Towards Remote Expert Supported Autonomous Assistant Robots in Shopping Environments

Niko Kleer
DFKI
Saarland Informatics Campus
Saarbrücken, Germany
niko.kleer@dfki.de

Martin Feick
DFKI
Saarland Informatics Campus
Saarbrücken, Germany
martin.feick@dfki.de

Florian Daiber
DFKI
Saarland Informatics Campus
Saarbrücken, Germany
florian.daiber@dfki.de

Michael Feld
DFKI
Saarland Informatics Campus
Saarbrücken, Germany
michael.feld@dfki.de

ABSTRACT

Autonomous robots that assist customers during their shopping trips may become ubiquitous in the near future. However, these systems are limited in their ability to appropriately reply to all of a customer's requests. Therefore, a dedicated human expert has to remotely "take over" and resolve the customer's issue to warrant a satisfying experience. As a result, a Transfer of Control from a machine to a human takes place. This creates the problem that the expert needs to quickly review what has happened prior to the takeover to provide optimal assistance. To address this, we designed and implemented an interactive summary concept allowing experts to quickly filter and review the most relevant information to solve the given issue. Finally, we provide an example use case illustrating how a remote expert may solve a request using our system.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; Interface design prototyping; • **Computer systems organization** → *Robotics*.

KEYWORDS

Shopping Assistant Robots; Human-Robot Interaction; Transfer of Control; Situation Summarization; Interface Prototyping

ACM Reference Format:
Niko Kleer, Martin Feick, Florian Daiber, and Michael Feld. 2024. Towards Remote Expert Supported Autonomous Assistant Robots in Shopping Environments. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24 Companion), March 11–14, 2024, Boulder, CO, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3610978.3640735

HRI '24 Companion, March 11–14, 2024, Boulder, CO, USA

ACM ISBN 979-8-4007-0323-2/24/03…$15.00
https://doi.org/10.1145/3610978.3640735

1 INTRODUCTION

Technological advances and the continuous integration of modern technology into our environment keep increasing the demand for professional service robots, as reported by the International Federation of Robotics (IFR) [15]. Shopping environments where such robots are intended to accompany customers during their shopping trip represent a prominent use case that has been explored extensively [6, 12, 25, 26]. However, enabling a robot to engage in a task-oriented conversation with a customer represents a challenging task. Imagine the following scenario: a robot accompanies a customer during a shopping trip. During the trip, the robot is capable of answering product-related questions such as providing their location in the store or the total number of units still in stock. However, there will always exist situations where the autonomous shopping assistant is unable to respond appropriately due to the complexity of the request or simply because the customer asks for a human's opinion because s/he does not wish to be consulted by a robotic system [6]. Whenever a robot is unable to cope with a customer's request, an expert has to take over and resolve the issue, leading to a *Transfer of Control* (ToC) from a machine to a human. If a local expert is nearby, the customer may directly call them. However, it is more likely that a dedicated remote expert monitors all shopping robots, and in case of a ToC, takes over to resolve the issue. This might result in the customer having to repeat their issue and possibly walk the expert through a considerable amount of contextual information related to the interaction with the robot. Consequently, the customer and the expert both lose valuable time. Even though these limitations are discussed in the literature [6, 10, 18], we are unaware of research targeted to resolve such situations. To resolve a customer's issue as quickly as possible, a summary comprising the most significant information about the issue would be useful. This raises numerous questions, including what information could be gathered and how it should be summarized. In particular, summarizing the human-robot conversation requires more sophisticated text summarization techniques.

This paper proposes a conceptual framework for situation summarization to enable a remote expert to resolve a customer's issue. To this end, we have implemented a prototypical system with an emphasis on the summarization of textual information gathered as a part of the human-robot conversation during the shopping trip.

2 RELATED WORK

The literature related to this work can be divided in two fields: robots assisting humans in shopping environments and text summarization techniques.

2.1 Robots in Shopping Environments

As the integration of autonomous robots in shopping environments represents a highly complex task, several field studies have previously been conducted. One of the most prominent examples of such a long-term field study was published by Gross et al. [12] who introduced TOOMAS, an interactive mobile shopping assistant [13]. Their robot was aimed at guiding customers through a large home store environment. The authors later evaluated their system from a user-centered perspective as well [7]. In another study carried out in a shopping mall by Kanda et al. [17, 18], a humanoid robot guided customers in an information providing task. Due to the size and complex layout of shopping malls, robots appear useful for helping visitors navigate through the mall [6]. The system was further aimed at building rapport in order to develop a relationship. The latter represents an important aspect with regard to whether a robot is perceived as a tool rather than a partner [3, 16, 26]. In another comprehensive field trial by Chen et al. [6], the shopping assistant robot KeJia was presented and evaluated. Their robot supported multimodal interaction using speech and a mobile application. Finally, a study regarding the acceptance of robots in shopping malls was presented by Niemelä et al. [22–25]. While the authors report on a generally positive sentiment towards the technology, there also exist many expectations from store owners for leveraging robots from a business perspective.

Based on the literature in this field, the authors seem to agree that human guidance in maze-like or crowded environments and the recognition of speech still represent a major challenge which is why many robots are teleoperated to an extent [6, 12, 17, 18]. Even though they acknowledge that robots might not be able to be fully autonomous in the near future, we are unaware of any published research targeted towards resolving a customer's issue. This is particularly the case when robots become even more sophisticated and are capable of engaging in longer conversations.

2.2 Text Summarization Techniques

Although this work aims to summarize a complete interaction between a customer and a robot, meaning it could be considered a multimodal summary, text summarization represents a particularly complex aspect. The automatic summarization of textual information has long been a field of interest. Due to the diversity of approaches and the continuous growth of the field, comprehensive surveys have been published [2, 8, 14, 21, 27]. General surveys commonly distinguish between extractive and abstractive summaries. Extractive summaries filter a collection of documents by determining sentences most beneficial for describing the entire collection. Abstractive approaches make an effort to develop an understanding of what a collection of documents is about. In both cases, the summary represents a considerably shorter version of the original text that conveys the most significant information. However, most text summarization techniques do not have to deal with challenges such as the temporal and spatial meaning of textual data as introduced during a Human-Robot Interaction (HRI) in a shopping environment. In case of an incident that is related to an entity such as a product from the store, keyword extraction or keyword-based text summarization can be considered relevant as well [4].

3 REQUIREMENTS AND CONCEPTUAL INTERFACE DESIGN

As this work is concerned with situations where a shopping assistance robot reaches its limitations during a human-robot interactive scenario, it is necessary to consider the requirements that allow a remote expert to resolve such an incident. These requirements relate to the data that is needed for quickly developing an understanding of the incident's cause. As this data should only consider crucial parts of the entire interaction between a customer and the robot, an intelligent summary has to be generated based on this interaction. In the next three sections, we will discuss the data that may enable a remote expert to understand a customer's issue, our conversation summarization approach, and a conceptual design for the interface operated by a remote expert.

3.1 Transfer of Control Parameters

It seems evident from the related work that there exists temporal and spatial data that may help a remote expert to resolve a problem [6, 17, 18, 20]. Although the data periodically collected by a robot differs according to the robot's capabilities as well as the application domain, the following data appears useful in general.

- The **conversation** between the customer and the robot.
- The **location** of the robot in the store while accompanying the customer during their shopping trip.
- A **timestamp** for inferring temporal relations by, for example, enabling linking location-based data to the transcribed sentences of the human-robot conversation.
- Finally, **products** that appeared as a part of the conversation or currently part of the customer's shopping cart. The reason for this is that these products carry the potential to lead to a product-related question a robot cannot answer.

Based on this data, the main goal lies in **generating an intelligent summary** that enables a remote expert to resolve a customer's issue as smoothly as possible, in case the robot reaches its limitations. Pre-processing and visualizing the data in such a manner so that the remote expert can make sense of the situation represents one of the main challenges in achieving this goal.

3.2 Conversation Summarization Approach

Filtering and summarizing conversation-based data for presenting the most substantial parts to a remote expert for resolving an issue represents a challenging task. Especially when shopping assistant robots incorporate dialog systems [6] or knowledge about the environment [17, 18]. In practice, this might imply that transferring the control to a remote expert must follow certain constraints. This is because asking a shopping assistant robot a completely unrelated question such as "what is the diameter of planet earth?" should not lead to an unnecessary ToC. Instead, enforcing constraints, such as **product-related questions may lead to transferring the control** to a remote expert, seems sensible. This section focuses

exactly on this constraint, generating a textual summary based on product-related questions that could not be answered by the robot.

Summarization approaches generally aim to turn a collection of documents $C = \{d_1, d_2, ..., d_n\}$ into a much smaller summarized textual description [2, 8, 14, 21, 27]. In our case, each spoken sentence can be considered a document. A substantial difference is, however, that there exists one document that causes the need for a ToC. Moreover, since all potentially relevant products are known by the robot, the products that caused the issue can be determined. The products can further be utilized to apply a filter to the documents, resulting in a subset of possibly relevant documents. At this point, this collection might still encompass a large number of documents. To simplify understanding the cause of ToC, these documents also need to be structured. The approach we propose is inspired by clustering-based summarization methods as this allows us to summarize information and recognize reoccurring patterns as well [8]. First, the approach relies on the well-established term frequency-inverse document frequency (TF-IDF) vector space model. After that, we apply the cosine-similarity measure

$$cos(\theta) = \frac{A \cdot B}{||A|| \cdot ||B||} = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \cdot \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

pairwise over all the TF-IDF vectors we retrieve. By comparing all vectors, we measure the relevance of our documents in relation to the cause of the ToC and are able to find temporal relations as well. Nevertheless, this representation of our documents cannot be presented in a useful manner. It is necessary to structure related documents and to determine their relevance to the cause of the ToC. To this end, we utilize the power of clustering approaches. Specifically, the matrix that results from the aforementioned pairwise comparisons functions as a distance matrix when applying a Hierarchical Agglomerative Clustering (HAC) approach. As a result, the documents are structured according to high cosine-similarities (i.e. they overlap in significant or contextual information). There exist two major challenges in achieving a sensible clustering. First, **conversation pre-processing** should ensure that the clustering method allows structuring the data appropriately. To this end, resolving references to mentioned products might be significant. Second, determining the **optimal number of clusters** represents a highly challenging task as the information that a remote expert gains based on a specific number of clusters might depend on an individual expert. After overcoming these challenges, integrating the textual summary into the overall summary is necessary and all information has to be visualized for the remote expert.

3.3 Conceptual Remote Expert Interface Design

Providing a visual representation of the intelligent summary represents a key aspect in enabling a remote expert to resolve a customer's issue. In the literature, we were able to find one example of such a user interface [10]. Even though the authors incorporate their interface for supervising and teleoperating a network of robots, elements such as a map or buttons for controlling a robot seem useful in our use case as well. Based on the ToC parameters discussed in Section 3.1, we designed a conceptual remote expert user interface (see Figure 1) that comprises the following panels and elements:

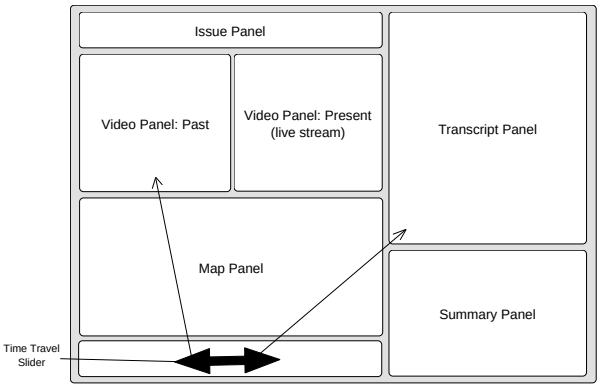

Figure 1: Conceptual remote expert interface design.

- **Issue Panel**: Provides general information about the cause of the ToC.
- **Video Panel**: Divided into a past- and live video panel. The video from the past enables the remote expert to revisit interactions between the customer and the robot. On the other hand, the live video allows the remote expert to directly interact with the customer.
- **Map Panel**: Similar to the interface presented by [10], we consider a map to be a substantial element. The map is supposed to visualize the robot's path during HRI. It is further possible to combine the textual summary with the geographical summary by marking relevant locations such as references to a product mentioned by a customer.
- **Transcript Panel**: The summary of the human-robot conversation might not include data that turns out to be important. Therefore, the remote expert should still have the opportunity to browse through all transcribed data.
- **Time Travel Slider**: The slider element is connected to the video and transcript panel, automatically jumping to the timestamp attached to the corresponding frame in the video and the text snippet in the transcript panel.
- **Summary Panel**: This panel uses our conversation summarization approach (see Section 3.2). By presenting the paragraphs that are considered most significant to a customer's issue, the remote expert should be able to quickly understand the general context of the problem.

4 EXAMPLE USE CASE

To demonstrate how a remote expert may resolve a customer's issue, we implemented a prototypical system based on the frameworks Python Django [9] and AngularJS [11], and created two different scenarios in collaboration with a retail focus group. One resembling a customer in a DIY market, and a second scenario in a regular retail store. For the latter, we recorded a 30-minute long realistic shopping tour at the Innovative Retail Laboratory (IRL) [19]. A Pepper robot [1] was used as a robotic platform. Below, we first outline the scenario before discussing how an expert may solve the issue using our system. Please note, in this conceptual work, we do not consider data privacy issues that may impact the design or available data in the interface based on customers' consent.

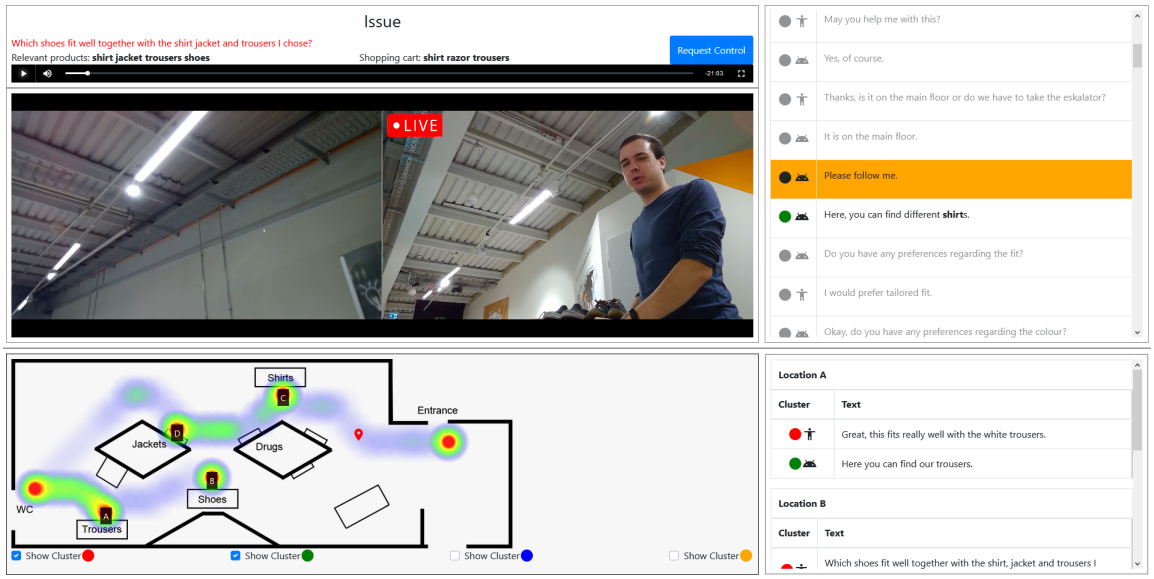

Figure 2: Interactive summary GUI consisting of issue, video, map, transcript and summary panel.

Scenario: A customer needs a new outfit for a wedding and asks the autonomous shopping assistant for help. The person needs items such as shoes, a shirt, a jacket and trousers. Together, they visit different sections of the retail store encountering numerous situations that the robot is familiar with. For instance, the customer may ask: "Where can I find this product?", "Do you have it in a different size?" or "Do you have this item in stock?". Such requests are common and should not cause an issue. However, there will always be situations where the robot cannot solve a problem or the customer requests a human's opinion. In our scenario, the customer and the shopping assistant have been on a shopping tour – selecting products, talking, and doing things unrelated to shopping itself (e.g. visiting the washroom), before the customer asks the following question: *"Which shoes fit well together with the shirt, jacket, and trousers I chose?"* triggering the ToC as the semantic complexity of the request exceeds the capabilities of the system.

Interactive summary: The remote expert receives a request as the robot is unable to assist the customer. First, s/he needs to develop an understanding of the situation to provide optimal assistance. Here, the interactive summary visualizes the ToC parameters discussed in Section 3.1. Initially, the expert could investigate the cause of the ToC and combine this with the most relevant events (see Figure 2 – indicated in red), which should give a good indication of the problem. The expert can quickly derive information about the shirt and trousers that are in the shopping cart. However, in our scenario, the expert does not have information about the jacket as it is not part of the shopping cart. Therefore, the interactive summary GUI allows the expert to (1) enable the second summary layer (see Figure 2 – indicated in green), and (2) quickly travel to the point in time where the customer and the autonomous shopping assistant visited the jacket section. Here, s/he can either review the (summarized) transcript or re-watch the full scene if necessary. At this point, all required information has been acquired to *take-over*.

5 DISCUSSION

Customer acceptance plays a significant role for shopping assistant robots to carry out their job as intended. The majority of results published in this field indicate a generally positive sentiment towards robots in shopping environments [16–18, 25, 26]. However, customers might feel uncomfortable to interact with such a robot because of design-related reasons. Chen et al. [6] report that some customers felt horror due to the appearance of their robot. As the use case we have observed in this paper strongly depends on the aspect of acceptance, a robot needs to appeal to the customers, especially from the perspective of the store owner.

Furthermore, although Large Language Models (LLMs) demonstrate great potential in many areas [5], their use has to be carefully considered. This is because the generated summary, including textual information, may depend on many factors that require knowledge about the environment, and linking summarized portions back to the environment may be challenging. Extensive priming of an LLM might be necessary to achieve the desired results. Since shopping assistant robots engage with customers, it is important to note that privacy concerns must also be addressed. This may include a mechanism for asking if s/he wishes their data to be transferred to the remote expert or additional anonymization safeguards, such as blurring the customer's face or changing her/his voice.

Finally, to measure the effectiveness of our proposed approach, future work should pursue a field study in a real shopping environment. To this end, its usability should be evaluated using established usability questionnaires and in close cooperation with the industry.

ACKNOWLEDGMENTS

We would like to thank Julian Wolter, Daniel Tabellion, and Moritz Wolf for their contributions to this research project. This work is supported by the German Federal Ministry of Education and Research (grant no. 01IW17004) as a part of TRACTAT.

REFERENCES

[1] Aldebaran United Robotics Group. 2023. *Pepper.* https://www.softbankrobotics.com/emea/en/pepper

[2] Mehdi Allahyari, Seyedamin Pouriyeh, Mehdi Assefi, Saeid Safaei, Elizabeth D. Trippe, Juan B. Gutierrez, and Krys Kochut. 2017. Text summarization techniques: a brief survey. *arXiv preprint arXiv:1707.02268* (2017).

[3] Francesca Bertacchini, Eleonora Bilotta, and Pietro Pantano. 2017. Shopping with a robotic companion. *Computers in Human Behavior* 77 (Dec. 2017), 382–395. https://doi.org/10.1016/j.chb.2017.02.064

[4] Santosh Kumar Bharti and Korra Sathya Babu. 2017. Automatic keyword extraction for text summarization: A survey. *arXiv preprint arXiv:1704.03242* (2017).

[5] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. 2023. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712* (2023).

[6] Yingfeng Chen, Feng Wu, Wei Shuai, and Xiaoping Chen. 2017. Robots serve humans in public places—KeJia robot as a shopping assistant. *International Journal of Advanced Robotic Systems* 14, 3 (2017). https://doi.org/10.1177/1729881417703569

[7] Nicola Doering, Sandra Poeschl, Horst-Michael Gross, Andreas Bley, Christian Martin, and Hans-Joachim Boehme. 2015. User-centered design and evaluation of a mobile shopping robot. *International Journal of Social Robotics* 7, 2 (2015), 203–225. https://doi.org/10.1007/s12369-014-0257-8

[8] Wafaa S. El-Kassas, Cherif R. Salama, Ahmed A. Rafea, and Hoda K. Mohamed. 2021. Automatic text summarization: A comprehensive survey. *Expert Systems with Applications* 165 (2021), 113679. https://doi.org/10.1016/j.eswa.2020.113679

[9] Django Software Foundation. 2023. *Get started with Django.* https://www.djangoproject.com

[10] Dylan F. Glas, Satoru Satake, Florent Ferreri, Takayuki Kanda, Norihiro Hagita, and Hiroshi Ishiguro. 2012. The network robot system: enabling social human-robot interaction in public spaces. *Journal of Human-Robot Interaction* 1, 2 (2012), 5–32. https://doi.org/10.5555/3109688.3109690

[11] Google. 2021. *AngularJS.* https://angularjs.org

[12] Horst-Michael Gross, Hans-Joachim Boehme, Christof Schroeter, Steffen Müller, Alexander König, Erik Einhorn, Christian Martin, Matthias Merten, and Andreas Bley. 2009. TOOMAS: interactive shopping guide robots in everyday use-final implementation and experiences from long-term field trials. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2005–2012. https://doi.org/10.1109/IROS.2009.5354497

[13] Horst-Michael Gross, Hans-Joachim Boehme, Christof Schröter, Steffen Müller, Alexander König, Christian Martin, Matthias Merten, and Andreas Bley. 2008. Shopbot: Progress in developing an interactive mobile shopping assistant for everyday use. In *2008 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, 3471–3478. https://doi.org/10.1109/ICSMC.2008.4811835

[14] Vishal Gupta and Gurpreet Singh Lehal. 2010. A survey of text summarization extractive techniques. *Journal of emerging technologies in web intelligence* 2, 3 (2010), 258–268. https://doi.org/10.4304/jetwi.2.3.258-268

[15] International Federation of Robotics. 2023. *Executive Summary World Robotics.* https://ifr.org/free-downloads/

[16] Yamato Iwamura, Masahiro Shiomi, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2011. Do elderly people prefer a conversational humanoid as a shopping assistant partner in supermarkets?. In *Proceedings of the 6th International Conference on Human-Robot Interaction*. IEEE, 449–456. https://doi.org/10.1145/1957656.1957816

[17] Takayuki Kanda, Masahiro Shiomi, Zenta Miyashita, Hiroshi Ishiguro, and Norihiro Hagita. 2009. An affective guide robot in a shopping mall. In *Proceedings of the 4th ACM/IEEE International conference on Human Robot Interaction*. Assocation for Computing Machinery, 173–180. https://doi.org/10.1145/1514095.1514127

[18] Takayuki Kanda, Masahiro Shiomi, Zenta Miyashita, Hiroshi Ishiguro, and Norihiro Hagita. 2010. A communication robot in a shopping mall. *IEEE Transactions on Robotics* 26, 5 (2010), 897–913. https://doi.org/10.1109/TRO.2010.2062550

[19] Frederik Kerber and Antonio Krüger. 2023. *IRL Innovative Retail Laboratory.* https://www.innovative-retail.de

[20] Takahiro Matsumoto, Satoru Satake, Takayuki Kanda, Michita Imai, and Norihiro Hagita. 2012. Do you remember that shop? computational model of spatial memory for shopping companion robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. IEEE, 447–454.

[21] Ani Nenkova and Kathleen McKeown. 2012. A survey of text summarization techniques. In *Mining text data*. Springer, 43–76. https://doi.org/10.1007/978-1-4614-3223-4_3

[22] Marketta Niemelä, Anne Arvola, and Iina Aaltonen. 2017. Monitoring the acceptance of a social service robot in a shopping mall: first results. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-robot Interaction*. Assocation for Computing Machinery, 225–226. https://doi.org/10.1145/3029798.3038333

[23] Marketta Niemelä, Päivi Heikkilä, and Hanna Lammi. 2017. A social service robot in a shopping mall: expectations of the management, retailers and consumers. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. Assocation for Computing Machinery, 227–228. https://doi.org/10.1145/3029798.3038301

[24] Marketta Niemelä, Päivi Heikkilä, Hanna Lammi, and Virpi Oksman. 2017. Shopping mall robots–opportunities and constraints from the retailer and manager perspective. In *International Conference on Social Robotics*. Springer, 485–494. https://doi.org/10.1007/978-3-319-70022-9_48

[25] Marketta Niemelä, Päivi Heikkilä, Hanna Lammi, and Virpi Oksman. 2019. A social robot in a shopping mall: studies on acceptance and stakeholder expectations. In *Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction*. Springer, 119–144. https://doi.org/10.1007/978-3-030-17107-0_7

[26] Alessandra Maria Sabelli and Takayuki Kanda. 2016. Robovie as a mascot: a qualitative study for long-term presence of robots in a shopping mall. *International Journal of Social Robotics* 8, 2 (2016), 211–221. https://doi.org/10.1007/s12369-015-0332-9

[27] Oguzhan Tas and Farzad Kiyani. 2007. A survey automatic text summarization. *PressAcademia Procedia* 5, 1 (2007), 205–213. https://doi.org/10.17261/Pressacademia.2017.591