim Gottscheber, Stefan Enderle, and David Obdrzalek (Eds.), Springer Berlin, Berlin, 96–106.

[71] Renata Tesch. 2013. *Qualitative Research: Analysis Types and Software*. Routledge, England, UK.

[72] Sam Thellman, Annika Silvervarg, Agneta Gulz, and Tom Ziemke. 2016. Physical vs. virtual agent embodiment and effects on social interaction. In *Intelligent Virtual Agents*. David Traum, William Swartout, Peter Khooshabeh, Stefan Kopp, Stefan Scherer, and Anton Leuski (Eds.), Springer International Publishing, Cham, 412–415.

[73] Ying Tian, Siyun Liu, and Jianying Wang. 2024. A corpus study on the difference of turn-taking in online audio, online video, and face-to-face conversation. *Language and Speech* 67, 3 (2024), 593–616. DOI: https://doi.org/10.1177/00238309231176768

[74] Dejan Todorović. 2006. Geometrical basis of perception of gaze direction. *Vision Research* 46, 21 (2006), 3549–3562. DOI: https://doi.org/10.1016/j.visres.2006.04.011

[75] Marynel Vázquez, Yofti Milkessa, Michelle M. Li, and Neha Govil. 2020. Gaze by semi-virtual robotic heads: Effects of eye and head motion. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, New York, 11065–11071. DOI: https://doi.org/10.1109/IROS45743.2020.9341209

[76] Joshua Wainer, David J. Feil-Seifer, Dylan A. Shell, and Maja J. Mataric. 2006. The role of physical embodiment in human-robot interaction. In *ROMAN 2006—The 15th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, New York, 117–122. DOI: https://doi.org/10.1109/ROMAN.2006.314404

[77] Induni N. Weerarathna, David Raymond, and Anurag Luharia. 2023. Human-robot collaboration for healthcare: A narrative review. *Cureus* 15, 11 (Nov. 2023), e49210.

[78] Eva Wiese, Patrick P. Weis, and Daniel M. Lofaro. 2018. Embodied social robots trigger gaze following in real-time HRI. In *2018 15th International Conference on Ubiquitous Robots (UR)*. IEEE, New York, 477–482. DOI: https://doi.org/10.1109/URAI.2018.8441825

[79] Frank Wilcoxon. 1945. Individual comparisons by ranking methods. *Biometrics Bulletin* 1, 6 (1945), 80–83.

[80] Haibin Yan, Marcelo H. Ang, and Aun Neow Poo. 2014. A survey on perception methods for human-robot interaction in social robots. *International Journal of Social Robotics* 6 (2014), 85–119.

[81] Kum Fai Yuen, Yiik Diew Wong, Fei Ma, and Xueqin Wang. 2020. The determinants of public acceptance of autonomous vehicles: An innovation diffusion perspective. *Journal of Cleaner Production* 270 (2020), 121904.

[82] Yanxia Zhang, Jonas Beskow, and Hedvig Kjellström. 2017. Look but don't stare: Mutual gaze interaction in social robots. In *Social Robotics*. Abderrahmane Kheddar, Eiichi Yoshida, Shuzhi Sam Ge, Kenji Suzuki, John-John Cabibihan, Friederike Eyssel, and Hongsheng He (Eds.), Springer International Publishing, Cham, 556–566.

[83] Yan Zhang and Barbara M. Wildemuth. 2009. Qualitative analysis of content. *Applications of Social Research Methods to Questions in Information and Library Science* 308, 319 (2009), 1–12.

[84] Jakub Złotowski, Diane Proudfoot, Kumar Yogeeswaran, and Christoph Bartneck. 2015. Anthropomorphism: Opportunities and challenges in human–robot interaction. *International Journal of Social Robotics* 7, 3 (Jun. 2015), 347–360. DOI: https://doi.org/10.1007/s12369-014-0267-6

Appendices

A Questionnaires

Table A1 shows the questions that were given to participants after the first run (leftmost column) and the second run (middle column). For those questions that did not change between the two runs, the cells are merged. The questionnaires were originally presented to participants in Swedish and have been translated to English for convenience here. The original documents exactly as they were given to participants can be found in the Supplementary Materials.

Table A1. The Questions from the Two Post-Run Questionnaires as Described in Section 3.2

Question in First Run	Question in Second Run	Response Type
What was the name of the stop where you were supposed to get off?		*Free text*
Did you manage to get off there?		*Yes, no*
Did something unexpected happen during your ride, i.e., apart from the bus letting on and dropping off passengers? If nothing happened, you may leave this empty.		*Free text*
Elin helped me understand why the bus had stopped when it stopped for technical reasons.		*1 (disagree completely)–7 (agree completely)*
When Elin talked to a passenger, the turn-taking (who started and stopped talking at what time) worked well.		*1 (disagree completely)–7 (agree completely)*
It was easy to understand what Elin said.		*1 (disagree completely)–7 (agree completely)*
Elin understood us passengers when we spoke.		*1 (disagree completely)–7 (agree completely)*
Elin gave reasonable answers to questions.		*1 (disagree completely)–7 (agree completely)*
Elin talked too much.		*1 (disagree completely)–7 (agree completely)*
	I felt that Elin was...	
Fake/Natural	More Fake/Natural than run 1?	*1 (left option)–7 (right option)*
Machine-Like/Human-Like	More Machine-Like/Human-Like than run 1?	*1 (left option)–7 (right option)*
Unconscious/Conscious	More Unconscious/Conscious than run 1?	*1 (left option)–7 (right option)*
Artificial/Lifelike	More Artificial/Lifelike than run 1?	*1 (left option)–7 (right option)*
	I felt that Elin was...	
Dead/Alive	More Dead/Alive than run 1?	*1 (left option)–7 (right option)*
Stagnant/Lively	More Stagnant/Lively than run 1?	*1 (left option)–7 (right option)*
Mechanical/Organic	More Mechanical/Organic than run 1?	*1 (left option)–7 (right option)*
Inert/Interactive	More Inert/Interactive than run 1?	*1 (left option)–7 (right option)*
Apathetic/Responsive	More Apathetic/Responsive than run 1?	*1 (left option)–7 (right option)*
	My opinion on Elin is ... (or) I felt that Elin was ...	
Dislike/Like	Dislike/Like more than run 1?	*1 (left option)–7 (right option)*
Unfriendly/Friendly	More Unfriendly/Friendly than run 1?	*1 (left option)–7 (right option)*
Unkind/Kind	More Unkind/Kind than run 1?	*1 (left option)–7 (right option)*
Unpleasant/Pleasant	More Unpleasant/Pleasant than run 1?	*1 (left option)–7 (right option)*
Awful/Nice	More Awful/Nice than run 1?	*1 (left option)–7 (right option)*
	I felt that Elin was ...	
Incompetent/Competent	More Incompetent/Competent than run 1?	*1 (left option)–7 (right option)*
Ignorant/Knowledgeable	More Ignorant/Knowledgeable than run 1?	*1 (left option)–7 (right option)*
Irresponsible/Responsible	More Irresponsible/Responsible than run 1?	*1 (left option) - 7 (right option)*
Unintelligent/Intelligent	More Unintelligent/Intelligent than run 1?	*1 (left option)–7 (right option)*
Foolish/Sensible	More Foolish/Sensible than run 1?	*1 (left option)–7 (right option)*
	When I interacted with Elin, I felt ...	
Anxious/Relaxed	More Anxious/Relaxed than run 1?	*1 (left option)–7 (right option)*
Calm/Agitated	More Calm/Agitated than run 1?	*1 (left option)–7 (right option)*
Still/Surprised	More Still/Surprised than run 1?	*1 (left option)–7 (right option)*

The questionnaire after the first run differed from that after the second run by whether the Godspeed questions were asked as absolute measures or as a comparison; this difference is shown with the split between the left and the middle column. The sub-questionnaires within Godspeed were broken up by sentences contextualizing the options on the left and right side of the scales; these contextualizing sentences are shown as italicized full-width lines.

B Semi-Structured Interview Questions

Below, the questions that were used for the semi-structured interview after the two rides are shown.

- Regarding the trip
  - What were your expectations with this trip? How do you feel now?
  - How do you feel about interacting with the robot, what was your experience like?
  - How do you feel riding with the autonomous bus, what was your experience like?
  - What did you do in the bus? (activities)
  - How was it to use the robot? (Did you get enough information about the ride and your stops?)
  - What do you think of the design of the bus?
  - Please describe how you used the service (bus and robot) What do you think about the bus driving itself? (Comfortable, uneasy, quick, etc.)
  - How do you think you would have felt if there was no driver onboard?
- Safety and security
  - How did you experience the safety aspect during this trip? (did you feel safe?)
  - What aspects of this trip made you feel safe? What aspects of this trip made you feel unsafe?
  - What are your thoughts regarding the virtual bus assistant and the Bus? Can it be trusted?
- General experiences
  - How was this experience differ from a normal bus ride?
  - What do you like about sharing mobility, what do you dislike about it?
  - What do these technologies (bus and robot for shared mobility) mean to you?
- Final thoughts
  - In your opinion, if these combination of innovations were available at a large scale in the city of Stockholm, how do you think this will affect people's everyday travel? For what purpose will you use this bus if it was available in the city everywhere?
  - Finally, were your expectations met? What was missing for you? What can be done for this service which will allow you to use this service regularly? Anything you'd like to add?

C LLM Prompt Used for Dialogue Act Classification

Table C1 shows the LLM prompt used for dialogue act classification, as described in Section 3.6.

Table C1. Chat Completion LLM Prompt Used for Dialogue Act Classification with GPT-4

Role	Message
System	I will provide you with a dialog between A, B, C, E och O onboard a self-driving bus. A, B, C are passengers. E (Elin) är the host. O is other. For every row in the dialog, I want you to provide an intent classification in English. Just answer with the classification in capital letters, no other text. You can choose between the following options: * GREETING: Instances where the characters are initiating a conversation or acknowledging each other's presence, like "Hej" (Hello), "Goddag" (Good day). * GIVE-INFORMATION: These are instances where a character provides details or answers, such as Elin (E) introducing herself as a bus assistant, providing information about bus stops, or informing about the next stop. * REQUEST-INFORMATION: Occasions where a character is seeking information from another. This includes Elin (E) asking passengers about their destinations. * CONFIRMATION: Short responses that acknowledge or confirm what someone else has said, for example, "Okej" (Okay), "Ja" (Yes), "Visst" (Sure). * THANKS: Expressions of gratitude, such as "Tack så mycket" (Thank you very much). * INSTRUCTION: When Elin (E) gives directions or instructions to the passengers, like advising them to sit down and buckle up. * FEEDBACK: Elin (E) asking passengers about what they think a robot like her should be able to do in a bus, or passengers giving such feedback. * FAREWELL: Ending conversations, such as "Hejdå" (Goodbye). * SMALL-TALK: Casual or off-topic conversations between characters, such as the discussion between passengers A and B about the bus feeling like a car or their plans. * APOLOGY: Instances where a character expresses regret for an inconvenience or mistake. * CLARIFICATION: When characters seek to clear up confusion or misunderstandings. * OTHER: Something that does no† fit any of the above.
User	E till A: Jag är Elin, en bussassistent
Assistant	GIVE-INFORMATION
User	A till E: Okej
Assistant	CONFIRMATION
User	E till A: Kan du säga var din destination är?
Assistant	REQUEST-INFORMATION
User	A till E: Jordgubbsvägen
Assistant	GIVE-INFORMATION

Throughout the same dialogue, user messages are appended with the next line in the dialogue. However, only the last 10 user/assistant messages are sent to reduce the length of the prompt. †: *(sic)*.

D Test Statistics for Wilcoxon Signed-Rank Tests from Section 4.2.3

Table D1 shows the test statistics W for the Wilcoxon signed-rank tests run to compare the questionnaire answers given after the runs. The test statistic for animacy falls between the critical values for $p = 0.05$ and $p = 0.02$ and is thus not significant when we Bonferroni-correct for the 11 comparisons done at the same time here. Thus, as mentioned in Section 4.2.3, no comparison is significant.

Table D1. Results of Wilcoxon Signed-Rank Tests for Questionnaire Answers Comparing the Two Runs of Our Experiment

Question	Statistic (W)	Sample Size (n)	Critical Values 5%	2%
Embodied Elin helped me understand why the bus had stopped when it stopped for technical reasons more than *disembodied Elin*.	378	19	46	37
When Elin talked to a passenger, the turn-taking (who started and stopped talking at what time) worked better with *embodied Elin* than with *disembodied Elin*.	521	27	107	92
It was easier to understand what *embodied Elin* said than *disembodied Elin*.	162	9	5	3
Embodied Elin understood us passengers when we spoke better than *disembodied Elin*.	528.5	30	137	120
Embodied Elin gave more reasonable answers to questions than *disembodied Elin*.	532.5	29	126	110
Embodied Elin talked too much to a greater extent than *disembodied Elin*.	240	28	116	101
How much more **anthropomorphism** did *embodied Elin* have than *disembodied Elin*?	683.5	52	434	397
How much more **animacy** did *embodied Elin* have than *disembodied Elin*?	398.5	52	434	397
How much more **likeability** did *embodied Elin* have than *disembodied Elin*?	674	53	434	397
How much more **perceived intelligence** did *embodied Elin* have than *disembodied Elin*?	327	42	294	266
How much more **perceived safety** did *embodied Elin* have than *disembodied Elin*?	443	48	396	362

In the Wilcoxon signed-rank test, the ranks of absolute values from the two classes are summarized, and the smaller of the two sums is the test statistic. Critical values are pre-calculated from the significance limit α and the sample size n, and the test statistic must be *smaller* than the critical value to count as statistically significant by that α [79]. To resolve the presence of ties, we followed the practice set up by Pratt [63], see Section 4.2.3.

Received 24 July 2024; revised 28 May 2025; accepted 25 July 2025