This project has received funding from the European Community's
Horizon 2020 Framework Programme under grant agreement 860410

Deliverable 2.4
The use of AI in AV human-factors research and human-factors requirements in AI-based AV design

Deliverable Status	**FINAL**
Type	**REPORT**
Dissemination level (according to the proposal)	**PUBLIC**

This project has received funding from the European Community's
Horizon 2020 Framework Programme under grant agreement 860410

Project Information

Project Name:	SHAPE-IT – Supporting the interaction of Humans and Automated vehicles: Preparing for the Environment of Tomorrow
Grant Agreement:	860410
Project Duration:	1 October 2019 – 31 March 2024
Coordinator	Chalmers University of Technology

Document Information

Due Date	30Nov2023	**Due date in months**	M50
Date of submission	30Nov2023	**Month of submission**	M50

Status	**Author's Name and Company/Organisation**	**Date**
Prepared by	Amna Pir Muhammad, Chi Zhang, Eric Knauss, Christian Berger, UGOT	15 Oct 2023
Reviewed by	Klaus Bengler, TU Munich	1 Nov 2023
Approved by	Jonas Bärgman, Chalmers	27 Nov 2023

Please cite as:
Berger, C., Zhang, C., Muhammad, A.P., Knauss, E. (2023). *The use of AI in AV human-factors research and human-factors requirements in AI-based AV design.* (Deliverable D2.4 in the EC ITN project SHAPE-IT). SHAPE-IT Consortium.
DOI: http://doi.org/10.17196/shape-it/2023/D2.4

Acknowledgements
This is a deliverable from the SHAPE-IT project – a project that is multi-disciplinary and with substantial interaction across the early-stage researchers (ESRs), supervisors across the six host universities, and the associated partners of SHAPE-IT (see list below). We consequently want to sincerely thank all individuals and organizations that have contributed to the SHAPE-IT project. We especially want to thank the associated partners, for making the secondments support the ESRs research, but also as a way for the ESRs to get insights into how other organizations work. Finally, this project would not have been possible without the Marie Skłodowska-Curie program. That is, this project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 860410.

List of Associated Partners

Asymptotic AI	https://www.asymptotic.ai
BMW	https://www.bmw.com
Institute of Transport Economics (TØI)	https://www.toi.no
SWOV	https://www.swov.nl
Toyota Motor Europe NV/SA (TME)	https://www.toyota-europe.com
Viscando	https://viscando.com
Volkswagen	https://www.vw.com
Volvo Cars	https://www.volvo.com
Zenseact	https://www.zenseact.com

Note 1: The term Early-Stage Researcher (ESR) is used extensively in this document. The ESRs are PhD candidates funded by the SHAPE-IT project.

Note 2: This document has been submitted to the EC for acceptance as a deliverable in the SHAPE-IT project. If changes are requested by the EC, the document will be updated.

Contents

This project has received funding from the European Community's
Horizon 2020 Framework Programme under grant agreement 860410

1 Summary

The design of automated vehicles (AVs) today is being enabled by the rise of new technologies, actually in particular recent advances in Artificial Intelligence (AI). Navigating the challenges and potential of this technology is crucial for the organizations that develop AVs, as well as for societies that rely on smart transportation.

In this report, we consider two perspectives on these technologies in AV research and design, with a particular focus on human factors (HF): (A) Human-Factors Requirements in AV Development, and (B) The Use of AI in Research about Vehicle-Human-Interaction. We describe each part separately; they are different enough to stand on their own, while both descriptions together make up this report.

We start with the first perspective – investigating how AI can facilitate HF research and practical use of AI to predict human behaviour for use by HF designers.

To support HF researchers and automation designers with tools for classifying and predicting interaction behaviours between AVs/vehicles and pedestrians in urban environments, we developed AI-based models (eg., Zhang et al., 2023) to predict the outcomes of pedestrian-vehicle interactions at unsignalised crossings. The models include random forest models, support vector machine models, and neural network models. The input consists of multiple features such as time to arrival (TTA), pedestrian waiting time, presence of a zebra crossing, and properties and personality traits of both pedestrians and drivers. The output consists of interaction outcomes such as crossing behaviour, crossing duration, and crossing initiation time. The predicted outcomes can contribute to a better understanding of the interactions. In addition, we analysed the interaction factors in order to support HF researchers and automation designers in their efforts to design safer interaction interface.

We reviewed a large selection of papers that used AI to predict pedestrian behaviour and interactions (Zhang and Berger, 2023). We proposed a framework of AI-based tools for predicting pedestrian behaviours and summarized some guidelines for using AI—especially deep learning methods for pedestrian behaviour and interaction prediction. Furthermore, our own body of work (Zhang et al., 2021, Zhang and Berger, 2022a, Zhang and Berger, 2022b, Zhang et al., 2023) provides detailed steps for developing an example of an AI model.

A key contribution of our research is metrics that allow the evaluation and assessment of AI's success at classifying and predicting pedestrian-vehicle interactions. In our study, we compared AI models with traditional linear models (Zhang et al., 2023). Further, we compared the performance of AI models and traditional methods with fewer input factors; traditional methods perform well when there are fewer, while AI-based methods perform better when dealing with more input factors. This finding provides information for optimal model selection in different scenarios. To summarize, our findings suggest that AI can help us understand the intentions of human actors and predict their next steps when they interact with AVs.

The second perspective investigates how HF research can facilitate AV development activities. We had anticipated that the reliance of AV on AI technology might play a major role in how developers need to think about HF (hence, this aspect is also reflected in the title of this report). Our reasoning was that AI-based AV provide a larger surface of interaction between humans and AVs, not only through the traditional human machine interface. However, early in the project we identified that there was a need to address not only the AI-based aspects of HF requirements in AV development, but also to address HF requirements overall in AV development – not the least within agile ways of working. We therefore decided to include AI-based AV development considerations as part of the larger scope of studying HF requirements in the context of AV development, with focus on agile processes. The agile angle was chosen as AV development increasingly incorporates agile and continuous development approaches. We find that it is conceptually unclear how to systematically incorporate HF in such a fast-paced environment. Further, the automotive industry used as our subject of study is lacking guidelines (as well as best practices) for incorporating HF into these ways of working. We propose the development and application of a HF requirements strategy to manage key implications, for which our research suggests useful templates and guidelines.

2 The use of AI in Human-Factors Research

2.1 Background and Motivation for Vehicle Human-Factors Research

Since the number of motor vehicles expanded significantly from 0.85 billion to 2.1 billion between 2000 and 2016, driving safety is becoming an important issue that attracts attention from industry and academia. Great efforts need to be put into reducing the number of people fatally injured in traffic. According to the World Health Organization, approximately 1.3 million people are killed each year in road traffic accidents, of which 310,000 are pedestrians, accounting for approximately 23% of all deaths worldwide (WHO, 2018). This is an unacceptably high figure that must be lowered.

Most pedestrian-vehicle incidents happen when people are crossing the street (Do et al. 2014). Pedestrians commonly interact with moving vehicles in this situation (Zhang and Berger, 2023). By comprehending the interactions between vehicles and humans (i.e., pedestrians), we can forecast pedestrian behaviour more accurately and lower the risk of collisions. Investigating vehicle-human interactions can also provide additional information about human behaviour that can be useful for developing AVs.

The vehicular automation industry has made numerous efforts to improve driving safety by utilizing technology. However, human error remains a significant risk factor, one of which is human error (Rumar, 1999). As technological advancements pave the way for ever-higher levels of automation on roads, the systematic study of HF (especially in the context of AV) is becoming increasingly important. Sensors in combination with prediction systems can provide information about vulnerable road users (VRUs) earlier, reducing the driver's cognitive load and providing the driving system with more time to react than if only sensors without a prediction component were used. It is crucial to comprehend and forecast how (automated) vehicles interact with pedestrians. Therefore, AV development must incorporate considerations about HF during the AV development lifecycle. By using this information, we can predict pedestrian behaviour more accurately, providing safer self-driving technologies and designing safer AVs.

2.2 AI Methods for Vehicle Human-Factors Research

The interactions between automated cars and pedestrians are very intricate, especially when we need to consider HF. There are several variables that can affect the interaction in real-world traffic. The following are a few examples of such variables:

- the actions and characteristics of pedestrians, including posture, direction of travel, age, gender, and even personality traits;
- the vehicle status, including speed, acceleration, direction, and size. If the car is manually driven, the interaction may also be impacted by the behaviour of the car;
- the surroundings, including the road's dimensions and topology, as well as traffic lights and signage.

The interactions between AVs and pedestrians have been analysed and predicted by researchers using a variety of techniques. In one example, Rasouli et al. (Rasouli et al., 2019) divided the interaction into two parts: understanding how vehicles and pedestrians communicate and understanding the intentions of pedestrians. Our research largely concentrated on the second part, understanding, and forecasting the intentions of pedestrians, taking the interaction as input features (Zhang et al., 2021, Zhang and Berger, 2022, Zhang et al., 2023).

Given that the variables mentioned above affect pedestrian behaviour and interactions, it is difficult for traditional analysis methods to predict pedestrian behaviour with any degree of consistency or accuracy. Artificial intelligence (AI) methods have the potential to be useful during the development of prediction systems (e.g., the training an AI-enabled system) or during the data analysis of pedestrian interaction.

Sarker (2021) describes AI as the simulation of human intelligence processes by computer systems (Sarker, 2021). These processes include learning (the acquisition of information and rules for using the information), reasoning (using the rules to reach approximate or definite conclusions), and self-correction. Machine Learning (ML) is a subset of AI that provides systems with the ability to automatically improve with experience without being explicitly programmed (Zhang, 2022). It involves using algorithms to parse data, learn from the data, and then make a prediction about something in the world. The ML models used in our research include logistic regression and linear regression, Support-Vector Machine (SVM), Random Forest (RF). A subset of ML, Deep Learning (DL), utilises artificial neural networks (ANN), which are inspired by the structure and function of the brain. DL uses neural networks with many layers, called deep neural networks, to extract latent information and learn complex patterns in large amounts of data. It has been applied to a variety of tasks, such as computer vision and natural language processing. The non-linearity of the model can be detected using the deep learning network's activation functions. The DL models used in our research include Recurrent Neural Networks (RNNs), including the variant Long Short-Term Memory (LSTM) networks; Generative Adversarial Networks (GANs); Convolutional Neural Networks (CNNs); and Transformer (TF) networks.

To help us understand and predict patterns of pedestrian behaviour, we use DL and other ML algorithms to recognize the interactions between pedestrians and vehicles.

2.3 Evaluation Metrics for AI-based Models

We use AI-based models to predict pedestrian-vehicle interactions, including the trajectories and crossing actions of pedestrians. The models consider the current pedestrian-vehicle and pedestrian-AV interactions in order to predict the pedestrians' future trajectories during the interaction. We can use the displacement error that measures the deviation between prediction and ground truth, defined as below following broadly accepted metrics in the community (C. Zhang, C. Berger, and M. Dozza, 2021), to report the prediction error and evaluate the model's performance:

The Average Displacement Error (ADE): the average distance gap between ground truth and prediction trajectories over all predicted time-steps.

$$ADE = \frac{\sum_{i \in n_p} \sum_{t=T_{obs}+1}^{T_{pred}} \left\| Y_t^i - \widehat{Y}_t^i \right\|_2}{n_p \times \left(T_{pred} - T_{obs} \right)}$$

The final displacement error (FDE): the average distance gap between ground truth and prediction trajectories for the last predicted time-step.

$$FDE = \frac{\sum_{i \in n_p} \left\| Y_t^i - \hat{Y}_t^i \right\|_2}{n_p}, \quad t = T_{pred}$$

To evaluate the accuracy of the prediction regarding the pedestrian's decision whether to cross during the interaction, we use the following metrics:

P is the number of positives

N is the number of negatives

TP denotes true positives

TN denotes true negatives

FP denotes false positives

FN denotes false negatives.

With these fundamental elements, we define accuracy (ACC) as how many predictions were correct out of all the predictions and precision measures how many of the positive predictions were correct, i.e., TP over the sum of TP and FP.

We define recall as the ratio between TP and the sum of TP and FN to describe the sensitivity of a predictor.

We define the F1 score as a balance between precision and recall as it considers both false positives and false negatives. The prediction accuracy and F1 score are used for evaluation as shown below:

$$ACC = \frac{TP + TN}{P + N}$$

$$F1 = \frac{2TP}{2TP + FP + FN}$$

We conducted research into pedestrian trajectory prediction and improved the state-of-the-art after evaluating the results of the interaction between the pedestrian and a vehicle. Our proposed neural network model improves the prediction accuracy and F1 score by 4.46% and 3.23%, respectively, and reduces the root mean squared error (RMSE) for crossing initiation time and crossing duration by 21.56% and 30.14%, respectively.

2.4 Overview of Research Outcomes

A method for predicting pedestrian trajectories which considers the social interactions between pedestrians as part of the model:
C. Zhang, C. Berger, and M. Dozza "Social-IWSTCNN: A Social Interaction-Weighted Spatio-Temporal Convolutional Neural Network for Pedestrian Trajectory Prediction in Urban Traffic Scenarios" In proceedings of the 2021 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2021. https://doi.org/10.1109/IV48863.2021.9575958

A method for predicting pedestrian trajectories which considers the interactions between pedestrians and vehicles as part of the model:
C. Zhang and C. Berger "Learning the Pedestrian-Vehicle Interaction for Pedestrian Trajectory Prediction" In 2022 the 8th International Conference on Control, Automation and Robotics (ICCAR). IEEE, 2022. https://doi.org/10.1109/ICCAR55106.2022.9782673

Analysing the interactions between pedestrians and automated vehicles using deep learning methods:
C. Zhang and C. Berger "Analyzing Factors Influencing Pedestrian Behavior in Urban Traffic Scenarios Using Deep Learning" In Transport Research Arena (TRA), 2022. Elsevier.
Analysing and predicting the interactions between pedestrians and vehicles at unsignalized crossings, considering the human personality traits, using machine learning methods:
C. Zhang, A. H. Kalantari, Y. Yang, Z. Ni, G. Markkula, N. Merat, and C. Berger "Cross or Wait? Predicting Pedestrian Interaction Outcomes at Unsignalized Crossings", In proceedings of the 2023 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2023.
https://doi.org/10.1109/IV55152.2023.10186616

3 Human-Factors Requirements in AV Development

For more than a decade, automotive industries are competing to bring AVs to the market. With increasing levels of vehicle automation, AVs are becoming more popular and capable. AVs rely on improvements in various AI methods and strategies to understand their circumstances better and make better driving decisions. In this context, AI is a technology that helps the AV imitate the way a human driver would operate a car.

Significantly, however, AVs promise many benefits, for example: fewer accidents, injuries, and deaths resulting from human-driver-caused crashes; increasing mobility of the young, adults, and elderly; and enabling drivers to engage in other activities while in the car (Fagnant & Kockelman 2015).

However, AVs also pose various challenges for humans, such as extra pressure and workload, driver engagement and re-engagement, AI decision-making capability, and testing & evaluation methodology (Billings, 2018; Merat et al., 2014).

The challenges are not limited to drivers; they also impact other individuals on the road who interact with AVs. The technology that permits a vehicle to sense, perceive, and "understand" its surroundings is steadily improving. However, an AV also needs to communicate its intentions and decisions unambiguously and transparently to its users and surroundings. HF research is highly relevant in this aspect and can significantly support this integral part of the engineering and assessment processes. HF researchers strongly advocate the integration of HF knowledge into the design of AVs to unlock the full potential of automation (Hancock, 2014, Hancock, 2017, Hancock, 2019, Lee, 2008, Navarro, 2019).

HF is a field of study that involves the investigation of human capabilities, limitations, and other human attributes, with the aim of applying these findings to enhance the performance, safety, and comfort of systems (HFES, 2023).

In fact, researchers advocate including HF knowledge in the early stages of development, specifically while developing automation systems (Chua & Feigh 2011; Håkansson & Bjarnason, 2020). Although there are guidelines on implementation considerations of HF aspects in vehicle development (Novakazi, 2023, Cao et al., 2021), they are focusing on HF and its importance. In contrast, the work presented here focuses on how to integrate HF related work in the overall design and development lifecycle.

Traditionally, this knowledge has been incorporated in the Requirements Engineering (RE) phase, which is the process of gathering, examining, recording, and validating requirements (Kotonya & Sommerville, 1998).

However, the increasing use of agile development approaches in the automobile sector considerably alters the function of RE. Agile development methods are a group of principles and practices based on incremental and iterative development, whereby self-organizing and cross-functional teams collaborate to create requirements and functionality (Moniruzzaman & Hossain, 2013). Agile approaches aim to deliver products to market faster. Due to competition, developers seek to deliver faster; as a result, they are at risk of focusing on technical aspects and overlooking others, such as those offered by HF. Moreover, because agile methodologies do not focus on the traditional processes, RE processes are not well integrated with agile methodologies and face different challenges (Meyer, 2014).
Without a clear role for RE in agile development, it may be challenging to include knowledge of human aspects as requirements. The lack of empirical research on how to include HF knowledge in agile development increases the difficulty; practitioners struggle due to a lack of specific guidance. Therefore, the research reported on in this deliverable (D2.4) investigates how we can efficiently provide AV developers with requirements based on HF expertise in large-scale, agile AV development.

3.1 Research Approach

The research approach used in this work mainly follows the empirical paradigm and focuses on investigating challenges and finding solutions in the real world. Empirical studies that focus on a real-world problem make it much easier to explore newly emerging topics of interest to the industry (Easterbrook et al., 2008). These studies can broaden our understanding of the problem and steer us toward promising solutions for managing requirements and HF in large-scale, agile AV development. While empirical investigations may employ qualitative or quantitative approaches (Creswell & Creswell, 2017), we primarily used qualitative research techniques for our exploratory investigation.

3.2 The Need for Developers to Understand Human Factors

When building complex, AI-intensive systems and products, it is important to focus on the technical aspects of the system along with HF (e.g., human capabilities and limitations, as well as different aspects of user experiences) (Heyn et al., 2021). Automated systems impose new types of challenges on humans, such as activating and deactivating automated features. Several studies have shown that when an AV is driving without human input (when both lateral and longitudinal control is handled by the vehicle), the humans (drivers) relax and lose focus on the roadway and driving—even when they have been informed that the system is not perfect. Humans are just not suited for monitoring tasks. However, if/when the system fails to perform the driving task and the drivers are notified that they should take over control, the situation can quickly become critical, possibly leading to a crash, if the human cannot focus and react in time. At the same time, if the system works as intended, disabling it may increase the risk of crashing, because its conflict and crash avoidance performance is substantially better than that of the human driver. This issue is well known, but still there are many aspects of vehicle automation design in which the human's role is unknown. A human-centric design philosophy, with knowledge of HF, can help identify these aspects and mitigate any negative consequences earlier, thus guiding development according to human capabilities and limitations. The ultimate result will be a safe, acceptable, and reliable system by design. Therefore, it is important to understand and incorporate HF knowledge during development, in order to successfully deploy AVs that reduce the number of accidents and improve mobility. For this concern, HF must be considered earlier in the development phases, right when the concepts are developed, i.e., in the RE phase. However, with agile development's tendency to neglect upfront analysis and heavy processes, it is challenging to include HF knowledge in the development of complex AI-intense systems such as AVs.

The research of ESR8 clearly establishes the need to develop a catalogue of good practices for managing HF knowledge in agile AV development.

3.3 Implications for Agile Ways of Working with Human Factors and Requirements Engineering

As a first step to overcoming the challenges of integrating HF into agile AV development, we have investigated the implications for agile ways of working, HF, and RE, when relying on all three in AV development (Muhammad, et al., 2023).

Implications for the agile way of working to incorporate HF. We found that agile AV development teams must have access to HF knowledge when making their local design decisions. This is implied by the need for agile teams to take responsibility for parts of the product, to incorporate all knowledge needed to develop those parts, and to be autonomous in their decisions within the scope of the parts' development.

Agile development strives to be responsive to changing goals and requirements. Consequently, the specific knowledge needs related to HF might not be anticipated up front. Thus, agile AV development teams must be able to acquire HF knowledge when needed, which makes it desirable to include HF experiments within the iterative work of agile teams. The implication that agile teams must be able to acquire HF knowledge raises questions about how to manage these experiments' design and evolution in agile work.

Given the lack of HF expertise in agile AV development teams, it is important that we formulate a strategy for agile AV development that takes HF into account. As the automotive industry shifts towards agile methodologies as well as continuous integration and delivery, novel collaborative approaches with suppliers are emerging. These approaches involve the close integration of suppliers into incremental work in order to meet distinct objectives. Consequently, our final implication for an agile approach is to methodically determine whether, and how, to incorporate suppliers into the scaled-agile development of AVs.

Implications for Working with Human Factors. Our findings indicate that HF experts (who have accumulated knowledge) should be in close contact with agile teams in order to raise awareness, enable relevant questions to be asked (regarding human behaviour and capabilities), and guide teams in the right direction. HF experts should also provide basic HF knowledge, in the form of checklists and design principles, to development teams.

Implications for Requirements Engineering. RE can support managing HF in large-scale, agile AV development, by effectively managing the knowledge acquired from experiments and by expressing design decisions in relation to HF requirements in the backlog. The other implication for RE is to increase the ability to prototype for requirement elicitation and validation based on the identified needs and HF checklists within agile teams. The last implication is to express the relationship between design decisions and HF knowledge, typically via trace links. In the context of the other implications, describing these relationships requires that system requirements be created together with the system/software, not before. It follows that the requirements, in the form of stories, would need to be provided during development rather than at the beginning of development.

3.4 Towards defining a Requirements Engineering Strategy that Incorporates Human Factors in Agile

Our implications discussed above indicate a need to evolve the agile way of working, the management of HF knowledge, as well as RE. To do so, practitioners are lacking concrete guidelines and best practices on how to incorporate HF knowledge in agile AV development. To address this challenge, we designed a template, shown in Table 1, that helps organizations develop a concrete requirements strategy (Muhammad et al., 2022) to integrate RE into agile development. This template comes with guidelines for creating a solution approach to outlining RE tasks within the context of an agile development process. It encompasses three complementary perspectives as building blocks: a structural view, an organizational view, and a view centred around work and feature flow.

The objective of establishing a requirements strategy is to foster a shared understanding of requirements (Cannon-Bowers & Salas, 2001) among these views, with a particular emphasis on creating a shared vocabulary and promoting the flow of information.

We recommend starting with the structural view, by defining the structure of requirements to establish a common terminology. Next, the responsibilities for managing requirements knowledge within the organization should be defined. The final building block is mapping both the structure and organizational roles to the agile workflow.

To design a requirements strategy aimed at addressing the challenges of RE in agile development from a structural perspective, it is imperative to understand the types of requirements, their levels of abstraction, and whether templates exist for these requirements. For instance: Are there high-level requirements? Can these requirements be broken down into more detailed specifications? Additionally, consider the need for traceability in this context.

Table 1: Template with Key Building Blocks of a Requirements Strategy

	Support for shared understanding		
Perspective	**Common Language**	**Knowledge flow**	**Examples**
Structural	Define reqts. Levels Define reqts. Types Define templates	Define structural decomp. Define traceability demands	Stakeholder, System, Component Requirements Requirements and Traceability Information Model User stories include customer value and goal
Organizational	Define ownership of reqts. types	Define roles and responsibilities	Training plan per type/role; Team responsibility sheet
Work and feature flow	Define lifecycle of types	Map structure to Workflow Map organization to workflow	Elicitation strategy, definition of done Stakeholder map, requirements review strategy

The organizational view focuses on roles and responsibilities (which must somehow be combined with the items in the structural perspective). We need to address questions such as who owns which requirements, which roles exist in the company and what their responsibilities are, and how these roles relate to the requirements. In this view, it is essential to prevent any gaps in responsibility; otherwise, there is a risk that individuals may assume others are handling specific tasks, which may than not be dealt with at all or only painfully late.

The third perspective involves aligning the RE strategy with the agile workflow and feature development process. In this perspective, it is crucial to align the structural and organizational aspects into clear guidelines on how they fit into the work and feature flow. This alignment can be partially achieved by establishing a definition of done criteria, for example. Moreover, it is also crucial to link the work and feature flows with the roles, responsibilities, and ownership of requirements. A stakeholder map can be a valuable tool, as it defines artifact ownership, communication recipients, and review stakeholders. Defining clear requirements review strategy can be highly beneficial, enhancing the quality of requirements and keeping reviewers updated on recent changes.

A requirements strategy should be created and systematically documented to ensure all objectives are properly addressed and understood by all stakeholders. This strategy should include methodologies, tools, and templates aimed at strategically addressing the challenges in RE within an organization. It should evolve over time to adapt to changing organizational needs, methods, and products. Additionally, the requirements strategy should serve to harmonize diverse stakeholders in terms of terminology, requirement categories, requirements abstraction levels, roles and duties, traceability, resource allocation, etc. (Zhang et al., 2013).

These guidelines aim to help organizations integrate RE effectively into agile development, bridging the gap between traditional upfront requirement phases and agile methodologies. The goal is to manage requirements effectively without contradicting the organization's agile objectives. This approach is adaptable and customizable for specific domains, making it valuable for any agile organization. Ultimately, a requirements strategy provides a framework for aligning RE activities with agile system development and can inform further research and best practices in the industry.

4. Overview of Research Outcomes

Identified properties for agile and human factors for AV development and provided some implications for human factors, agile way of working, and requirements engineering. (Note that this paper is included as an Appendix; under Creative Commons license CC BY 4.0 DEED, published by Elsevier):

> Muhammad, A.P., Knauss, E. and Bärgman, J. (2023) “Human factors in developing automated vehicles: A requirements engineering perspective,” Journal of Systems and Software, vol. 205, p. 111810. https://doi.org/https://doi.org/10.1016/j.jss.2023.111810

Identified the current challenges and practices in large-scale agile AV development:

> Muhammad, A.P., Knauss, E., Bärgman, J. and Knauss, A. (2023) “Managing Human Factors in Automated Vehicle Development: Towards Challenges and Practices,” in 31st International Requirements Engineering Conference (RE). IEEE.

Defined requirements strategy to address challenges related to requirements engineering agile development:

Muhammad, A. P., Knauss, E., Batsaikhan, O., Haskouri, N. E., Lin, Y.-C. and Knauss, A. (2022) "Defining requirements strategies in agile: a design science research study," in International Conference on Product-Focused Software Process Improvement. Springer, pp. 73–89

Problem description:

Muhammad, A.P. (2021) "Methods and guidelines for incorporating human factors requirements in automated vehicles development." in REFSQ Workshops

General exploration of problem areas related to requirements engineering in developing AI-intense systems.

Heyn, H.-M., Knauss, E., Muhammad, A. P., Eriksson, O., Linder, J., Subbiah, P., Pradhan, S. K. and Tungal, S. (2021) "Requirement engineering challenges for ai-intense systems development," in 2021 IEEE/ACM 1st Workshop on AI Engineering-Software Engineering for AI (WAIN). IEEE, pp. 89–96.

5. References

Billings C. E.. (2018) Aviation automation: The search for a human-centered approach. CRC Press.

Cannon-Bowers, J. A. and Salas, E. (2001) "Reflections on shared cognition," Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior, vol. 22, no. 2, pp. 195–202.

Cao, Y., Griffon, T., et al., (2021) "Code of Practice for the Development of Automated Driving Functions," L3Pilot public deliverable D2.3, www.L3Pilot.eu.

Chua Z. K. and Feigh, K. M. (2011) "Integrating human factors principles into systems engineering," in 2011 IEEE/AIAA 30th Digital Avionics Systems Conference. IEEE, pp. 6A1–1.

Creswell J. W. and Creswell J. D. (2017) Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

Do, A. H., Balk, S. A. and Shurbutt, J. W. (2014) "Why did the pedestrian cross the road?" Public Roads, vol. 77, no. 6, 2014.

Fagnant D. J. and Kockelman K. (2015) "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations," Transportation Research Part A: Policy and Practice, vol. 77, pp. 167–181.

Hancock, P. A. (2014) "Automation: how much is too much?" Ergonomics, 57 (3), pp. 449-454

Hancock, P. A. (2017) "Imposing limits on autonomous systems," Ergonomics, vol. 60, no. 2, pp. 284–291.

Hancock, P. A. (2019) "Some pitfalls in the promises of automated and autonomous vehicles," Ergonomics, 1, DOI: 10.1080/00140139.2018.1498136

Heyn, H.-M., Knauss, E., Muhammad, A. P., Eriksson, O., Linder, J., Subbiah, P. Pradhan, S. K., and Tungal, S. (2021) "Requirement engineering challenges for ai-intense systems development," in 2021 IEEE/ACM 1st Workshop on AI Engineering-Software Engineering for AI (WAIN). IEEE, 2021, pp. 89–96.

Human Factors and Ergonomics Society (2023) "Definitions of human factors and ergonomics,", accessed on 30 Mar, 2023. [Online]. Available: https://www.hfes.org/About-HFES/What-is-Human-Factors-and-Ergonomics

Håkansson E. and Bjarnason, E. (2020) "Including human factors and ergonomics in requirements engineering for digital work environments," in 2020 IEEE First International Workshop on Requirements Engineering for Well-Being, Aging, and Health (REWBAH). IEEE, pp. 57–66.

Kotonya G. and Sommerville, I. (1998), Requirements engineering: processes and techniques. Wiley Publishing, 1998.

Lee, J. D. (2008) "Review of a pivotal human factors article: 'humans and automation: use, misuse, disuse, abuse'," Hum. Factors, 50 (3), pp. 404-410

Merat, N. , Jamson, A. H., Lai, F. C., Daly, M. and Carsten, O. M. (2014) "Transition to manual: Driver behaviour when resuming control from a highly automated vehicle," Transportation research part F: traffic psychology and behaviour, vol. 27, pp. 274–282.

Meyer, B. (2014) Agile! Bertrand Meyer: The Good, the Hype and the Ugl. Springer, 2014.
Easterbrook, S. Singer, J.., Storey, M.-A. and Damian, D. (2008) "Selecting empirical methods for software engineering research," in Guide to advanced empirical software engineering. Springer, 2008, pp. 285–311.

Moniruzzaman, A. and Hossain, D. S. A. (2013) "Comparative study on agile software development methodologies," arXiv preprint arXiv:1307.3356, 2013.

Muhammad, A. P., Knauss, E., and Bärgman, J. (2023) "Human factors in developing automated vehicles: A requirements engineering perspective," Journal of Systems and Software, vol. 205, p. 111810.

Muhammad, A. P., Knauss, E., Batsaikhan, O., Haskouri, N. E., Lin, Y.-C. and Knauss, A. (2022) "Defining requirements strategies in agile: a design science research study," in International Conference on Product-Focused Software Process Improvement. Springer, pp. 73–89

Navarro, J. (2019) "A state of science on highly automated driving," Theor. Issues Ergon. Sci., 20 (3), pp. 366-396

Novakazi, F. (2023). *Design for Perception - A human-centric approach to the design of driving automation systems based on the driver's perception* Chalmers University of Technology]. Gothenburg, Sweden. https://research.chalmers.se/en/publication/538287

Rasouli, A., & Tsotsos, J. K. (2019). Autonomous vehicles that interact with pedestrians: A survey of theory and practice. IEEE transactions on intelligent transportation systems, 21(3), 900-918.

Rumar, K. (1999) "Transport safety visions, targets and strategies: beyond 2000," 1st European Transport Safety Lecture. European Transport Safety Council, Brussels, Tech. Rep.

Sarker, I. H. (2021) “Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions,” SN Computer Science, vol. 2, no. 6, pp. 1–20

WHO, “Global status report on road safety 2018: Summary,” (2018) World Health Organization, Report.

Zhang, X., Auriol, G., Eres, H., and Baron, C. (2013) “A prescriptive approach to qualify and quantify customer value for value-based requirements engineering,” International Journal of Computer Integrated Manufacturing, vol. 26, no. 4, pp. 327–345,.

Zhang, C., Berger, C. and Dozza, M. (2021) “Social-IWSTCNN: A social interaction-weighted spatio-temporal convolutional neural network for pedestrian trajectory prediction in urban traffic scenarios,” In: Proceedings of the IEEE Intelligent Vehicles Symposium (IV), pp. 1515-1522.

Zhang, C. and Berger, C. (2022a) “Learning the pedestrian-vehicle interaction for pedestrian trajectory prediction,’ In: Proceedings of the 8th International Conference on Control, Automation and Robotics (ICCAR), pp. 230-236.

Zhang, C. (2022b). Predicting Pedestrian Behavior in Urban Traffic Scenarios Using Deep Learning Methods. Doctoral Licentiate Thesis, Chalmers University of Technology.

Zhang, C. and Berger, C. (2022) ”Analysing Factors Influencing Pedestrian Behavior in Urban Traffic Scenarios Using Deep Learning.” Accepted in Transport Research Arena (TRA). Elsevier.

Zhang, C. and Berger, C. (2023) “Pedestrian Behavior Prediction Using Deep Learning Methods for Urban Scenarios: A Review,” In: IEEE Transactions on Intelligent Transportation Systems.

Zhang, C. Kalantari, A. H. Yang, Y., Ni, Z. , Markkula, G. Merat, N., and Berger, C. (2023) ”Cross or Wait? Predicting Pedestrian Interaction Outcomes at Unsignalized Crossings,” In: Proceedings of the IEEE Intelligent Vehicles Symposium (IV).

Pedestrian Behavior Prediction Using Deep Learning Methods for Urban Scenarios: A Review

Chi Zhang and Christian Berger

Abstract—**The prediction of pedestrian behavior is essential for automated driving in urban traffic and has attracted increasing attention in the vehicle industry. This task is challenging because pedestrian behavior is driven by various factors, including their individual properties, the interactions with other road users, and the interactions with the environment. Deep learning approaches have become increasingly popular because of their superior performance in complex scenarios compared to traditional approaches such as the social force or constant velocity models. In this paper, we provide a comprehensive review of deep learning-based approaches for pedestrian behavior prediction. We review and categorize a large selection of scientific contributions covering both trajectory and intention prediction from the last five years. We categorize existing works by prediction tasks, input data, model features, and network structures. Besides, we provide an overview of existing datasets and the evaluation metrics. We analyze, compare, and discuss the performance of existing work. Finally, we point out the research gaps and outline possible directions for future research.**

Index Terms—**Pedestrian behavior prediction, trajectory, intention, deep learning, neural networks, automated vehicles, survey.**

I. Introduction

According to World Health Organization (WHO)'s report on road safety [1], about 1.35 million people are fatally injured by road crashes every year. Pedestrians constitute 23% of all road traffic deaths globally, which is unacceptably high. As the most vulnerable road users, pedestrians are important participants and need protection. Given that human errors are one of the main factors in most road traffic crashes [2], automated vehicles (AVs) may have the potential to reduce these figures and improve road safety. Hence, it is essential to predict the behavior of pedestrians for AVs to better understand the AV's surroundings for making better and safer driving decisions and preventing potential hazardous situations. In recent years, the interest in AVs has attracted increasing attention to research related to pedestrian behavior prediction.

Predicting pedestrians' behavior is a great challenge. In contrast to the vehicles, whose behavior prediction has been well studied and reviewed by Lefèvre et al. [3] and Mozaffari et al. [4] for instance, pedestrians are more agile and can change their speed and direction unexpectedly [5] with unknown or hardly predictable moving patterns [6]. Pedestrian behavior is driven by complicated influencing factors. These factors include not only the properties of the pedestrians themselves such as the motion states, destination, age, and gender [7], but also the interactions with other pedestrians [8] and vehicles [9]–[11]. Furthermore, the environment can also influence the intention of pedestrians both explicitly and implicitly. The non-linearity arising from pedestrian interactions and the complexity of multiple influencing factors hinder accurate prediction using conventional knowledge-based models such as social force [12] and constant velocity model [13]. Deep learning is a subset of machine learning based on artificial neural networks with multiple layers. Inspired by the biological neuron, artificial neural networks are composed of nodes with linear weights and bias, and non-linear activation functions. Deep learning methods are powerful tools that can be used to extract high-level features from data, and can deal with the non-linearity of the data. Therefore, researchers are exploring the potential of deep learning models to represent and extract pedestrians' behavior patterns in a data-driven manner. In this paper, we analyze and categorize existing research and discuss how current challenges have been addressed so far.

As deep learning methods are data-driven, datasets are important for developing models. The report on pedestrian safety by WHO [14] has shown that about 70% of pedestrian fatalities occur in urban areas in the European Union, and in the United States, this number is about 76%. Pedestrian-vehicle collisions occur more in urban areas than rural areas in these countries, and hence, most of the publicly available datasets for developing pedestrian behavior prediction models used by researchers are collected in urban areas. Therefore, we review prediction methods and datasets in urban scenarios.

The scope of this paper covers studies that predicted pedestrian behavior, including the future trajectory and crossing intention. We focus on deep learning-based models. When it comes to datasets and model inputs, we focus on urban scenarios, and cover various inputs such as camera images, light detection and ranging (LiDAR) point clouds, or the speed of the ego vehicle to name a few. Various factors that influence pedestrian behavior are covered, such as pedestrians' own past motion states, interactions with other pedestrians and vehicles, and influences of the environment.

There are several published papers that reviewed existing works on pedestrian behavior prediction. Hirakawa et al. [15] surveyed vision-based methods for pedestrian path prediction, where deep learning-based methods were only covered by a small extent. Rudenko et al. [16] reviewed the work related to human motion trajectory prediction and categorized existing methods by the modeling approach and contextual cues. Korbmacher and Tordeux [17] reviewed pedestrian trajectory prediction methods, compared deep learning methods and

Chi Zhang and Christian Berger are with the Department of Computer Science and Engineering, University of Gothenburg, 41756 Gothenburg, Sweden. (e-mail:chi.zhang@gu.se, christian.berger@gu.se)

knowledge-based methods. These papers only covered the trajectory prediction and omitted the important prediction of intention that can be used for pedestrian-vehicle collision avoidance. Shirazi and Morris [9] focused on pedestrian intention at intersections and analyzed how crossing behavior is influencing intersection participants. Ohn-Bar et al. [10] provided a survey on interactions between humans and autonomous vehicles. Rasouli and Tsotsos [11] reviewed pedestrian behavior studies of both classical pedestrian–driver interactions and more recent autonomous vehicles and pedestrian interactions, but mainly focused on analyzing human factors and interactions instead of deep learning-based behavior prediction. Ridel et al. [18] reviewed and classified existing pedestrian behavior prediction models, but they classified previous works from only a single criterion, and many recently suggested deep learning methods were not covered. Most of the previous review papers focused on a single task, either the analysis of trajectories [15]–[17] or intention [9], or interactions between pedestrians and vehicles [10], [11], which did not cover the aspects in this paper's scope. Moreover, most of these papers classified the existing literature by a single criterion [17], [18], and only include methods with some particular input data [15].

To overcome the drawbacks listed above, we review, categorize, and analyze the existing research on pedestrian behavior including both the trajectory and intention prediction in this paper. We propose four criteria for classification to consider existing works from different dimensions. The main contributions of this paper are:

- We present a detailed analysis of the existing literature on pedestrian behavior prediction, including trajectory prediction, intention prediction, and the joint prediction of both. We categorize existing approaches from four criteria including a) prediction tasks b) input data, c) the features that are considered in existing models, and d) network structures, and emphasize the advantages and drawbacks of existing approaches.
- We include the most recently proposed existing publicly available datasets and commonly used evaluation metrics. We compare the trajectory and intention prediction tasks on the most commonly used open datasets and present state-of-the-art algorithms.
- We point out research gaps and outline the potential directions for future works.

II. Methodology and Taxonomy

A. Methodology

Our methodology to find and collect existing papers is based on direct search and snowballing. We used IEEE Xplore digital library and Google Scholar for direct search to include both scientific databases and open-access pre-prints. We used "pedestrian behavior prediction" OR "pedestrian trajectory prediction" OR "pedestrian intention prediction" filtered by: "deep learning" OR "network" as initial search strings. We did not set the time range explicitly, but after searching, the results originated mainly from 2016 to 2021. Then we went through the results to select relevant papers meaning that the research targets are pedestrians instead of drivers or robots, the research goal is behavior prediction instead of detection, tracking, or vehicle/robot path planning, and the methods are deep learning. We selected 50 papers from direct searching. Then we did backward and forward snowballing (as proposed in [19]) with their citations and references to include relevant publications, and got 42 papers from snowballing. We review 92 papers in total, including 44 on trajectory prediction, 17 on intention prediction, 6 on joint prediction, 18 on datasets and benchmarks, and 7 on literature review. The number of papers over the years[1] and the distribution of the papers is shown in Fig. 1. The rising trend of the papers indicates the growing interest in this field.

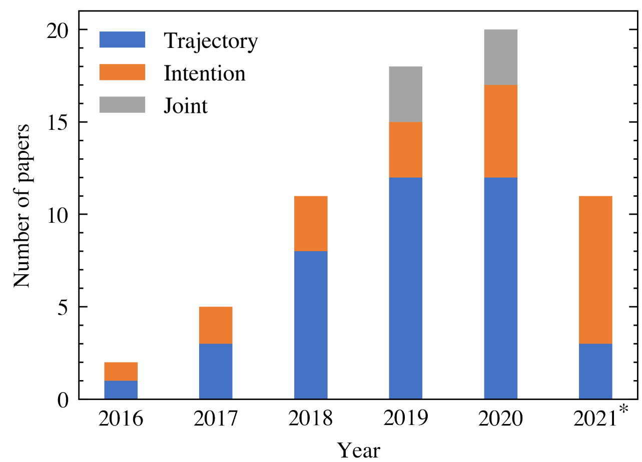


Fig. 1. The number of papers over the years and the distribution of the papers. The rising trend of the papers indicates the growing interest in deep learning-based pedestrian behavior prediction. *Note that in 2021 only the papers published in the **first half** of the year are included.

B. Taxonomy and Overview

We address our expansion of the taxonomy proposed by Hirakawa et al. [15] and Rudenko et al. [16], and categorize existing studies by the following four criteria. With the help of this taxonomy, one can easily get started with a model's desired input and output, decide the features that they want to consider in the model, and find a reasonable network structure.

1) Prediction tasks: The prediction tasks define the problem that a model is addressing, and a model's expected output. We classify previous models by three kinds of prediction tasks, including a) trajectory prediction, b) intention prediction, and c) joint prediction that predict both trajectory and intention.
2) Input data: The input data show the information provided by sensors or annotations that are used as model inputs. We classify previous models by three kinds of input data that provide different types of information, including: a) the past trajectories of pedestrians from annotations, b) the information provided by sensors, and c) other supplementary information such as the map information, the information of the ego vehicle, etc.

[1]Note that in 2021 only the papers published in the first half of the year are included.

3) Model features: There are many factors that influence the future behavior of pedestrians. It is hard to consider all factors, so previous studies tried to cover those factors that influence pedestrians most as model features. Model features are the observations and factors that were considered by previous studies in models as stimuli to the future behavior of pedestrians. We classify previous models by three types of model features, including a) the observed information of target pedestrians, b) the information of other agents that interact with target pedestrians, and c) the information of the environment.
4) Network structures: The network structures show how previous studies learned the moving pattern from observed information. There are several typical structures used in existing prediction models that can be classified into sequential networks and non-sequential networks.

We summarize the pedestrian behavior prediction framework in Fig.2 and show how these four criteria are related. We review and classify the existing works in detail from the proposed categories: prediction tasks as in Sec. III, input data as in Sec. IV, model features as in Sec. V, and network structures as in Sec. VI. Then, we outline the evaluation metrics and the datasets used in existing research in Sec. VII, and compare the performances on publicly available datasets to point out the research gaps and outlines potential research directions in Sec. VIII. Finally, we present our conclusions in Sec. IX.

III. PREDICTION TASKS

In this section, we classify previous studies based on prediction tasks, including trajectory prediction, intention prediction, and joint prediction that predicts both. We cover different output representations and training strategies for each type of task. Table I summarizes different types of prediction tasks, model features, and input data of existing studies.

A. Trajectory Prediction

a) Task definition: The trajectory prediction methods provide low-level information of pedestrian behavior with detailed spatial and temporal information. This information can be used for collision avoidance or helping autonomous vehicles to plan their future path. We define the trajectory of a pedestrian as a sequence of x-y coordinate positions including their temporal order. A person's position in a scene is represented by the x-y-coordinate $X = (x, y)$. Given a set of n pedestrians with their observed positions over time steps t, $X_t^i = (x_t^i, y_t^i)$ where $i \in \{1, \ldots, n\}$, $1 \leq t \leq T_{obs}$, and other information I such as the information of the surrounding environment and objects, we aim to predict the likely trajectories of the target pedestrians $\hat{Y}_t^i = (\hat{x}_t^i, \hat{y}_t^i)$ in the future time steps $T_{obs} + 1 \leq t \leq T_{pred}$.

b) Output representation: There are different kinds of output representations for trajectory prediction. Many researchers treated trajectory prediction as a regression problem. The output can be represented as: a) positions of (x, y) coordinates, b) uni-modal distributions, and c) multi-modal distributions. Representing output as positions is used by many studies, such as [21]–[24], [49]. Such models are simple compared to those models predicting distributions, and can get deterministic results, but they cannot include the randomness nature of the pedestrian movement. Uni-modal distributions are very popular for trajectory prediction and are used by studies such as [27], [28], [39], [41], [45], [52], [59], [62]. Compared to multi-modal distribution models, the uni-modal prediction requires less computational cost, but the model may learn an "average behavior" that is not plausible. Multi-modal distributions can overcome the drawback of converging to average behaviors by outputting several plausible behaviors, and are used by studies such as [31], [33], [35], [38], [40], [44], [47], [50], [51], [58]. But this representation requires higher computational resources with more complicated frameworks such as GANs, and are hard to converge.

Instead of treating the trajectory and positions as a continuous variable and directly regressing values, the trajectory prediction can also be represented as a discrete variable. The output can be represented as: a) discretizing the frame scene into grids, and b) discretizing the pedestrian velocity into bins. Grid-based representations are used by studies such as [48], [56]. Using a grid-based representation to encode the location information enables a parameter-free approximation of distributions, but the discretization over the whole scene may require high dimensionality. Therefore, grids are more often used for representing local occupation information for interaction with neighbors or the environment as in [27], [49]. The trajectory prediction can also be treated as a classification task by quantizing the input data into classes and represented by one-hot encoding. Giuliari et al. [24] used 1000 bins to represent the velocity of pedestrians and predicted the future velocity by classification. But the authors claimed that the classification generally gets worse results than regression models because of quantization errors. In addition to predicting only future trajectories, some work outputs both destination and trajectory prediction [86], or outputs the pedestrians' walking behavioral response in each footstep [57].

c) Training strategies: For trajectory prediction, mean square error (MSE), also called L2 loss, is commonly used, especially for position representations, as in studies [21]–[24], [32], [35], [49], [56]. For uni-modal distribution representations, the negative log-likelihood loss is used, as in studies [27], [28], [39], [41], [52], [59], [62]. For the multi-modal distributions representations such as GAN-based models, the adversarial loss is used, together with L2 loss to measure the distance between generated samples and the ground-truth, as in studies [31], [40], [44], [50], [51]. Amirian et al. [33] also used information loss in addition to discrimination loss and adversarial loss. Eiffert et al. [58] used adversarial loss with the negative log-likelihood loss for the generator.

B. Intention Prediction

a) Task definition: The intention prediction methods provide high-level information on pedestrian behavior. The intention or action can be predicted in different time horizons. Understanding and predicting the pedestrian intention, especially the crossing intention, is crucial for higher "Society of Automotive Engineers" (SAE) Levels aiming at automated

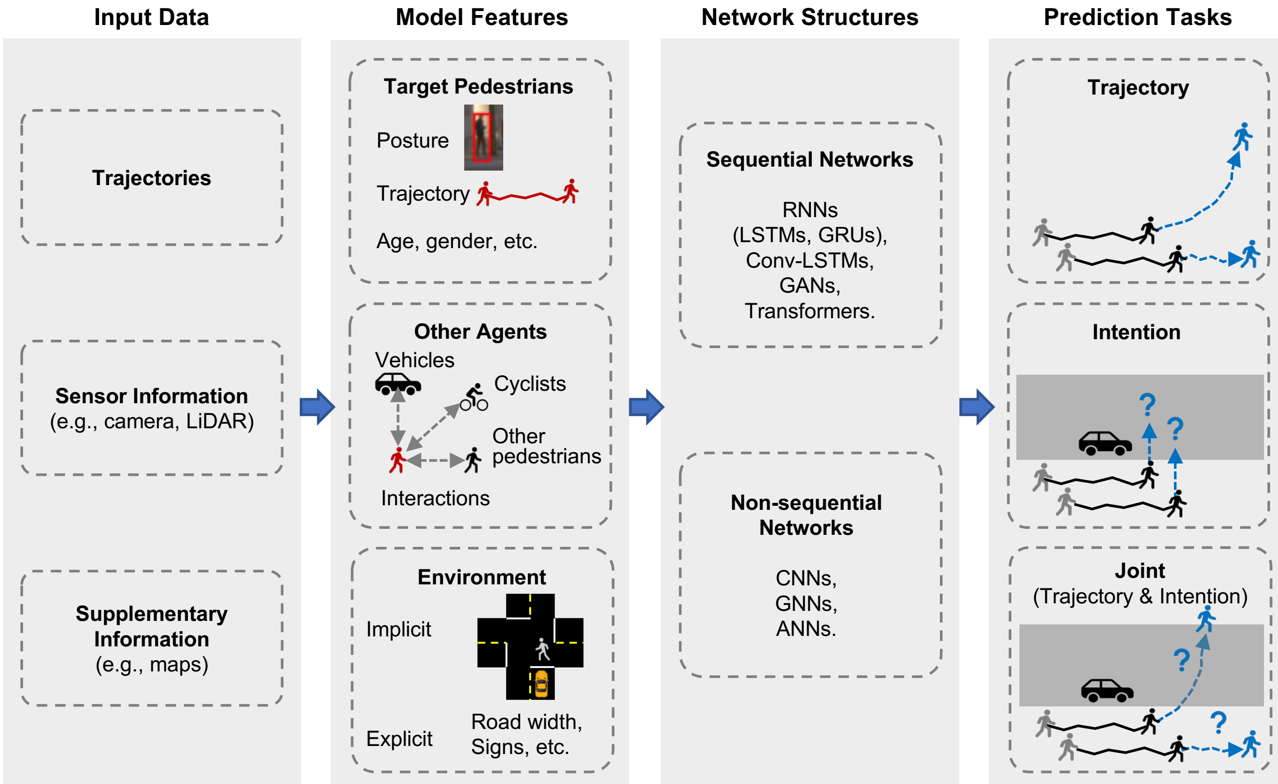


Fig. 2. Four criteria of the pedestrian behavior prediction system for categorizing existing studies. Input data are fed into the model. Model features are the stimuli of network structures. Different network structures are utilized to extract spatial and temporal information, and output different types of prediction tasks.

driving. With the precise prediction of pedestrian intention in advance, automated vehicles can make better decisions and reduce the risk for potentially hazardous situations. Given the observed information of a pedestrian such as trajectories and postures, we aim to predict the intention of a pedestrian. The intention can be defined as discrete behavior types in the future. Many studies use "intention" interchangeably with "actual actions in the future", because labeling the "intention" of a pedestrian is usually a hard problem. Rasouli et al. [84] addressed and labeled intention by asking multiple annotation participants to observe the video of pedestrians and label the crossing intention, and then took the average. In this paper, we do not distinguish the intention and actual action.

b) Output representation: The intention prediction is a classification problem. Many studies treated the problem as a binary classification with crossing or non-crossing (C/NC) action, such as in [7], [64], [70], [75], [76]. Some other studies predicted multi-classification with several different action types. For instance, Fang et al. [69] predicted four types of behaviors including crossing, stopping, bending, and starting, using several binary classifications for multi-classification. Rasouli et al. et al. [73] included four types of behaviors including walking, standing, looking towards the traffic, and not looking. Goldhammer et al. [81] classified pedestrians' motion states into waiting, starting, moving, and stopping. The multi-classification usually includes the whole process of crossing with a certain order, and contains more information.

c) Training strategies: Rasouli et al. [73] used sigmoid cross entropy loss for classification. Besides, many studies used deep learning networks to extract features, and then use other machine learning classification methods. For example, the studies [69], [87] used SVM with hinge loss, and the studies [69], [70] used random forest (RF) for classification.

C. *Joint Prediction*

Pedestrian intention can be predicted jointly with trajectory prediction. There are mainly two kinds of joint prediction frameworks. One kind is that the trajectory and intention prediction tasks share the same feature extracting module. The extracted features are fed into two separate streams for different prediction tasks. For instance, Liang et al. [83] predicted both, the future positions (xy-coordinates) as well as estimating the possibilities of future activity labels simultaneously in one network. The trajectory generator and activity prediction modules share the features extracted from the images. In this framework, the trajectory and intention prediction share the same network, which can save computational resources.

Another kind is that the trajectory and intention are separately predicted, but the information is used to refine each other as suggested by Huang [80]. In works [81], [82], [84], [85], the researchers extracted features for the two tasks separately, and combined the two tasks based on the intention prediction results information to improve the trajectory prediction results.

TABLE I
MODEL FEATURES AND INPUT DATA OF PEDESTRIAN BEHAVIOR PREDICTION

Prediction Tasks	Model Features			Input Data			Papers
	Target Pedestrians	Other Agents	Envir-onment	Traj-ectory	Sensor Data	Supplementary information	
Trajectory (44 papers)	Trajectory	-	-	Yes	-	-	[20]–[26]
	Trajectory	Social interaction	-	Yes	-	-	[27]–[45]
	Trajectory, skeleton cue	Social interaction	-	Yes	Camera images	-	[46]
	Trajectory	-	Implicit	Yes	Camera images	-	[47], [48]
	Trajectory	Social interaction	Implicit	Yes	Camera images	-	[49]–[54]
	Trajectory	Person-ORU interaction	Implicit	Yes	Camera images	-	[55]
	Trajectory	Social interaction; Person-ORU interaction	Implicit	Yes	-	Scene image map	[40]
	Trajectory, motion states, destination	Social interaction	Implicit	Yes	-	Grid-based map, destination	[56]
	Trajectory, motion states, destination	Social interaction	Explicit	Yes	Camera images	Destination	[57]
	Trajectory, category	Social interaction; Person-ORU interaction	-	Yes	-	Agent Category	[58]–[60]
	Trajectory, category, direction	Person-ORU interaction	-	Yes	-	Agent Category	[61]
	Trajectory, velocity, agent shape	Person-ORU interaction	-	Yes	Camera images	Agents' states, traffic concentration	[62]
	Trajectory, appearance cue, VR information	Person-ORU interaction	Implicit	Yes	Camera images	Pedestrian VR information, vehicle's states	[63]
Intention (17 papers)	Motion states	Vehicle factors	Explicit	-	Lidar images	Static map	[64], [65]
	Trajectory	-	-	Yes	Lidar images	-	[66]
	Appearance cue	Vehicle factors	Explicit	-	Camera images	Vehicles' states	[67]
	Skeleton cue, motion states, individual information	Vehicle factors	Explicit	-	Camera images	-	[68]
	Skeleton and/or appearance cue	-	-	-	Camera images	-	[69]–[72]
	Appearance cue	-	Implicit	-	Camera images	-	[73]–[77]
	Appearance cue	Person-ORU interaction	Implicit	-	Camera images	-	[78]
	Speed, age, gender	Vehicle factors	Explicit	-	Lidar images, Camera images	Age, gender, environmental parameters	[7]
	Trajectory, skeleton and appearance cue	Vehicle states	Implicit	Yes	Camera images	Bounding boxes, the speed of ego-vehicle	[79]
Joint (6 papers)	Trajectory	-	-	Yes	-	-	[80]
	Trajectory, motion states	-	-	Yes	-	-	[81]
	Trajectory, skeleton cue, velocity	-	Explicit	Yes	Camera images	-	[82]
	Skeleton and appearance cue	Social interaction; Person-ORU interaction	Implicit	-	Camera images	-	[83]
	Trajectory, appearance cue	Vehicle factors	Implicit	Yes	Camera images	Bounding-boxes	[84], [85]

The combination of the two streams can then utilize more information to get better performance.

D. Summary of Prediction Tasks

The existing works for different prediction tasks are listed in Table I. We notice that there are more papers on trajectory prediction than the other two tasks. The application of different tasks is one of the reasons for this imbalance. The trajectory prediction can be used for many scenarios, not only for the automated vehicles in urban scenarios, but also for the development of social-aware robots in indoor scenarios, while the crossing intention prediction is mainly used for traffic scenarios. Therefore, there were more researchers from different research fields focused on trajectory prediction compared with intention prediction. There are other reasons related to the prediction methods and datasets that are used by these tasks. We discuss them in Sec. VI-C and Sec. VII-C.

IV. INPUT DATA

Previous models used various types of input data. The pre-processed data such as trajectories and raw sensor data such as camera images can be used for training. Besides, other information such as the map and road parameters can be used to provide environment information. In this section, we classify previous studies based on the type of input data that provide different kinds of information.

Existing methods use one or multiple data sources as input to predict pedestrian trajectories:

1) Past trajectories, which can provide information of a pedestrian's motion state. It is used by most trajectory prediction methods.
2) Sensor data, such as the sequences of scene images recorded by the camera, and the point clouds recorded by LiDAR. The sensor data can provide more information of the pedestrian's posture and appearance, as well as provide the environmental context information.
3) Other supplementary information, including the pedestrian information (e.g., age and gender), the vehicle state (e.g., the speed and heading angle), and environment information (e.g., the road information and maps).

The input data of previous studies is shown in Table I, showing that different prediction tasks require different input data. The trajectory prediction requires trajectories as input. The trajectory can be labeled from either camera-recorded videos or LiDAR point cloud videos, or even generated from the simulation. In the studies that only require trajectories, raw sensor data is not required. For those trajectory prediction methods that require sensor data, the camera images and LiDAR point clouds can be used to provide visual behavior information. The intention prediction usually requires raw sensor data, that can provide visual or posture behavior cues for a pedestrian's intention. For joint prediction, both trajectory and raw sensor data can be utilized because this type of task requires both trajectory and visual behavior information. When the model needs the environment or other information, the supplementary information such as maps of the environment, the types of the object, and even virtual reality (VR) information can be required. With different types of input data, different features can be considered for modeling. More details about the model features are presented in Sec. V. As most of the existing studies used publicly available datasets for training and evaluation, we introduce more details about the sensors in Sec. VII-B for models that used raw sensor data.

V. Model Features

In this section, we categorize previous studies based on what features of pedestrian behavior have been considered in the model. Many factors can influence pedestrian behavior. Rasouli and Tsotsos [88] divided the factors that influence pedestrian behavior into pedestrian factors and environmental factors. Kotseruba et al. [85] analyzed the implicit and explicit factors that influence the pedestrians' crossing behavior, including the environment, communication with others, and their own states. Researchers consider one or several of these influencing factors as model features. In this paper, based on the internal and external stimuli of pedestrian behavior defined by Rudenko et al. [16] and the influencing factors mentioned in [85], [88], we divide existing works by three types of model features. Fig. 3 shows the classification of model features, and the number of

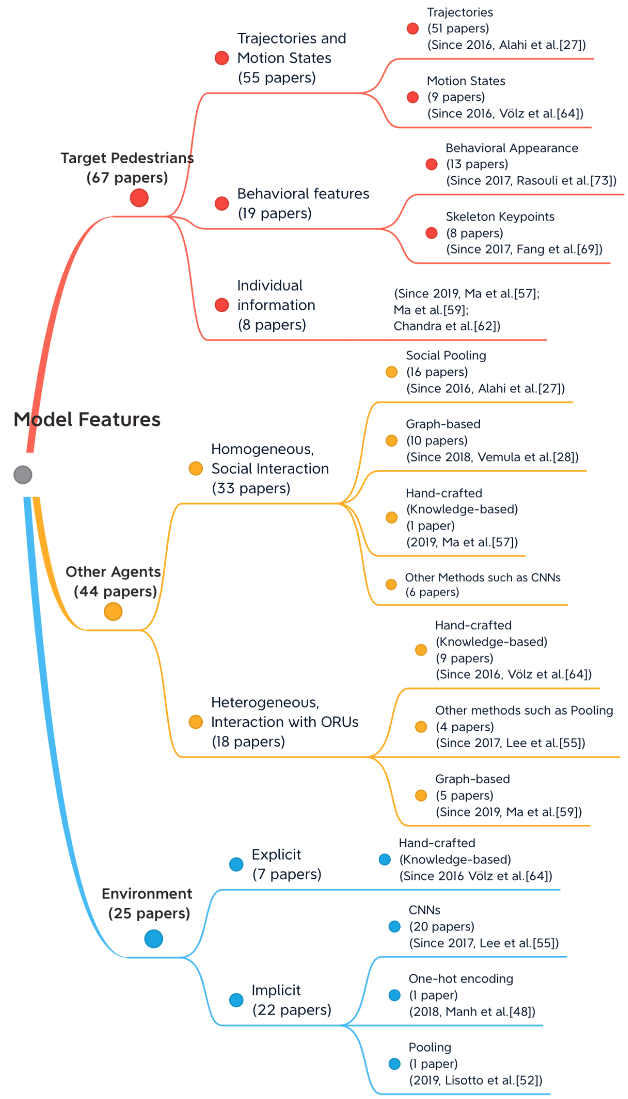


Fig. 3. The classification of the model features. The number of papers that use the corresponding features and the year that firstly used the factors/methods are listed. Please note that a paper can use **multiple** model features.

papers that used the corresponding features. Existing works use one or several combinations of these features:

1) The features related to *target pedestrians*, including trajectories and motion states, behavioral cues such as posture and appearance, as well as individual information such as the age and gender, etc.;
2) The features related to *other agents*, including homogeneous interaction, i.e., the social interactions between pedestrians; and heterogeneous interaction, i.e., the interaction between pedestrians and other road users (ORUs). Note that in this paper, we mean other types of road users except pedestrians when we say "ORUs";
3) The features related to the *environment*, including explicit factors and implicit interactions with context scenes.

TABLE II
INFORMATION OF TARGET PEDESTRIANS USED IN PREDICTION. PLEASE NOTE THAT A PAPER CAN USE **MULTIPLE** MODEL FEATURES.

Target Pedestrian Information		Papers	Summary
Trajectories and Motion States (55 papers)	Trajectories (51 papers)	Trajectories only: Trajectory prediction: [20]–[24], [34] Intention prediction: [66] Joint prediction: [80]	**Advantages:** Contain the historical temporal information. The predicting models are usually simple and require less computing resources. **Drawbacks:** These models have not considered the interaction with other road agents and environment.
		Together with other factors: Trajectory prediction: [25]–[33], [35]–[63] Intention prediction: [79] Joint prediction: [81], [82], [84], [85]	**Advantages:** The models consider other features can get more accurate results. **Drawbacks:** The predicting models are complicated and require more computing resources.
	Motion States (9 papers)	Trajectory prediction: [56], [57], [61], [62] Intention prediction: [7], [64], [65], [68] Joint prediction [81]	**Advantages:** Provide simple but strong information. They are easy to get, and do not require complicated feature extraction and labelling. **Drawbacks:** These properties do not include other implicit information, and only related to current states. They are usually used together with other inputs.
Behavior features (19 papers)	Appearance-based (13 papers)	Trajectory prediction: [63] Intention prediction: [67], [72]–[79] Joint prediction: [83]–[85]	**Advantages:** The images can provide more information than just trajectories, the posture and appearance can reveal the future action. **Drawbacks:** These models require more powerful computing resources.
	Skeleton-based (8 papers)	Trajectory prediction: [46] Intention prediction: [68]–[71], [79] Joint prediction: [82], [83]	
Individual information (8 papers)	Category and type	Trajectory prediction: [58], [59], [61]	**Advantages:** These factors influence the pedestrian behavior and using them as features enables researchers to get accurate results. **Drawbacks:** Can be hard to get. If based on assumptions, it may not be precise.
	Others	Trajectory prediction: [63] Intention prediction: [7], [57], [62], [68]	

A. *Target Pedestrians*

The states of target pedestrians are essential model features for predicting their future behaviors. A summary is listed in Table II.

1) Trajectories and motion states:

a) Trajectories: Most of the trajectory prediction models include the history of pedestrian trajectories, sometimes together with other model features. Trajectory prediction studies [20]–[24] considered only pedestrians' past trajectories for predicting their future trajectories. They extracted the features through embedding layers, and fed the features into deep learning structures for prediction. In addition to only trajectories, studies [25], [26], [44] encoded intermediate destinations from the trajectories and predict future trajectories conditioned on the destinations. For intention prediction, Zhao et al. [66] used trajectories extracted from roadside LiDAR sensors to predict the crossing intention. For joint prediction, Huang et al. [80] used trajectory to predict future intention and trajectory simultaneously, with the predicted results refining each other.

Although the context-based data are good indicators to include, the prediction can be faster by using only the trajectory as input. With recurrent networks, the temporal information of target pedestrians can be extracted from the trajectories, which usually provide rich historical information. There are several advantages of using only trajectories for prediction: It requires less annotation effort than annotating more semantic information on images, and the predicting framework is usually simple and requires less computing resources than those methods which consider the interaction with other road agents and with the environment. The drawbacks are that these methods have not considered the interaction with other road agents and the environment that could also affect the future behavior of pedestrians. The trajectories considered together with other model features usually take the trajectory as part of the input. These models extract the trajectory feature in an individual branch and utilize other compensation information resulting in higher accuracy.

b) Motion states: The motion states such as the velocity and position are also important features for human behavior prediction. For trajectory prediction, Ma et al. [57] focused on a microscopic level instead of estimating the positions at each time-step, and predicted the future trajectories by learning pedestrian's walking behavior at each footstep, considering the velocity and the step length as important inputs. Song et al. [56] also considered velocity as one of the features. Carrasco et al. [61] used orientation to build a graph representation for feature extraction. Chandra et al. [62] used position, velocity, and other factors as model features to define the state space of each road agent. For intention prediction methods, many studies [7], [64], [65], [68] included velocity to decide whether a pedestrian wants to cross the road or not. For the joint prediction, Goldhammer et al. [81] considered pedestrians' trajectory and velocity, as well as their ego-coordinate for prediction.

The motion states can provide simple but strong information about the moving behavior of pedestrians. The velocity and position information is easy to get, and does not require complicated feature extraction and labeling. However, these properties do not include other implicit information, and are only related to states at the current time step. Therefore, the motion states are usually used together with several other inputs as complements.

2) Behavior features: As proposed by Schmidt and Färber [89], using only trajectory information for intention prediction is insufficient. The behavioral features, especially the appearance and posture, usually indicate a pedestrian's intention, and are used by many intention prediction and joint prediction works. The CNNs are usually used to extract visual cue information and/or get the key-point features of pedestrians. The behavioral information from the images can provide more behavior information of pedestrians than just trajectories, but requires more powerful computing resources.

a) Appearance-based: For intention prediction, the appearance behavioral feature can be extracted implicitly from images, usually using CNNs [72], [73], [78]. Three-dimensional CNNs (3D-CNNs) have been utilized to extract spatio-temporal features and recognize pedestrians' crossing intentions in [67], [75], [76]. For the joint prediction, Rasouli et al. [84] and Kotseruba et al. [85] used separate streams for intention estimation, which extracts posture features from the local context and appearance with CNNs.

b) Skeleton-based: The postures of pedestrians are strong behavioral cues that can indicate their intentions. The postures can be represented and estimated by skeleton keypoints using pre-trained CNN-based networks, as in For intention prediction studies [46], [68]–[71], [79] and joint prediction studies [82], [83]. The hourglass network [90] and OpenPose [91] are utilized to extract pose features.

3) Individual information: Category, destination, and agent shape/size, age, gender, and the theory of mind information are considered in many existing papers. The trajectory prediction models that involve multi-agents [58], [59], [61] required the category of the target pedestrians in prediction. Ma et al. [59] and Carrasco et al. [61] used the category and coordinates of the agent as vertex information. Both methods build graph representations of the instances, and consider all types of agents in traffic, that can also be used as pedestrian predictors. In work [58] denotes the vehicle and pedestrian type, and used the information for vehicle-human interaction, which can be explained in detail in the following sections.

Individual information such as age and gender can provide supplementary information for pedestrian behavior prediction, and they are significant factors that influence pedestrian behavior [92], [93]. For intention prediction, age and gender are included as important model features in work [7] to provide necessary human factors-related information. Ma et al. [57] assumed the destination is a vertical line of the crossroad, and used the distance from the destination to the target pedestrian as input features. Chandra et al. [62] also considered the road agent's shape and size as implicit constraints in the trajectory prediction. Kim et al. [63] proposed the multiple stakeholder perspective structure (MSPM) that considered the information not only from the driver's view using sensors mounted on a vehicle, but also included the information from the pedestrian's view using VR devices. These individual factors can influence pedestrian behavior and enable the researchers to get more accurate results using them as model features. However, compared to the trajectories and images, many factors are much harder to get. The destination, age, and gender usually require questionnaires or additional annotation. Otherwise, they can be based on assumptions or output from previous perception modules but may not be precise enough.

B. Other Agents

In this section, we discuss the influence of the other agents on pedestrians' behaviors. The information and interaction with other agents are included by 65% of existing papers that we reviewed. A summary is presented in Table III.

1) Homogeneous - social interaction between pedestrians: According to Moussaid et al. [8], pedestrians' future behavior is not only dependent on their past states, but also driven by social interactions with other pedestrians nearby. Social interaction is an important factor for modeling pedestrians' future trajectories.

a) Hand-crafted features: For trajectory prediction, Ma et al. [57] used hand-crafted features to model the social relationship between pedestrians. They utilized relative positions and relative velocities between the pedestrian and the seven nearest neighbors in front of the target pedestrian as input features. While these hand-crafted features succeeded in this task, they are often hard to generalize to new scenarios. Therefore, deep learning methods are developed to be more powerful structures for extracting social interactions.

b) Social pooling and its variants: Social-LSTM [27] modeled social interactions in a learning-based approach for trajectory prediction. Instead of using knowledge-based methods as in social force [12], the authors proposed a social pooling layer over the hidden states of LSTMs to model the interactions between pedestrians. Several works including [29], [30], [32], [44], [46], [49], [52] followed the social pooling trend and improved the interaction mechanism by attention pooling using various attention mechanisms. Fernando et al. [29] improved the social pooling module with a soft and hard-wired attention mechanism. Xu et al. [30] utilized a weighted spatial affinity function with calculated weights to determine the social interactions over the spatial features. Zhang et al. [32] proposed a state refinement module for future predictions. Sophie [51] assumed that people pay more attention to closer objects and sorted the attention by distance.

Later works [31], [33], [45] improved the interaction module with a more complicated pooling structure. Social-GAN [31] pointed out that local interaction information is not always sufficient, and hence, they use a multi-layer perceptron (MLP) followed by a max-pooling structure to capture the global social interaction information. Amirian et al. [33] improved the interaction module by using an attention pooling that relies on hand-crafted interaction features inspired by neuroscience and biomechanics. Zhang et al. [45] proposed the Social Interaction Extractor to learn interaction weights with a sub-network structure. Kothari et al. [39] categorized the existing interaction module into grid-based methods and non-grid-based methods, and proposed a grid-based directional pooling method and the DirectConcat method that achieved improvement. Bhujel et al. [53] calculated the social attention from the hidden state with designed physical and social attention functions. Col-GAN [36] proposed an attention module that used MLPs to learn the interaction and used a weighted sum to calculate the interaction feature.

The social pooling module enables the existing work to consider social interaction. The structure is simpler than the graph-based models with fewer parameters to learn.

TABLE III
INFORMATION OF OTHER ROAD USERS (ORUS) USED IN PEDESTRIAN BEHAVIOR PREDICTION

Other Agent Information		Papers	Summary
Homogeneous - Social Interaction Between Pedestrians (33 papers)	Hand-crafted features (1 paper)	Trajectory prediction: [57]	**Advantages:** Explainable. **Disadvantages:** Hard to generalize to new scenarios.
	Social pooling and its variant (16 papers)	Trajectory prediction: [27], [29]–[33], [36], [39], [44]–[46], [49], [51]–[53], [56]	**Advantages:** The social pooling module considers social interaction in the model. It is relatively simple compare with graph based models. **Disadvantages:** They mainly deal with symmetric interactions.
	Graph-based representation (10 papers)	Trajectory prediction: [28], [34], [35], [37], [41] [40], [50], [54], [58], [59]	**Advantages:** They can extract non-symmetric interactions. **Disadvantages:** The construction of the graph is computational- and time-consuming.
	Other methods (6 papers)	Trajectory prediction: [38], [42], [43], [56], [60] Joint prediction: [83]	**Comments:** The social norm is considered using sampling methods. Agent-aware attention and LSTMs are used to model social and time dimensions simultaneously. CNNs are applied on grid-based map.
Heterogeneous - Interaction with Other Road Users (18 papers)	Hand-crafted features (9 papers)	With single vehicle: Trajectory prediction: [63] Intention prediction: [7], [64], [65], [67], [79] Joint prediction: [84], [85] Vehicle volume: Intention prediction: [68]	**Advantages:** Explainable. **Disadvantages:** Can be hard to generalize to new scenarios.
	Graph-based representation (5 papers)	Trajectory prediction: [40], [58], [59], [61], [78]	**Comments:** The graph based module can extract non-symmetric interactions between pedestrians and other road users.
	Other methods (4 papers)	Trajectory prediction: [55], [60], [62] Joint prediction: [83]	**Comments:** Grid-based pooling, CNNs, One-hot coding, and reinforcement learning can be combined into the network.

c) Graph-based representation: The symmetric pooling (max or average pooling) operation assumes that the interactions between pedestrians are symmetric, which, however, is not always the case. To extract non-symmetric interactions, researchers use a graph to represent the relationship between pedestrians. In such graphs, the vertices represent the states of the pedestrians and the edges represent the spatial or temporal relationships between pedestrians.

For trajectory prediction, Vemula et al. [28] represented the social attention by a spatio-temporal graph representation, using soft attention with calculated weights over hidden states of each node. Zhang et al. [34] used the graph representation and applied the social graph network directly on MLP embedded features from agents' locations and velocity status. STGAT [35] and Social-BiGAT [50] applied the graph attention networks (GAT) as proposed by Veličković et al. [94] to extract the social interactions between pedestrians over the hidden states of LSTMs. STGAT [35] calculates the relationship for each time step to get the state of pedestrians, while Social-BiGAT [50] calculates the interaction after extracting the hidden states from all observed time steps. Hu et al. [40] proposed an interaction branch with a graph structure, namely neural motion message passing (NMMP), which calculates k times the interacted actor embedding with graph neural network on the hidden states of each agent. Yu et al. [37] exploited a spatio-temporal graph transformer (STAR) to model the spatio-temporal interaction between pedestrians. Social-STGCNN [41] and STGT [54] used graph convolutional networks (GCNs) [95], which are defined as convolution operations over graphs to extract the spatio-temporal social interaction feature.

The graph-based module can extract non-symmetric interactions and get better results than the pooling structures, but the instruction of the graph takes more computational resources, and hence, can be more time-consuming.

d) Other methods: For the trajectory prediction, Social-NCE [38] considers unfavorable events like discomfort and collision situations when learning socially aware motion representations. The authors proposed a safety-driven sampling method, called the multi-agent contrastive sampling, to select negative samples from the neighborhood of other agents in the future. Yuan et al. [42] proposed AgentFormer that can simultaneously model the time and social dimensions using an agent-aware attention mechanism. Tra2Tra [43] proposed a spatial-temporal attention module, that embedded the spatial feature from the coordinates of all pedestrians, and used an LSTM network to extract the temporal dependency between spatial features. Song et al. [56] considered the target pedestrian's neighbors by considering their neighbors' speed. The speed is filled in cells of a grid-based map, and CNNs are used to extract the spatial relationship with the neighbors.

2) Heterogeneous - interaction with other road users (ORUs): The future behavior of pedestrians is influenced by the interaction with ORUs such as vehicles according to Shirazi et al. [9].

a) Hand-crafted features: In Schmidt and Färber's research [89], parameters such as the distance and velocity of the vehicles can influence the crossing intention. For the intention prediction, many researchers used hand-crafted features as inputs, such as in studies [7], [64], [67], [79], including vehicle's velocity or speed, relative velocity and distance between the pedestrian and vehicle, or time to collision (TTC). Zhang et al. [68] used vehicle-related information for crossing intention prediction, including vehicle volume, the green light time for vehicles, and the number of vehicles. For the joint prediction, Rasouli et al. [84] and Kotseruba et al. [85] utilized the ego-vehicle information including the speed and heading angle as complementary inputs.

b) Graph-based representation: As the interaction between different types of traffic agents is usually non-symmetric, graph-based methods can model heterogeneous interactions. For the trajectory prediction, Eiffert et al. [58] proposed a graph vehicle-pedestrian attention network (GVAT) to include both human-human interactions and human-vehicle interactions. Ma et al. [59] used a 4-dimensional graph that consists of the instance layer and the category layer to represent the traffic sequence and to calculate their interaction. The instance layer represents the individual interaction, while the category layer ensures the motion pattern of different categories. Hu et al.'s framework [40] can jointly predict the trajectory of pedestrians and vehicles by the proposed NMMP module with a graph representation. Carrasco et al. [61] built the graph with coordinates, categories, headings as vertices, and exploited the graph attention layer to include the interaction. For the intention prediction, Liu et al. [78] captured the interaction between the pedestrians and other road users using graph convolution to include both spatial and temporal context.

c) Other methods: For trajectory prediction, Lee et al. [55] modeled the interaction for multi-agents with a spatial grid-based pooling layer, which is similar to the social-pooling layer. Chandra et al. [62] took sequences of images as input to predict the trajectories of heterogeneous traffic agents including pedestrians, using CNNs for extracting the appearance and behavioral information of different road agents. Li et al. [60] considered the existence of a vehicle, and combined reinforcement learning into the prediction. For joint prediction, Liang et al. [83] modeled the interaction between pedestrians and other road users in the scene by explicitly modeling the geometric relation with a knowledge-based function defined by the authors that considered the geometric distance and the box size, and modeled the object type using one-hot encoding.

C. Environment

The interaction with the environment also influences pedestrians' behaviors. The environmental information is included by 36% of the existing papers that we reviewed. To include the interactions with the environment scene as model features, studies either took explicitly defined environment features as inputs, or use sequences of camera images or a navigation map to learn the pedestrians' interaction with the surrounding environment implicitly. In this section, we present how the researchers address pedestrian-environment interactions. The summary is shown in Table IV.

1) Explicit features: Explicit features are manually defined and usually explainable. Many researchers utilize information about zebra crosswalks and the curbs. For trajectory prediction, Ma et al. [57] used the distance to the left and right boundaries of the crossroad as input features for the prediction. For intention prediction, the distance between the vehicle and the crosswalk, the distance between the pedestrian and the crosswalk, and the distance between the pedestrian and the curb are important factors and were used by Völz et al. [64] and Zhang et al. [7] as model features. Yang et al. [67] considered the existence of stop signs, zebra markings, and traffic lights in local traffic scenes. They used the prior weight to represent different scenes. In addition to the geometry-related environment features, Zhang et al. [68] used temperature as an important factor to predict the crossing intention at red-light. For joint prediction, Wu et al. [82] used the crossable information at a crossroad to change the sampling weight when predicting the trajectory.

2) Implicit features: The implicit features are not explicitly defined and are usually extracted from images or semantic maps.

a) CNN-based feature extractor: As CNNs are capable of extracting image features, pre-trained CNNs can be used to extract appearance features and implicit human-scene interaction features from a sequence of images. The trajectory prediction works [47], [49]–[51], [54], [55] followed this direction using pre-trained CNNs. Bhujel et al. [53] utilized CNN features and a physical attention function to learn the probability that a location is the right place to focus for predicting the next position. Instead of sensor image data, Hu et al. [40] used a 2D bird's-eye-view scene image map as input to provide prior knowledge about the traffic condition and rules, and extracted the environment information with the CNN structure [96] to extract scene embedding. Song et al. [56] used a grid-based map with occupied cells to indicate the fixed obstacles in the scenes using CNNs to extract the environmental features. For intention prediction, Rasouli et al. [73], Hoy et al. [74], and Kotseruba et al. [79] used CNNs to extract visual context features implicitly. Liu et al. [78] segmented the images into pedestrians and objects with binary masks using a segmentation model [97]. Then, they captured the context feature by encoding the segmented binary masks with the ResNet backbone. Works [75], [76] utilized 3D-convolutional networks for image feature extraction in the observed time period. For joint prediction, Liang et al. [83] used a pre-trained scene segmentation model [98] for environmental feature extraction. The integer scene semantic features are transformed into binary masks, then two convolutional layers are applied to the mask features to get CNN features. Rasouli et al. [84] and Kotseruba et al. [85] used CNNs to extract the local visual context around the pedestrian with a bounding box implicitly along with the appearance feature.

b) Other methods: For trajectory prediction, Scene-LSTM [48] takes the scene information into consideration by using grid cells to represent the input scene image. The calculated hidden states of each grid cell are used as input to a scene data filter to pass the scene constraints information to get better trajectory prediction results. Lisotto et al. [52] utilized a semantic map and the navigation map, and applied semantic and navigation pooling to extract the environmental interaction feature. The semantic map, which contains the scene context, is generated from the image using semantic segmentation, and the navigation map which embodies the most frequently crossed areas is generated from the observed data by counting the crossing frequency of squared patches.

D. Summary of Model Features

Model features play important roles in pedestrian behavior prediction. As we summarized in Fig. 3, a method can use

TABLE IV
INFORMATION OF ENVIRONMENT USED IN PEDESTRIAN BEHAVIOR PREDICTION

Environment	Descriptions	Papers	Summary
Explicit (7 papers)	Hand-crafted features	Trajectory prediction: [57] Intention prediction: [7], [64], [65], [67], [68] Joint prediction: [82]	**Comments:** They are manually defined, simple and usually explainable.
Implicit (22 papers)	CNN based feature extractor (20 papers)	Trajectory prediction: [40], [47], [49]–[51], [53]–[56] Intention prediction: [63], [73]–[79] Joint prediction: [83]–[85]	**Comments:** CNNs are capable of extracting image features, and can be used to extract the interaction between pedestrians and the environment implicitly.
	Others (2 papers)	Trajectory prediction: [48], [52]	**Comments:** One-hot encoding and pooling can be used to encode the location information. But when encoding location information with one-hot vectors, the dimensionality might become very high.

multiple model features. For the target pedestrians, the trend is also to include more information. In 2016, the trajectories and motion states are included [27]. In 2017, the behavioral features are included [69], [73], and in 2019, the individual information are added [57], [59], [62]. For the interaction with other agents, the social interactions are included mainly in trajectory prediction. The social pooling methods was proposed in 2016 [27], and the graph-based model was proposed in 2018 [28]. In 2019, Ma et al. [57] added knowledge-based information to model the interaction. The interaction with other road users such as vehicles is included mainly in the intention prediction works. In 2016, researchers started to use hand-crafted features to model the interaction [64]. in 2017, the learning-based feature extractor such as pooling method [55] and graph-based methods [59] are proposed. For the environment feature, researchers first model it with hand-crafted features explicitly in 2016 [64], then used a CNN-based model to learn it in 2017 [55]. In 2018 and 2019, other attempts on one-hot encoding [48] and pooling [52] are tried by researchers. The hand-crafted features used in existing works are explainable but hard to generalize, while the learning-based features have achieved more accurate results but are difficult to explain. Future works can focus on how to combine these features.

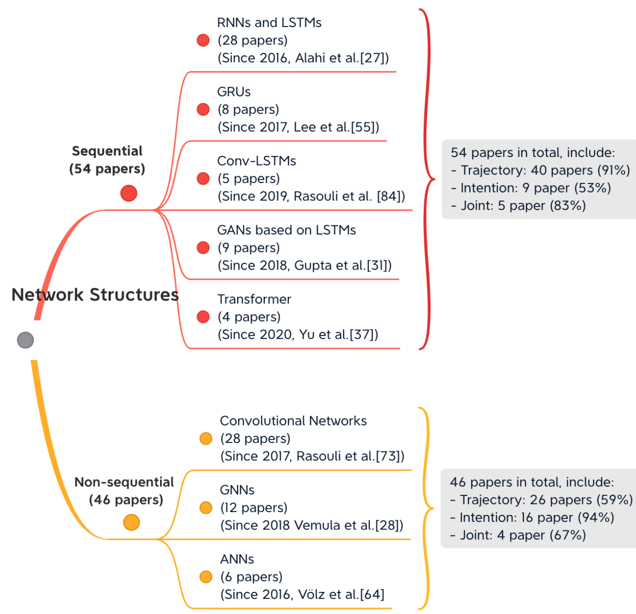


Fig. 4. The classification of the network structures. The number of papers that used corresponding methods and the year that firstly used the network structures are listed. Please note that a paper can use **multiple** network structures. For example, a model can use CNNs for extracting the visual information, and use LSTMs for the temporal prediction. The distribution of the papers is summarized in the boxes on the right side.

VI. NETWORK STRUCTURES

In this section, we list commonly used network structures, and classify them into sequential networks and non-sequential networks. These structures can be combined to form a prediction model. For instance, a model can use CNNs for extracting visual information, and use LSTMs for temporal prediction. Fig. 4 shows the classification of the network structures. Table V presents the summary of the network structures used by existing research.

A. Sequential Networks

The sequential networks typically deal with time-series information by assuming the moving state at one time step is conditionally dependent on previous states. Traditional models used for predicting the pedestrian's future action such as hidden Markov models (HMM) [99], [100], partially observable Markov decision processes (POMDP) [101], and Gaussian processes [5], [102], [103] require accurate and precise segmentation and tracking of pedestrians. However, this is challenging due to the difficulty of extracting reliable image features as outlined by Völz et al. [64]. With the help of deep learning, the models are able to extract features from images with CNNs and to extend the long-term memory with Recurrent neural networks (RNNs) including long short-term memory (LSTMs) and gate recurrent units (GRUs), convolutional LSTMs (Conv-LSTMs), and transformer networks (TFs) to overcome the limitation of traditional models.

1) Recurrent neural networks (RNNs) and long short-term memory (LSTMs): RNNs and their improved version, LSTMs are preferred by many researchers because of their strong ability to handle the trajectory sequence information. For trajectory prediction, Vemula et al. [28] used spatio-temporal graph within the RNN structure. Alahi et al. [27] utilized LSTMs to learn the motion state of a pedestrian and proposed Social-LSTM model to predict a pedestrian's trajectory. Later trajectory prediction methods such as [20]–[22], [29], [30], [32], [34],

TABLE V
NETWORK STRUCTURES FOR PEDESTRIAN BEHAVIOR PREDICTION. PLEASE NOTE THAT A PAPER CAN USE **MULTIPLE** NETWORK STRUCTURES.

Network Structures (Earliest Used Time)		Papers	Summary
Sequential Networks (54 papers)	RNNs and LSTMs (Since 2016, Alahi et al. [27])	Trajectory: RNNs: [28]; LSTMs: [20]–[22], [27], [29], [30], [32], [34], [35], [38], [39], [43], [48], [49], [52], [53], [59], [62], [63], [84] Intention: LSTMs: [7], [71], [72], [77] Joint: LSTMs: [80], [82], [83]	**Advantages:** RNNs (including LSTMs, GRUs) are more capable to handle long term prediction than the traditional models. **Disadvantages:** They cannot be parallelized, and cannot handle too long sequences.
	GRUs (Since 2017, Lee et al. [55])	Trajectory: [25], [26], [47], [55] Intention: [74], [78], [79] Joint: [85]	
	Conv-LSTMs (Since 2019, Rasouli et al. [84])	Trajectory: [56], [63] Intention: [75], [76] Joint: [84]	**Advantages:** Conv-LSTMs can extract spatial and temporal features simultaneously. **Disadvantages:** The computational cost is higher than for LSTMs.
	GANs (Since 2018, Gupta et al. [31])	Trajectory: [31], [33], [36], [40], [46], [50], [51], [58], [60]	**Advantages:** The GANs as generative models can predict multiple plausible trajectories. **Disadvantages:** Hard to train, and requires techniques for convergence.
	Transformers (Since 2020, Yu et al. [37])	Trajectory: [24], [37], [42], [54]	**Advantages:** They can handle long sequences and allow parallelization. **Disadvantages:** Implemented with a fixed-length, not flexible enough.
Non-sequential Networks (46 papers)	Convolutional Networks (Since 2017, Rasouli et al. [73])	Trajectory: [23], [40], [41], [45], [47], [49]–[51], [53]–[56], [61] Intention: [46], [67]–[73], [75], [76], [78], [79] Joint: [83]–[85]	**Comments:** CNNs can be used for both extracting spatial features and sequential features. For the sequential prediction, as there is not dependency of the previous time steps, the prediction error do not accumulate like the RNNs, and it allows parallel computation.
	GNNs (Since 2018, Vemula et al. [28]	Trajectory: [28], [34], [35], [37], [40], [41], [50], [54], [58], [59], [61] Intention: [78]	**Comments:** GNNs can be used for extracting non-symmetric interactions and capturing spatio-temporal features.
	Other ANNs (Since 2016, Völz et al. [64])	Trajectory: [44], [57] Intention: [64]–[66] Joint: [81]	**Advantages:** Structures are simple; can handle the non-linearity. **Disadvantages:** For 2D image input, ANNs will lose the spatial information, and require a huge amount of trainable parameters. For sequential input, ANNs cannot capture sequential information.

[35], [39], [43], [48], [49], [52], [53], [59], [62] followed this trend of using LSTM-based methods to cope with time-series information.

For intention prediction, Zhang et al. [7] used LSTMs with the attention mechanism for prediction that outperforms the SVM model. Pop et al. [77] proposed a multi-task network that combines the CNNs for extracting visual features and the LSTM network for estimating the time to cross the street. The FuSSI-Net proposed by Piccoli et al. [71] used a CNN-based network for detection and skeleton keypoints extraction, and then used LSTMs to extract temporal information. For joint prediction, Huang et al. [80] proposed warp LSTM to deal with neighboring time steps in place of global positions and to allow for long-term trajectory prediction. They proposed the mutable intention filter to generate potential intentions, and then predicted the intention-aware trajectories. Lorenzo et al. [72] employed CNNs to extract pedestrians' behavioral features and applied various RNNs including LSTMs, GRUs, and the bidirectional variants of LSTMs and GRUs for crossing probability prediction. Kim et al. [63] proposed the MSPM model, that includes a driver perspective network and a pedestrian perspective network. The driver perspective network used LSTMs to encode the speed and trajectory information of the driver's perspective and other structures for image feature extraction, and used LSTMs to predict a pedestrian's behavior.

For joint prediction, Liang et al. [83] extracted the feature with CNNs, and then the extracted features are fed into a trajectory generator and activity predictor separately. In the trajectory generator, LSTMs are used for sequence prediction, while in the activity predictor, two separate convolution layers are used on a multi-scale Manhattan Grid for classification and regression to predict the label and location. Wu et al. [82] first extracted skeleton features with CNN-based methods and then used LSTMs to predict behavior classes (i.e. standing, walking, running), and used the dynamic Bayesian network to identify crossing intention. The predicted intention information is used for deciding the weights for trajectory sampling to improve the results. Rasouli et al. [84] used LSTMs in the pedestrian trajectory and vehicle speed prediction stream, and used LSTMs together with other structures in the intention estimation stream.

2) Gate recurrent units (GRUs): GRUs are another improved version of RNNs that are also popularly used for sequential prediction. For intention prediction, Hoy et al. [74] explored a variant of variational recurrent neural networks (VRNNs), namely the deep variational Bayes filters [104] for extracting tracking features, using GRU layers in VRNN cells with CNNs' extracted visual features as inputs. Liu et al. [78] used GRUs for behavior prediction after using a CNN-based segmentation model [97] for appearance feature encoding. Kotseruba et al.'s later work [79] used 3D-CNN for local visual context extraction and used GRUs for non-visual features encoding from bounding boxes, poses, and ego-vehicle speed. For joint prediction, Kotseruba et al. [85] employed GRUs for trajectory prediction, connected with the intention feature extracted from images using CNNs, and fed into a fully connected layer for future action classification.

GRUs can be combined with generative models for pedestrian trajectory prediction. These multi-modal models can provide multiple feasible results by incorporating prior knowledge into pedestrian behavior learning. Recently, conditional variational autoencoders (CVAEs) with sequential encoders and decoders have been adopted to predict multi-modal distributions. The BiTraP [25], SGNet [26], CGNS [47] and DESIRE [55] used GRU encoder-decoders based on CVAE method for trajectory prediction with multi-modal goal estimation. Social-NCE [38] applied LSTM model based on the noise-contrastive estimation (NCE) methods [105] by introducing a social contrastive loss, namely the InfoNCE loss [106].

RNNs and their variants including LSTMs and GRUs use hidden states to represent the time-varying motion properties. They are more capable of dealing with long-term prediction than traditional models because of their capability of learning the dependencies between temporally correlated data. However, the sequential computation of RNN-based models inhibits parallelization. Besides, the networks cannot do well for long sequences because the "temporal distance" between two sample positions is linear, and the network tends to "forget" the information of the previous sample in the sequence. Furthermore, it is hard to explain the physical meaning of the hidden states that represent the moving states.

3) Convolutional LSTMs (Conv-LSTMs): Conv-LSTMs as proposed by Shi et al. [107] have been used to extract spatial and temporal information. For trajectory prediction, Kim et al. [63] used CNNs, Conv-LSTMs, and LSTMs for encoding image information in the driver perspective network. Song et al. [56] used a grid-based map with social and scene information filled in the cells, and used deep conv-LSTM to predict the future trajectories. For intention prediction, Gujjar et al. [76] and Chaabane et al. [75] used 3D-CNN layers as the encoder and conv-LSTM layers as the decoder in their encoder-decoder structure. For joint prediction, Rasouli et al. [84] proposed the PIE model that used LSTMs, CNNs, and conv-LSTMs for prediction. In the intention estimation stream, CNNs are used for appearance behavioral feature extraction with conv-LSTMs as the encoder, and LSTMs as the decoder.

4) Generative adversarial networks (GANs) based on LSTMs: The previously mentioned models follow a uni-modal distribution. As there could be multiple socially acceptable trajectories, Gupta et al. [31] proposed Social-GAN, which assumed that the pedestrian trajectories follow a multi-modal distribution, which means that multiple future trajectories are potentially plausible. They utilized the GANs with an LSTM-based generator for trajectory prediction. Social-BiGAT [50] and studies [33], [40], [46], [51], [58] followed this trend and used LSTMs as generators of the GANs, with various structures of extracting the interactions with other objects. Li et al. [60] utilized Social-GAN and combined it with reinforcement learning in their prediction. The Col-GAN [36] used a GAN structure with LSTM encoder-decoder as the generator. But instead of using an LSTM-based discriminator like Social-GAN [31] and Sophie [51], they used CNNs as the discriminator and classify the segments of a trajectory are real or fake.

The GANs can predict multiple plausible and socially acceptable trajectories given a partial history instead of predicting only one "average behavior". The drawback of the GANs is that they are usually hard to train and require techniques to make the model converge.

5) Transformer networks (TFs): The TFs [108] can alleviate the previously mentioned problems of RNN-based models. The TFs used the attention mechanism to help memorize the information in long sequences. The attention mechanism can create shortcuts between the context vector and the entire source input instead of only the last hidden state. TFs made ground-breaking progress recently in the Natural Language Processing domain and are becoming popular to be adopted for predicting pedestrian behaviors because of their capability of long-term prediction. Giuliari et al. [24] adopted both, the original TF and bidirection transformer (BERT) for trajectory prediction. The authors considered only the individual trajectory as model features yet still gained better performance than previous LSTM- and CNN-based methods. Yu et al. [37] further considered social interaction using graph-based representation to achieve more accurate results. The AgentFormer [42] applied the agent-aware transformer in a multi-agent trajectory prediction framework based on CVAE and modeled the future trajectory distribution conditioned on past trajectories and contextual information. Syed et al. [54] proposed the STGT model that used a CNN model (PSP-Net [109]) for segmentation and extracting the image environmental features, and the transformer is used for sequence prediction.

The TFs avoid recursion and allow parallel computation to reduce training time. With the attention mechanism, the TFs get more accurate results than RNNs. However, the transformers are implemented with a fixed length, and cannot model dependencies that are longer than the fixed length. Some other improvement versions of TFs such as the TransformerXL [110] and the compressive Transformer [111] could be used in the trajectory prediction or other sequence prediction tasks.

B. Non-sequential Networks

The non-sequential networks are used to extract spatial and interaction features. Besides, they can also model the temporal information by directly modeling the final state or distribution over the entire history of observed states without the assumption of conditional dependency on previous states.

1) Convolutional networks: CNNs are used in many models to extract implicit appearance features from images as discussed in Sec. V. Trajectory prediction studies used pre-trained CNNs to extract implicit features of the environment as in [40], [47], [49]–[51], [53]–[56]. Intention prediction studies used CNNs to extract appearance behavioral features ad in [72], [73], [78], [79], and skeleton behavioral features as in [46], [68]–[71], [79]. 3D-CNNs are used to extract spatio-temporal features as in [67], [75], [76]. For joint prediction, CNNs are used to extract posture features as in [84], [85] and environment features as in [83].

In addition to extracting spatial features from images, CNNs can also be used to extract sequential features for pedestrian trajectory prediction. Many methods use hidden states of LSTMs to represent the pedestrian motion states. However, Nikhil and Morris [23] pointed out that trajectories are continuous in nature and do not have a complicated "state". The feature extraction of hidden states in previous models is indirect and the physical meaning of hidden states is difficult to interpret. Bai et al. [112] noticed that recurrent architectures have limitations in inefficient parameters and the training can be inefficient. Therefore, instead of using LSTMs, Nikhil and Morris [23] proposed an algorithm using CNNs to predict the trajectories for computational efficiency, which yields competitive results with a faster speed. Lea et al. [113], [114] proposed temporal convolutional networks (TCNs) that dealt with time series and extracted features by convolutional layers on the temporal dimension. Mohamed et al. [41] proposed the Social-STGCNN model, which reached faster speed and better results on trajectory prediction, by using TCNs to extract spatio-temporal features from the spatial and social interaction features, and utilized CNNs as an extrapolator on the time dimension. Zhang et al. [45] proposed the Social-IWSTCNN, which followed the trend of using CNNs and TCNs for prediction. The convolutional-based methods enable parallelization and without the dependencies on the previous time step, the prediction can be faster and the prediction errors do not accumulate like with RNNs.

2) Graph neural networks (GNNs): GNNs are neural networks over graph-represented data. GNNs have achieved significant success in human action recognition [115]–[117]. GNNs can be used in pedestrian behavior prediction for extracting spatial and temporal interaction between pedestrians and other objects and are especially suitable for modeling non-symmetric interactions and spatio-temporal features as mentioned in Sec. V.

Graph convolutional networks (GCNs) proposed by Kipf and Welling [95] define the convolution operations over graphs. Social-STGCNN [41], STGT [54] use GCNs to extract the spatio-temporal social interaction features for trajectory prediction. Liu et al.'s [78] used GCNs captured the interaction between pedestrians and other road users using graph convolution to include both spatial and temporal context. In particular, the graph attention networks (GATs) as proposed by Veličković et al. [94] improved weighted message passing between nodes and are applied by STGAT [35], Social-BiGAT [50] and studies [58], [61]. Yu et al. [37] improved GAT by applying a transformer boosted attention mechanism and proposed spatio-temporal graph transformer (STAR) model. These methods model the interaction not only based on the current frame but also consider the influence of other time steps. Besides, commonly used network structures can be applied to graph representations. For trajectory prediction, Hu et al. [40] proposed a neural motion message passing (NMMP) structure, which used MLP embeddings to pass messages between nodes and edges. Zhang et al. [34] proposed the social graph network that applied a one-layer MLP on egde and nodes of a graph. Vemula et al. [28] applied structural RNN [118] on edges and nodes of spatio-temporal graphs to model the spatio-temporal interaction between pedestrians. Ma et al. [59] applied LSTM on the nodes of a 4-dimensional graph to model the interaction of different instances and categories.

3) Other artificial neural networks (ANNs): For the trajectory prediction, Ma et al. [57] used an ANN with hidden layers to model the mechanism of decision-making that employed human experience to make the approach more realistic for the prediction of microscopic pedestrian walking behavior. For the intention prediction, Völz et al. [64] designed a dense neural network using 15 hand-crafted features over five time steps, and the dense network outperformed the LSTM and SVM methods. Zhao et al. [66] compared the intention prediction with Naive Bayes methods using trajectories as input, and claimed the results of ANN is worse than the Bayes methods. This may be because they only include the trajectories as inputs, which is too simple to demonstrate the power of neural networks, and other networks such as RNNs can be used for sequence prediction and CNNs can be used for image inputs. CVAEs can also be combined with ANNs. PCENet [44] considered the intermediate stochastic destinations of the pedestrians into prediction by using an endpoint CVAE, where the prediction is conditioned on the features extracted from the past encoder using MLPs. For joint prediction, Goldhammer et al. [81] proposed the PolyMLP model that uses an MLP network to predict polynomial approximation of time series.

The structures of ANNs are simple and can handle non-linearity. ANNs can be used for multiple tasks when the number of input features is small, especially for the intention prediction with hand-crafted features. However, for a 2D image that is a common kind of input in pedestrian behavior prediction, ANNs will lose the spatial information because of squeezing the image into a 1D vector, and can require a huge amount of trainable parameters, where CNNs could be the better choice because they share weights and can keep the spatial information. Besides, ANNs cannot capture sequential information in the input data, where RNNs could handle better.

C. Summary of Network Structures

From the distribution of the papers in Fig. 4, we see that sequential methods are mainly used for trajectory prediction. This is because trajectory prediction requires time series information. Trajectory prediction also employed GNNs for extracting interactions with other road users. The intention prediction usually used non-sequential networks, because they usually need the visual behavior features, which are extracted by CNNs. The joint prediction used both sequential networks and non-sequential networks, as they needed both spatial and temporal information.

The prediction methods also influenced the development of different prediction tasks. For the sequential methods that are commonly used by the trajectory prediction, research followed the trend from LSTMs in 2016 [27], GRUs in 2017 [55], to GANs in 2018 [31] and Conv-LSTMs in 2019 [84], and to the recently used Transformers in 2020 [37]. Each time the development of sequential methods stimulated the research on trajectory prediction. In contrast, for the intention prediction, most works used non-sequential. These models rely on the CNNs to process the images, which usually require more computing resources. This influences the development of the intention prediction. In future work, we need to investigate how much effort we should put into intention prediction. We need to trade off the additional gain from adding intention information for the application domain (e.g., for increased safety in an operational design domain for an autonomously driving vehicle) and the cost of increased computing resources and accuracy and reliability for the perception system.

VII. Evaluation and Datasets

In this section, we firstly present the evaluation metrics that are commonly used for pedestrian behavior prediction. Then, we provide a review of the most commonly used datasets. There are some benchmarks for the trajectory prediction [39], [119] and intention prediction [79] that evaluated parts of the existing works.

A. Evaluation Metrics

1) Trajectory prediction: The evaluation metrics for trajectory prediction are listed below.

- The average displacement error (ADE) (or the mean squared error (MSE)): The average distance between ground-truth and prediction trajectories over all predicted time steps, as defined below, where the predicted position for i^{th} pedestrian at time-step t is $\hat{Y}_t^i = (\hat{x}_t^i, \hat{y}_t^i)$, and the ground-truth is Y_t^i, $i \in \{1, ..., n\}, T_{obs} + 1 \leq t \leq T_{pred}$.

$$ADE = \frac{\sum_{i \in n} \sum_{t=T_{obs}+1}^{T_{pred}} \|Y_t^i - \hat{Y}_t^i\|_2}{n \times (T_{pred} - T_{obs})} \quad (1)$$

- The final displacement error (FDE): The average distance between ground-truth and prediction trajectories for the final predicted time-step, as defined below:

$$FDE = \frac{\sum_{i \in n} \|X_t^i - \hat{X}_t^i\|_2}{n}, t = T_{pred} \quad (2)$$

Some other evaluation metrics such as the collision rate and negative log-likelihood are mentioned in the TrajNet++ benchmark [39]. The average non-linear displacement error is also used by some papers [27], [29], [30], [48], which is the MSE at the non-linear regions of a trajectory.

2) Intention prediction: The evaluation metrics for intention prediction are listed below, with the number of positives P, negatives N, true positives TP, true negatives TN, false positives FP, and false negatives FN.

- Accuracy (ACC): $ACC = (TP + TN)/(P + N)$
- F1 score (F_1): $F_1 = 2TP/(2TP + FP + FN)$
- Precision: $Precision = TP/(TP + FP)$
- Recall (True Positive Rate): $Recall = TP/(TP + FN)$
- Average precision (AP): $AP = \sum_{k=1}^{n} (P(k)\Delta r(k))$. AP is defined as the area under the precision-recall curve, where k is the rank in the sequence of retrieved documents, n is the number of retrieved documents, $P(k)$ is the precision at cut-off k in the list, and $\Delta r(k)$ is the change in recall from items $k - 1$ to k.

3) Joint prediction: For the joint prediction, the intention and trajectory results can be evaluated separately.

B. Datasets

High-quality and large-scale datasets are crucial for data-driven deep learning algorithms. Yin et al. [120] and Kang et al. [121] explored publicly available datasets to investigate their properties for developing autonomous driving features. In this part, we briefly introduce the publicly available datasets that that are commonly used for pedestrian behavior prediction. Table VI lists the publicly available datasets that are used by existing works and the summaries.

1) Trajectory prediction: **ETH** [122] and **UCY** [123] datasets are widely used for evaluating pedestrian trajectories prediction. These two datasets contain five scenes of bird's-eye-view (BEV) videos collected in various scenarios, including crowded urban scenes. The ETH dataset contains two scenes with 750 annotated pedestrians, and UCY dataset contains three components with 786 annotated pedestrians. However, these two datasets are limited to pedestrians in crowds, and do not consider other road users.

KITTI [124] dataset contains driving scenarios collected by multi-sensors from the vehicle's view. The data is collected with a 64-layer LiDAR and two high-resolution stereo cameras (grayscale and color) with a resolution of 1392×512 pixels at 10 fps. It contains over 200,000 3D objects annotated in synchronized and calibrated LiDAR and stereo images. This dataset enables 3D detection and tracking estimation, and can also be used for pedestrian trajectory prediction.

Daimler [5] dataset consists of 68 sequences of images captured from the vehicle's view, of which 12,485 images contain pedestrians. The videos are recorded with a stereo camera with a resolution of 1176×640 pixels at 16 fps. The dataset contains four typical types of pedestrian behaviors, including crossing, stopping, starting, and bending in, and can be used to evaluate pedestrian trajectory prediction and intention classification.

New York Grand Central (GC) Dataset [125] contains more than 12,000 trajectories annotated in a one-hour-long BEV video. The video is recorded at 25 fps with a resolution of 1920×1080 pixels. This dataset includes crowd pedestrian scenes but is not collected in traffic scenarios.

Stanford Drone Dataset (SDD) [126] contains 20 scenes of BEV videos collected in a university campus. The videos are captured with a 4k camera on a quadcopter platform with a resolution of 1400×1904 pixels. It includes over 11,000 unique pedestrians and other road users, such as vehicles and bikers with their interactions captured.

Waymo Open Dataset [127] contains 1,150 scenes collected by multi-sensors from the vehicle's view in traffic scenarios. The sensors include five LiDAR sensors, and five high-resolution pinhole cameras. Three front cameras have a resolution of 1920×1280 pixels, two side cameras have a resolution of 1920×1040 pixels. The LiDAR on top has a scan range of 75m, the other four LiDAR have a scan range of 20m. Each scene is 20 seconds long, containing 2D and 3D objects labeled in LiDAR and camera images sampled at 10 Hz. The objects include pedestrians, cyclists, vehicles, and signs. This dataset has become increasingly popular for detection and tracking evaluation, and can also be used for evaluating trajectory prediction.

TABLE VI
EVALUATION METRICS AND DATASETS FOR PEDESTRIAN BEHAVIOR PREDICTION

Dataset (Year)	Citation (Total/ Per year/Last year)	Prediction Tasks and Used in Papers	Summary
ETH (2009) [122]; UCY (2007) [123]	1188 / 99 / 364 710 / 51 / 265	Trajectory: [23], [25]–[28], [30]–[37], [40]–[44], [47]–[54], [58]	Collected in crowded urban scenes in bird's-eye-view (BEV). There are five scenes with more than 1500 people. Drawbacks: Do not include other traffic agents, and they are not collected in traffic scenarios.
KITTI (2012) [124]	7952 / 884 / 3520	Trajectory: [55]	Collected in traffic scenarios from vehicle's view. The data is collected with a 64-layer LiDAR and two high-resolution stereo cameras (grayscale and color) with a resolution of 1392×512 pixels at 10 fps. It contains over 200,000 3D objects annotated in synchronized LiDAR and stereo images.
Daimler (2013) [5]	214 / 27 / 75	Trajectory: [21], [22], [74]	Collected in traffic scenarios from the vehicle's view. The videos are recorded with a stereo camera with a resolution of 1176×640 pixels at 16 fps. It consists of 68 sequences of stereo images, with four types of pedestrian behaviors. It can be used to evaluate trajectory and intention prediction.
New York Grand Central (GC) (2015) [125]	209 / 35 / 59	Trajectory: [29], [30]	Collected in New York grand central in BEV. The video is recorded at 25 fps with a resolution of 1920×1080 pixels. It consists of more than 12,000 trajectories in a one-hour video. Drawbacks: Do not include other traffic agents, and they are not collected in traffic scenarios.
SDD (2016) [126]	485 / 97 / 284	Trajectory: [40], [44], [47], [51], [55]	Collected in a university campus in BEV. The videos have a resolution of 1400×1904 pixels at 30 fps. It contains 20 scenes with over 11,000 pedestrians, and other road users such as vehicles and bikers.
Waymo (2020) [127]	453 / 453 / 449	Trajectory: [45]	Collected in traffic scenarios from vehicle's view. It consists of 1,150 scenes collected by multi-sensors including five LiDAR sensors and five high-resolution pinhole cameras. Three front cameras have a resolution of 1920×1280 pixels, two side cameras have a resolution of 1920×1040 pixels. The dataset contains 2D and 3D objects (pedestrians, cyclists, vehicles, and signs) labeled in LiDAR and camera images sampled at 10 Hz. There are over 23k 3D-tracked pedestrians and 45k 2D-tracked pedestrians labeled.
JAAD (2017) [73]	128 / 32 / 93	Trajectory: [25], [26], [63] Intention: [67], [70]–[73], [75]–[79] Joint: [84]	Collected in traffic scenarios from the vehicle's view. There are over 300 video clips. The HD videos are recorded with on-board monocular camera at 30 fps. Most of the videos have a resolution of 1920×1080 pixels. The duration is between 5 to 15 seconds. The dataset contains approximately 82,000 frames and 2,000 unique pedestrian samples. The number of pedestrians with behavior annotations is 686.
PIE (2019) [84]	86 / 43 / 86	Trajectory: [25], [26], [63] Intention: [79] Joint: [84], [85]	Collected in traffic scenarios from the vehicle's view. There are six sets consisting of over 6 hours of driving videos. The HD videos with a resolution of 1920×1080 pixels are recorded with on-board monocular camera at 30 fps. The average duration is 10 min. The dataset contains approximately 290,000 annotated frames. The number of pedestrians with behavior annotations is 1842. The annotations include the bounding boxes with occlusion flags, crossing intention confidence, and text labels for pedestrians' actions.
ActEV/VIRAT (2018) [128]	97 / 32 / 42	Joint: [83]	Collected in traffic scenarios in BEV. Includes 455 videos from 12 traffic scenes, with more than 12 hours of recordings. Most of the videos have a high resolution of 1920×1080 pixels.

To evaluate existing pedestrian trajectory prediction algorithms, Sadeghian et al. [119] built the TrajNet benchmark, which is based on selected trajectories from the ETH, UCY, and SDD datasets and uses the ADE and FDE evaluation metrics, and is expanded to TrajNet++ by Kothari et al. [39] with larger-scale data and more evaluation metrics.

For the trajectory prediction, there are datasets that only contains pedestrians, such as the Subway Station dataset [129] and the CUHK Crowd Dataset [130] used by Xu et al. [30]; and the Town Center Dataset [131] used by Xue et al. [49]. Besides, there are several datasets that contain urban traffic, such as ApolloScape [132] as used by Ma et al. [59], Interaction Dataset [133] as used by Li et al. [47], and nuScenes [134] as used by Yao et al. [25]. But these datasets are mainly designed for detection or for vehicle behavior prediction instead of pedestrian behavior prediction.

2) Intention prediction: For the intention prediction, many previous works are based on data collected by the authors themselves [7], [64], because they can design what information to include in the data collection. We outline the publicly available datasets that are commonly used for pedestrian intention prediction.

Joint Attention for Autonomous Driving (JAAD) [73] dataset contains over 300 video scenes, and each scene ranges from 5 to 15 seconds in duration. The videos are recorded with three types of onboard cameras at 30 fps. 60 clips are collected in North America by a camera with a resolution of 1920x1080, 276 clips are collected in Europe by a camera with a resolution of 1920x1080, and 10 clips are collected in Europe by a camera with a resolution of 1280×720 pixels. This dataset contains approximately 82,000 frames and 2,000 unique pedestrian samples comprising a total number of 337,000 bounding boxes with behavioral and contextual tags. The number of pedestrians with behavior annotations is 686.

Pedestrian Intention Estimation (PIE) [84] dataset contains over 6 hours of driving footage captured from the vehicle's view, and the videos are split into approximately 10 minutes long pieces and grouped into 6 sets. The HD videos with a resolution of 1920×1080 pixels are recorded with an onboard camera at 30 fps. The dataset contains approximately 290,000 annotated frames. The number of pedestrians with behavior annotations is 1842. The dataset provides pedestrian behaviors and continuous sequences at the point of crossing. The pedestrians are annotated with the bounding boxes with occlusion flags, and crossing intention confidence and text tags for their actions.

3) Joint prediction: The **JAAD** and **PIE** datasets can be used for evaluating both trajectory and intention prediction, as well as joint prediction.

The **ActEV/VIRAT** [128] dataset includes 455 videos at 30 fps from 12 traffic scenes in BEV with more than 12 hours of recordings, and can be used for the evaluation of both trajectory and intention prediction. Most of the videos have a resolution of 1920×1080 pixels.

Other datasets such as the one proposed by Kooij et al. [135], which consists of sequences including single pedestrians with the intention to cross the street, can be used to evaluate the trajectories at crossing areas and intention prediction.

C. Summary and Discussion of Datasets

Table VI lists the publicly available datasets that are used by existing works and the summaries. We presented the number of citations of each dataset in the table to show the popularity of the dataset, including the number of total citations, the citation per year after released, and the citation in the last year. The KITTI dataset and Waymo Open dataset can also be used for other tasks such as detection and tracking, so there are more citations. ETH and UCY datasets are the most popular for trajectory prediction. SDD is also popular as it contains the annotation of pedestrians and other road users and can be used to study the interactions. JAAD and PIE datasets are the most popular for intention prediction. These two datasets can also be used for joint prediction.

The ETH and UCY datasets, the most commonly used datasets for trajectory prediction, were proposed in 2007 and 2009. While the JAAD and PIE datasets, the most commonly used datasets for intention prediction, were proposed in 2017 and 2019, which are ten years later than the datasets for trajectory prediction. This is because the information of pedestrian intention is more implicit compared to trajectories, and hence, the labeling of intention is more difficult compared to the labeling of trajectories. On the other hand, the dataset used for training and evaluation can influence the development of the prediction models. The earlier appearance of the commonly used dataset for trajectory prediction is another reason for more papers on this topic compared to intention prediction.

We also looked into the places where the data was captured and found they are mainly collected in North America, Europe, and Asia, including the USA, Canada, Germany, Switzerland, Bulgaria, Cyprus, and China. There are few datasets with urban scenarios captured in South America, Africa, and Oceania. Future research could focus on developing more datasets for these places. Furthermore, the comparability of findings across datasets is another issue that needs to be tackled to enable the transferability of results as well as applicability for certain geographic regions.

VIII. Comparison and Discussion

A. Performance of Existing Models

In this section, we compare the performance of some of the reviewed prediction methods. To align and compare the results, we select the works that used the most common publicly available datasets and metrics. The joint prediction is evaluated separately for trajectory and intention, so we compare them with the trajectory and intention prediction on corresponding datasets.

1) Trajectory prediction: For the trajectory prediction, we compare the ADE and FDE values in meters, with 3.2s observation time and 4.8s prediction time on the ETH and UCY datasets. In Table VII, we list the evaluation results, model features, and summarize the methods used for feature extraction and modeling. From the first LSTM-based network for trajectory prediction, Social-LSTM [27], to the most recent model, AgentFormer [42], the ADE has improved from 0.72m to 0.18m, and the FDE improved from 1.54m to 0.29m.

The models intended to consider more model features to improve the accuracy, including the consideration of social interaction and the interaction within a scene. For the social interactions, the social pooling method improved to more complicated attention pooling networks, and afterwards, the graph-based spatio-temporal attention network took place. Recently, researchers have focused on the interactions with other road users, i.e., the heterogeneous interaction, to model real traffic scenarios. The graph-based representation is a powerful tool to model non-symmetric interactions. The environment and appearance features encoded by CNNs from the images help to improve the results. Besides, the instant destination is increasingly popular to be considered while predicting in goal-driven networks.

For prediction methods, instead of only using sequential or non-sequential methods, many models combine the CNNs and the sequential models to extract both the spatial and temporal features. The multi-modal GAN and CVAE models that can provide multiple plausible predictions are becoming increasingly popular compared to the uni-modal methods that predict a single distribution. The recurrent LSTM models are gradually replaced by the TCN models and TF models that have made a breakthrough in performance and can be paralleled to reduce training time. The current state-of-the-art algorithm AngentFormer [42] used the TF-based CVAE model and use agent-aware attention to model the spatio-temporal interaction at the same time.

2) Intention prediction: For the intention prediction, we compare the AP and ACC for the C/NC classification on the most commonly used JAAD dataset. Table VIII lists the selected algorithms, their observation and prediction time horizon, the evaluation results, model features, and the summary. From the baseline method provided in the JAAD dataset [73] to the most recent intention prediction work [67], the AP is increased from 0.63 to 0.90.

Early works considered the appearance and skeleton of pedestrians and the environment context. Recent research included the vehicle states and the interaction with other road users to improve the precision. Off-the-shelf CNN-based

TABLE VII
COMPARISON FOR TRAJECTORY PREDICTION

Paper, Author	Year	ADE / FDE	Model Features	Summary of Network Structures
Social-LSTM [27] (Alahi et al.)	2016	0.72 / 1.54	Trajectory, social interaction	LSTMs for sequence prediction; social pooling to model social interaction.
Social-GAN [31] (Gupta et al.)	2018	0.58 / 1.18	Trajectory, social interaction	LSTM-based GAN for multi-modal sequence prediction; social pooling network to model social interaction.
[23] (Nikhil et al.)	2018	0.59 / 1.22	Trajectory	CNNs instead of LSTMs for sequence prediction, enables parallelization.
SNS [52] (Lisotto et al.)	2019	0.36 / 1.81	Trajectory, social interaction, environment	LSTMs for sequence prediction; social, navigation and semantic pooling to model social interaction and environmental interaction.
Sophie [51] (Sadeghian et al.)	2019	0.54 / 1.15	Trajectory, social interaction, environment	LSTM-based GAN for multi-modal sequence prediction; CNNs for environmental feature extraction; soft-attention to model social interaction.
[34] (Zhang et al.)	2019	0.48 / 0.99	Trajectory, social interaction	LSTM encoder-decoder for sequence prediction; social graph network to model social interaction.
CGNS [47] (LI et al.)	2019	0.49 / 0.97	Trajectory, social interaction, environment	GRU-based CVAE for multi-modal sequence prediction; CNNs for environmental feature extraction; soft-attention to model social interaction.
Social-BiGAT [50] (Kosaraju et al.)	2019	0.48 / 1.00	Trajectory, social interaction, environment	LSTM-based GAN for multi-modal sequence prediction; CNNs for environmental feature extraction; GAT to model social interaction.
[83] (Liang et al.) (Joint Prediction)	2019	0.46 / 1.00	Person behavior, social interaction, Person-ORU interaction environment	LSTM for sequence prediction; CNNs for environmental and appearance feature extraction; geometric relation function for person-object interaction modeling.
SR-LSTM [32] (Zhang et al.)	2019	0.45 / 0.94	Trajectory, social interaction	LSTMs for sequence prediction; social-aware information selection and state refinement module to model social interaction.
Social-ways [33] (Amirian et al.)	2019	0.46 / 0.83	Trajectory, social interaction	LSTM-based Info-GAN for multi-modal sequence prediction; attention pooling to model social interaction.
STGAT [35] (Huang et al.)	2019	0.43 / 0.83	Trajectory, social interaction	LSTM encoder-decoder for sequence prediction; GAT for social interaction modeling.
Social-STGCNN [41] (Mohamed et al.)	2020	0.44 / 0.75	Trajectory, social interaction	TCNs and CNNs for sequence prediction, enables parallelization; spatio-temporal GCNs to model social interaction.
NMMP [40] (Hu et al.)	2020	0.41 / 0.82	Trajectory, social interaction, Person-ORU interaction	LSTM-based GAN for multi-modal sequence prediction; graph-based NMMP module to model the interaction with other road users.
[58] (Eiffert et al.)	2020	0.34 / 0.77	Trajectory, social interaction, Person-ORU interaction	LSTM-based GAN for multi-modal sequence prediction; Mixture Density Networks (MDN) and GVAT module to model the interaction with other road users.
Transformer (TF) [24] (Giuliari et al.)	2020	0.31 / 0.55	Trajectory	TF for sequence prediction, enables parallelization for encoder-phase.
STAR [37] (Yu et al.)	2020	0.26 / 0.53	Trajectory, social interaction	TF for sequence prediction; GCNs to model social interaction.
PECNet [44] (Mangalam et al.)	2020	0.29 / 0.48	Trajectory, social interaction, destinations	CVAE for multi-modal sequence prediction with an endpoint encoder for destinations; social pooling to model social interaction.
Tra2Tra [43] (Xu et al.)	2021	0.20 / 0.54	Trajectory, social interaction	LSTM for sequence prediction; LSTM-based spatio-temporal attention module to model social interaction.
SGNet [26] (Wang et al.)	2021	0.18 / 0.35	Trajectory, destinations	GRU-based CVAE for multi-modal sequence prediction; a stepwise goal estimator (SGE) for destination estimation.
Bitrap [25] (Yao et al.)	2021	0.18 / 0.35	Trajectory, destination	GRU-based CVAE for multi-modal sequence prediction with a GRU-based encoder and goal estimation, and a bi-directional decoder.
AgentFormer [42] (Yuan et al.)	2021	**0.18 / 0.29**	Trajectory, social interaction	TF-based CVAE for multi-modal sequence prediction; agent-aware TF to model social interaction on both time and social dimensions.

segmentation and detection models are used for appearance and environmental feature extraction. 3D-CNNs can be used to extract both spatial and temporal information. A longer observation time improves the results [70], [78] showing that time series-related information contributes to the intention prediction. Recent work combined the CNN-based model with sequential models, including LSTMs and conv-LSTMs, to better extract the temporal information. The current state-of-the-art model [67] used a 3D-CNN to extract spatial and temporal behavioral feature, and encode the environmental and vehicle interaction feature with an additional distance encoding module.

B. Research Gaps and Future Opportunities

Next, we discuss the current research gaps in pedestrian behavior prediction that could be improved for future research.

1) Trajectory prediction: Most existing trajectory prediction studies relied on past trajectories, and did not take full use of the appearance and skeleton behavioral features like intention prediction studies. Only a few of them (e.g., [46]) consider the pedestrians' visual behavioral features. In future works, the visual behavioral features can be considered even more. Another problem of existing trajectory prediction is that the prediction considers the "perfect" detection and tracking (i.e.,

TABLE VIII
COMPARISON FOR INTENTION PREDICTION

Paper, Author	Year	Observation / Prediction Time (sec)	AP / ACC	Model Features	Summary of Network Structures
ATGC [73] (Rasouli et al.)	2017	0.33-0.5 / 0.03 (Next frame)	0.63 / –	Appearance cue, environment	Used CNNs to extract behavioral features for prediction.
[70] (Fang et al.)	2018	0.46 / 0.03 (Next frame)	– / **0.88**	Skeleton cue	Used CNNs and skeleton fitting to extract skeleton-based behavior features for prediction.
Res-EnDec [76] (Gujjar et al.)	2019	0.53 / 0.53	0.81 / –	Appearance cue, environment	Used 3D-CNNs as the encoder and conv-LSTMs as the decoder for generating future video, and appended a binary classifier to the generator for intention classification.
[75] (Chaabane et al.)	2020	0.53 / 0.53	0.87 / –	Appearance cue, environment	Used 3D-CNNs as the encoder and depth-wise separable conv-LSTMs as the decoder for generating future video, and appended a binary classifier to the generator for intention classification.
[72] (Lorenzo et al.)	2020	– / 1.0	0.83 / –	Appearance cue	Used CNNs to extract behavioral features, and applied LSTMs, GRUs, and the bidirectional variants of LSTMs and GRUs for crossing probability prediction.
[78] (Liu et al.)	2020	1.0 / 1.0	– / 0.79	Appearance cue, Person-ORU interaction, environment	Used a CNN-based model for image parsing, and encoding the appearance features. Used graph convolution for spatio-temporal interaction extraction. GRUs are used for capturing the temporal features and for behavior prediction.
PCPA [79] (Kotseruba et al.)	2021	0.53 / 0.03 (Next frame)	0.86 / 0.85	Skeleton and appearance cue, vehicle state, environment	Used 3D-CNNs for local visual context extracting, and used GRUs for non-visual features encoding. The temporal attention and modality attention modules are applied to learn the interaction.
[67] (Yang et al.)	2021	0.53 / 0.03 (Next frame)	**0.90** / –	Appearance cue, vehicle states, environment	Used 3D-CNNs to Extract spatial and temporal behavioral features, and used a distance encoding module to extract environmental contextual cues and vehicle features.
PIE [84] (Rasouli et al.) (Joint Prediction)	2019	0.5 / 0.03 (Next frame)	– / 0.79	Appearance cue, vehicle states, environment	Used LSTMs, CNNs and conv-LSTMs for joint prediction. For the trajectory prediction and vehicle speed prediction stream, the authors used LSTMs with temporal attention in encoder inputs, and self-attention in decoder inputs. For the intention estimation, CNNs are used for appearance behavioral feature extraction, and conv-LSTMs are used as encoders, and LSTMs are used as decoders.

the ground-truth of past trajectories). However, this is usually not feasible in practice. Future work should look at how to predict under conditions of imperfect detection and tracking and how to develop an end-to-end prediction from raw sensor data. Besides, existing works have used static graph-based models to extract spatio-temporal features. As dynamic graph-based models have shown a potential of better reflecting the spatio-temproal features compared with the static graph-based model in traffic flow prediction as used by Peng et al. [136], in future trajectory prediction works, researchers can also consider using dynamic graph-based models.

2) Intention prediction: Only a few works (e.g., [7], [57], [67]) considered the traffic rules and signals while predicting the crossing behaviors. The existence of crosswalks and traffic signal lights are easy to get while strongly influencing the crossing behavior. Hence, such factors can be combined with other implicit environmental context features for intention prediction in future works. The interaction with vehicles and other road users can influence the pedestrian's decision. Unlike trajectory prediction, which considered various interactions between different traffic agents, most intention prediction studies used hand-crafted features to define the relationship with a single vehicle as shown in Table III. In future works, the graph-based or attention-pooling method can also be employed to extract the interaction relationships in crossing intention prediction.

As discussed, intention prediction usually requires large computational resources. More research could focus on investigating whether adding the intention prediction can bring noticeable improvements to an application domain.

3) Joint prediction: The predicted results of trajectory and intention can be used to improve each other. Future works can focus on joint prediction, which could use past trajectories and interaction information that is usually used in trajectory prediction, and appearance behavioral cue that is typically used in intention prediction. The two prediction branches can share the extracted features to compensate for each other.

4) Hybrid models: The behavior of pedestrians in urban traffic usually includes interactions between multiple road users. As we summarized in Table III, the interactions can either be learned implicitly by deep learning models that can include as much information as possible without requiring expert knowledge but that are hard to explain, or be represented by using knowledge-based hand-crafted features that are explainable but requires prior knowledge instead. In future works, we can develop hybrid models to take advantage of both approaches. For example, we can use conventional models with parameters learned from deep learning networks such as the Deep Social Force proposed by Kreiss [137], or implement the conventional knowledge-based model as a layer in the deep learning network.

5) Benchmark: As we reviewed and summarized in Sec VII, existing works use various datasets and metrics. The most popularly used datasets for trajectory prediction, ETH and UCY, are limited to crowds but not designed for traffic scenarios, and hence, they are not suitable to represent the performance for automated driving usage. The recently proposed popular benchmark, TrajNet, and TrajNet++ are not designed for automated driving scenarios and do not cover enough traffic scenes. For intention prediction, many researchers still use self-collected datasets on a selected intersection, which makes it difficult for others to replicate and compare the work. For joint prediction, many existing works evaluate the trajectory and intention prediction separately with different datasets for comparison with previous works. Existing benchmarks either focus on trajectory or intention prediction. In future works, a benchmark can be defined and explored for the behavior prediction that includes both tasks and to thoroughly compare the performance for the joint prediction.

IX. Conclusions

In this paper, we have presented a thorough review of pedestrian behavior prediction models that use deep learning methods extracted from 92 papers. Compared with previous literature review papers, the original contributions of our review paper are as follows:

- Both trajectory and intention predictions are considered and analyzed, instead of only focusing on a single type of task;
- We have categorized existing works by three different criteria to provide a perspective from different dimensions, instead of reviewing the papers from a single criterion;
- We introduced widely used datasets containing urban scenarios and we have evaluated and compared previous methods on such publicly available datasets.
- We included the most recent papers from 2016 to 2021.

We have discussed the model features used by existing models, and how they extracted these features. We have presented, categorized, and discussed the prediction methods used by existing works. The advantages and drawbacks of using different model features, and the properties of different prediction methods are discussed in detail. We have discussed why there is more research on trajectory prediction than intention prediction, how much effort we should put into intention prediction, which prediction methods we should use for which task, and the distribution of the datasets in the world. Finally, we outline the research gaps and possible research directions for improving the performance of prediction algorithms for urban scenarios.

Acknowledgment

This research is funded by the European research project "SHAPE-IT – Supporting the Interaction of Humans and Automated Vehicles: Preparing for the Environment of Tomorrow". This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 860410. The authors would like to thank Prof. Marco Dozza for his valuable comments. Zhongjun Ni is acknowledged for his suggestions on visualization.

References

[1] WHO, "Global status report on road safety 2018: Summary," World Health Organization, Report, 2018.

[2] K. Rumar, "Transport safety visions, targets and strategies: beyond 2000," *1st European Transport Safety Lecture. European Transport Safety Council, Brussels, Tech. Rep*, 1999.

[3] S. Lefèvre, D. Vasquez, and C. Laugier, "A survey on motion prediction and risk assessment for intelligent vehicles," *ROBOMECH journal*, vol. 1, no. 1, pp. 1–14, 2014.

[4] S. Mozaffari, O. Y. Al-Jarrah, M. Dianati, P. Jennings, and A. Mouzakitis, "Deep learning-based vehicle behavior prediction for autonomous driving applications: A review," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[5] N. Schneider and D. M. Gavrila, "Pedestrian path prediction with recursive bayesian filters: A comparative study," in *German Conference on Pattern Recognition*. Springer, 2013, pp. 174–183.

[6] S. Ferguson, B. Luders, R. C. Grande, and J. P. How, "Real-time predictive modeling and robust avoidance of pedestrians with uncertain, changing intentions," in *Algorithmic Foundations of Robotics XI*. Springer, 2015, pp. 161–177.

[7] H. Zhang, Y. Liu, C. Wang, R. Fu, Q. Sun, and Z. Li, "Research on a pedestrian crossing intention recognition model based on natural observation data," *Sensors*, vol. 20, no. 6, p. 1776, 2020.

[8] M. Moussaïd, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz, "The walking behaviour of pedestrian social groups and its impact on crowd dynamics," *PloS one*, vol. 5, no. 4, p. e10047, 2010.

[9] M. S. Shirazi and B. Morris, "Observing behaviors at intersections: A review of recent studies & developments," in *2015 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2015, pp. 1258–1263.

[10] E. Ohn-Bar and M. M. Trivedi, "Looking at humans in the age of self-driving and highly automated vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 1, pp. 90–104, 2016.

[11] A. Rasouli and J. K. Tsotsos, "Autonomous vehicles that interact with pedestrians: A survey of theory and practice," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 900–918, 2019.

[12] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," *Physical review E*, vol. 51, no. 5, p. 4282, 1995.

[13] C. Schöller, V. Aravantinos, F. Lay, and A. Knoll, "What the constant velocity model can teach us about pedestrian motion prediction," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1696–1703, 2020.

[14] W. H. Organization *et al.*, "Pedestrian safety: a road safety manual for decision-makers and practitioners," 2013.

[15] T. Hirakawa, T. Yamashita, T. Tamaki, and H. Fujiyoshi, "Survey on vision-based path prediction," in *International Conference on Distributed, Ambient, and Pervasive Interactions*. Springer, 2018, pp. 48–64.

[16] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: A survey," *The International Journal of Robotics Research*, vol. 39, no. 8, pp. 895–935, 2020.

[17] R. Korbmacher and A. Tordeux, "Review of pedestrian trajectory prediction methods: Comparing deep learning and knowledge-based approaches," *arXiv preprint arXiv:2111.06740*, 2021.

[18] D. Ridel, E. Rehder, M. Lauer, C. Stiller, and D. Wolf, "A literature review on the prediction of pedestrian behavior in urban scenarios," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, Conference Proceedings, pp. 3105–3112.

[19] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, 2014, pp. 1–10.

[20] J. Bock, T. Beemelmanns, M. Klösges, and J. Kotte, "Self-learning trajectory prediction with recurrent neural networks at intelligent intersections." in *VEHITS*, 2017, pp. 346–351.

[21] K. Saleh, M. Hossny, and S. Nahavandi, "Intent prediction of vulnerable road users from motion trajectories using stacked lstm network," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2017, Conference Proceedings, pp. 327–332.

[22] ——, "Intent prediction of pedestrians via motion trajectories using stacked recurrent neural networks," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 4, pp. 414–424, 2018.

[23] N. Nikhil and B. Tran Morris, "Convolutional neural network for trajectory prediction," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, Conference Proceedings.

[24] F. Giuliari, I. Hasan, M. Cristani, and F. Galasso, "Transformer networks for trajectory forecasting," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 10 335–10 342.

[25] Y. Yao, E. Atkins, M. Johnson-Roberson, R. Vasudevan, and X. Du, "Bitrap: Bi-directional pedestrian trajectory prediction with multi-modal goal estimation," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1463–1470, 2021.

[26] C. Wang, Y. Wang, M. Xu, and D. J. Crandall, "Stepwise goal-driven networks for trajectory prediction," *arXiv preprint arXiv:2103.14107*, 2021.

[27] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social lstm: Human trajectory prediction in crowded spaces," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, Conference Proceedings, pp. 961–971.

[28] A. Vemula, K. Muelling, and J. Oh, "Social attention: Modeling attention in human crowds," in *2018 IEEE international Conference on Robotics and Automation (ICRA)*. IEEE, 2018, Conference Proceedings, pp. 4601–4607.

[29] T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Soft+ hardwired attention: An lstm framework for human trajectory prediction and abnormal event detection," *Neural networks*, vol. 108, pp. 466–478, 2018.

[30] Y. Xu, Z. Piao, and S. Gao, "Encoding crowd interaction with deep neural network for pedestrian trajectory prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, Conference Proceedings, pp. 5275–5284.

[31] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social gan: Socially acceptable trajectories with generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, Conference Proceedings, pp. 2255–2264.

[32] P. Zhang, W. Ouyang, P. Zhang, J. Xue, and N. Zheng, "Sr-lstm: State refinement for lstm towards pedestrian trajectory prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, Conference Proceedings, pp. 12 085–12 094.

[33] J. Amirian, J.-B. Hayet, and J. Pettré, "Social ways: Learning multi-modal distributions of pedestrian trajectories with gans," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, Conference Proceedings, pp. 2964–2972.

[34] L. Zhang, Q. She, and P. Guo, "Stochastic trajectory prediction with social graph network," *arXiv preprint arXiv:1907.10233*, 2019.

[35] Y. Huang, H. Bi, Z. Li, T. Mao, and Z. Wang, "Stgat: Modeling spatial-temporal interactions for human trajectory prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, Conference Proceedings, pp. 6272–6281.

[36] S. Liu, H. Liu, H. Bi, and T. Mao, "Col-gan: Plausible and collision-less trajectory prediction by attention-based gan," *IEEE Access*, vol. 8, pp. 101 662–101 671, 2020.

[37] C. Yu, X. Ma, J. Ren, H. Zhao, and S. Yi, "Spatio-temporal graph transformer networks for pedestrian trajectory prediction," in *European Conference on Computer Vision*. Springer, 2020, Conference Proceedings, pp. 507–523.

[38] Y. Liu, Q. Yan, and A. Alahi, "Social nce: Contrastive learning of socially-aware motion representations," *arXiv preprint arXiv:2012.11717*, 2020.

[39] P. Kothari, S. Kreiss, and A. Alahi, "Human trajectory forecasting in crowds: A deep learning perspective," *arXiv preprint arXiv:2007.03639*, 2020.

[40] Y. Hu, S. Chen, Y. Zhang, and X. Gu, "Collaborative motion prediction via neural motion message passing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, Conference Proceedings, pp. 6319–6328.

[41] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel, "Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, Conference Proceedings, pp. 14 424–14 432.

[42] Y. Yuan, X. Weng, Y. Ou, and K. Kitani, "Agentformer: Agent-aware transformers for socio-temporal multi-agent forecasting," *arXiv preprint arXiv:2103.14023*, 2021.

[43] Y. Xu, D. Ren, M. Li, Y. Chen, M. Fan, and H. Xia, "Tra2tra: Trajectory-to-trajectory prediction with a global social spatial-temporal attentive neural network," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1574–1581, 2021.

[44] K. Mangalam, H. Girase, S. Agarwal, K.-H. Lee, E. Adeli, J. Malik, and A. Gaidon, "It is not the journey but the destination: Endpoint conditioned trajectory prediction," *arXiv preprint arXiv:2004.02025*, 2020.

[45] C. Zhang, C. Berger, and M. Dozza, "Social-iwstcnn: A social interaction-weighted spatio-temporal convolutional neural network for pedestrian trajectory prediction in urban traffic scenarios," in *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2021, pp. 1515–1522.

[46] J. Zhong, H. Sun, W. Cao, and Z. He, "Pedestrian motion trajectory prediction with stereo-based 3d deep pose estimation and trajectory learning," *IEEE access*, vol. 8, pp. 23 480–23 486, 2020.

[47] J. Li, H. Ma, and M. Tomizuka, "Conditional generative neural system for probabilistic trajectory prediction," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 6150–6156.

[48] H. Manh and G. Alaghband, "Scene-lstm: A model for human trajectory prediction," *arXiv preprint arXiv:1808.04018*, 2018.

[49] H. Xue, D. Q. Huynh, and M. Reynolds, "Ss-lstm: A hierarchical lstm model for pedestrian trajectory prediction," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018, Conference Proceedings, pp. 1186–1194.

[50] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. Reid, H. Rezatofighi, and S. Savarese, "Social-bigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks," in *Advances in Neural Information Processing Systems*, 2019, Conference Proceedings, pp. 137–146.

[51] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi, and S. Savarese, "Sophie: An attentive gan for predicting paths compliant to social and physical constraints," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, Conference Proceedings, pp. 1349–1358.

[52] M. Lisotto, P. Coscia, and L. Ballan, "Social and scene-aware trajectory prediction in crowded spaces," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2019, Conference Proceedings.

[53] N. Bhujel, E. K. Teoh, and W.-Y. Yau, "Pedestrian trajectory prediction using rnn encoder-decoder with spatio-temporal attentions," in *2019 IEEE 5th International Conference on Mechatronics System and Robots (ICMSR)*. IEEE, 2019, pp. 110–114.

[54] A. Syed and B. Morris, "Stgt: Forecasting pedestrian motion using spatio-temporal graph transformer," in *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2021, Conference Proceedings.

[55] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. Torr, and M. Chandraker, "Desire: Distant future prediction in dynamic scenes with interacting agents," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 336–345.

[56] X. Song, K. Chen, X. Li, J. Sun, B. Hou, Y. Cui, B. Zhang, G. Xiong, and Z. Wang, "Pedestrian trajectory prediction based on deep convolutional lstm network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3285–3302, 2020.

[57] Y. Ma, E. W. Lee, Z. Hu, M. Shi, and R. K. Yuen, "An intelligence-based approach for prediction of microscopic pedestrian walking behavior," *IEEE transactions on intelligent transportation systems*, vol. 20, no. 10, pp. 3964–3980, 2019.

[58] S. Eiffert, K. Li, M. Shan, S. Worrall, S. Sukkarieh, and E. Nebot, "Probabilistic crowd gan: Multimodal pedestrian trajectory prediction using a graph vehicle-pedestrian attention network," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5026–5033, 2020.

[59] Y. Ma, X. Zhu, S. Zhang, R. Yang, W. Wang, and D. Manocha, "Trafficpredict: Trajectory prediction for heterogeneous traffic-agents," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, Conference Proceedings, pp. 6120–6127.

[60] K. Li, M. Shan, K. Narula, S. Worrall, and E. Nebot, "Socially aware crowd navigation with multimodal pedestrian trajectory prediction for autonomous vehicles," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020, pp. 1–8.

[61] S. Carrasco, D. F. Llorca, and M. A. Sotelo, "Scout: Socially-consistent and understandable graph attention network for trajectory prediction of vehicles and vrus," in *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2021, Conference Proceedings.

[62] R. Chandra, U. Bhattacharya, A. Bera, and D. Manocha, "Traphic: Trajectory prediction in dense and heterogeneous traffic using weighted interactions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, Conference Proceedings, pp. 8483–8492.

[63] K. Kim, Y. K. Lee, H. Ahn, S. Hahn, and S. Oh, "Pedestrian intention prediction for autonomous driving using a multiple stakeholder perspective model," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 7957–7962.

[64] B. Völz, K. Behrendt, H. Mielenz, I. Gilitschenski, R. Siegwart, and J. Nieto, "A data-driven approach for pedestrian intention estimation," in

2016 IEEE 19th international conference on intelligent transportation systems (itsc). IEEE, 2016, Conference Proceedings, pp. 2607–2612.

[65] B. Völz, H. Mielenz, I. Gilitschenski, R. Siegwart, and J. Nieto, "Inferring pedestrian motions at urban crosswalks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 2, pp. 544–555, 2018.

[66] J. Zhao, Y. Li, H. Xu, and H. Liu, "Probabilistic prediction of pedestrian crossing intention using roadside lidar data," *IEEE Access*, vol. 7, pp. 93 781–93 790, 2019.

[67] B. Yang, W. Zhan, P. Wang, C. Chan, Y. Cai, and N. Wang, "Crossing or not? context-based recognition of pedestrian crossing intention in the urban environment," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[68] S. Zhang, M. Abdel-Aty, Y. Wu, and O. Zheng, "Pedestrian crossing intention prediction at red-light using pose estimation," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[69] Z. Fang, D. Vázquez, and A. M. López, "On-board detection of pedestrian intentions," *Sensors*, vol. 17, no. 10, p. 2193, 2017.

[70] Z. Fang and A. M. López, "Is the pedestrian going to cross? answering by 2d pose estimation," in *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018, Conference Proceedings, pp. 1271–1276.

[71] F. Piccoli, R. Balakrishnan, M. J. Perez, M. Sachdeo, C. Nunez, M. Tang, K. Andreasson, K. Bjurek, R. D. Raj, E. Davidsson *et al.*, "Fussi-net: Fusion of spatio-temporal skeletons for intention prediction network," in *2020 54th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2020, pp. 68–72.

[72] J. Lorenzo, I. Parra, F. Wirth, C. Stiller, D. F. Llorca, and M. A. Sotelo, "Rnn-based pedestrian crossing prediction using activity and pose-related features," in *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2020, pp. 1801–1806.

[73] A. Rasouli, I. Kotseruba, and J. K. Tsotsos, "Are they going to cross? a benchmark dataset and baseline for pedestrian crosswalk behavior," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, Oct 2017, Conference Proceedings, pp. 206–213.

[74] M. Hoy, Z. Tu, K. Dang, and J. Dauwels, "Learning to predict pedestrian intention via variational tracking networks," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 3132–3137.

[75] M. Chaabane, A. Trabelsi, N. Blanchard, and R. Beveridge, "Looking ahead: Anticipating pedestrians crossing with future frames prediction," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, Conference Proceedings, pp. 2297–2306.

[76] P. Gujjar and R. Vaughan, "Classifying pedestrian actions in advance using predicted video of urban driving scenes," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, Conference Proceedings, pp. 2097–2103.

[77] D. O. Pop, A. Rogozan, C. Chatelain, F. Nashashibi, and A. Bensrhair, "Multi-task deep learning for pedestrian detection, action recognition and time to cross prediction," *IEEE Access*, vol. 7, pp. 149 318–149 327, 2019.

[78] B. Liu, E. Adeli, Z. Cao, K.-H. Lee, A. Shenoi, A. Gaidon, and J. C. Niebles, "Spatiotemporal relationship reasoning for pedestrian intent prediction," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3485–3492, 2020.

[79] I. Kotseruba, A. Rasouli, and J. K. Tsotsos, "Benchmark for evaluating pedestrian action prediction," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 1258–1268.

[80] Z. Huang, A. Hasan, K. Shin, R. Li, and K. Driggs-Campbell, "Long-term pedestrian trajectory prediction using mutable intention filter and warp lstm," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 542–549, 2020.

[81] M. Goldhammer, S. Köhler, S. Zernetsch, K. Doll, B. Sick, and K. Dietmayer, "Intentions of vulnerable road users—detection and forecasting by means of machine learning," *IEEE transactions on intelligent transportation systems*, vol. 21, no. 7, pp. 3035–3045, 2019.

[82] H. Wu, L. Wang, S. Zheng, Q. Xu, and J. Wang, "Crossing-road pedestrian trajectory prediction based on intention and behavior identification," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020, pp. 1–6.

[83] J. Liang, L. Jiang, J. C. Niebles, A. G. Hauptmann, and L. Fei-Fei, "Peeking into the future: Predicting future person activities and locations in videos," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, Conference Proceedings, pp. 5725–5734.

[84] A. Rasouli, I. Kotseruba, T. Kunic, and J. K. Tsotsos, "Pie: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, Conference Proceedings, pp. 6262–6271.

[85] I. Kotseruba, A. Rasouli, and J. K. Tsotsos, "Do they want to cross? understanding pedestrian intention for behavior prediction," in *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2020, pp. 1688–1693.

[86] E. Rehder, F. Wirth, M. Lauer, and C. Stiller, "Pedestrian prediction by planning using deep neural networks," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 5903–5908.

[87] B. Völz, H. Mielenz, G. Agamennoni, and R. Siegwart, "Feature relevance estimation for learning pedestrian behavior at crosswalks," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. IEEE, 2015, pp. 854–860.

[88] A. Rasouli and J. K. Tsotsos, "Joint attention in driver-pedestrian interaction: from theory to practice," *arXiv preprint arXiv:1802.02522*, 2018.

[89] S. Schmidt and B. Faerber, "Pedestrians at the kerb–recognising the action intentions of humans," *Transportation research part F: traffic psychology and behaviour*, vol. 12, no. 4, pp. 300–310, 2009.

[90] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *European conference on computer vision*. Springer, 2016, pp. 483–499.

[91] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "Openpose: realtime multi-person 2d pose estimation using part affinity fields," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 1, pp. 172–186, 2019.

[92] M. M. Hamed, "Analysis of pedestrians' behavior at pedestrian crossings," *Safety science*, vol. 38, no. 1, pp. 63–82, 2001.

[93] H. Guo, Z. Gao, X. Yang, and X. Jiang, "Modeling pedestrian violation behavior at signalized crosswalks in china: A hazards-based duration approach," *Traffic injury prevention*, vol. 12, no. 1, pp. 96–103, 2011.

[94] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in *International Conference on Learning Representations (ICLR)*, 2017, Conference Proceedings.

[95] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[96] A. Howard, A. Zhmoginov, L.-C. Chen, M. Sandler, and M. Zhu, "Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation," in *CVPR*, 2018.

[97] J. Li, A. Raventos, A. Bhargava, T. Tagawa, and A. Gaidon, "Learning to fuse things and stuff," *arXiv preprint arXiv:1812.01192*, 2018.

[98] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[99] Q. Zhu, "Hidden markov model for dynamic obstacle avoidance of mobile robot navigation," *IEEE Transactions on Robotics and Automation*, vol. 7, no. 3, pp. 390–397, 1991.

[100] R. Kelley, A. Tavakkoli, C. King, M. Nicolescu, M. Nicolescu, and G. Bebis, "Understanding human intentions via hidden markov models in autonomous mobile robots," in *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, 2008, pp. 367–374.

[101] T. Bandyopadhyay, C. Z. Jie, D. Hsu, M. H. Ang, D. Rus, and E. Frazzoli, "Intention-aware pedestrian avoidance," in *Experimental Robotics*. Springer, 2013, pp. 963–977.

[102] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Gaussian process dynamical models for human motion," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 2, pp. 283–298, 2007.

[103] V. Karasev, A. Ayvaci, B. Heisele, and S. Soatto, "Intent-aware long-term prediction of pedestrian motion," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 2543–2549.

[104] M. Karl, M. Soelch, J. Bayer, and P. Van der Smagt, "Deep variational bayes filters: Unsupervised learning of state space models from raw data," *arXiv preprint arXiv:1605.06432*, 2016.

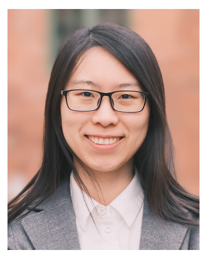
[105] M. Gutmann and A. Hyvärinen, "Noise-contrastive estimation: A new estimation principle for unnormalized statistical models," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 297–304.

[106] A. v. d. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.

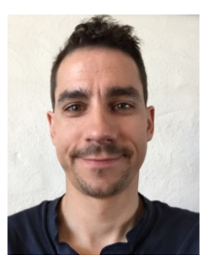
[107] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," *Advances in neural information processing systems*, vol. 2015, pp. 802–810, 2015.

[108] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in

31st Conference on Neural Information Processing Systems (NIPS), 2017, Conference Paper.
[109] A. Syed and B. T. Morris, "Cnn, segmentation or semantic embeddings: Evaluating scene context for trajectory prediction," in *International Symposium on Visual Computing*. Springer, 2020, pp. 706–717.
[110] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. V. Le, and R. Salakhutdinov, "Transformer-xl: Attentive language models beyond a fixed-length context," *arXiv preprint arXiv:1901.02860*, 2019.
[111] J. W. Rae, A. Potapenko, S. M. Jayakumar, and T. P. Lillicrap, "Compressive transformers for long-range sequence modelling," *arXiv preprint arXiv:1911.05507*, 2019.
[112] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.
[113] C. Lea, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks: A unified approach to action segmentation," in *European Conference on Computer Vision*. Springer, 2016, Conference Proceedings, pp. 47–54.
[114] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, Conference Proceedings, pp. 156–165.
[115] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.
[116] S. Qi, W. Wang, B. Jia, J. Shen, and S.-C. Zhu, "Learning human-object interactions by graph parsing neural networks," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 401–417.
[117] M. Li, S. Chen, X. Chen, Y. Zhang, Y. Wang, and Q. Tian, "Actional-structural graph convolutional networks for skeleton-based action recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3595–3603.
[118] A. Jain, A. R. Zamir, S. Savarese, and A. Saxena, "Structural-rnn: Deep learning on spatio-temporal graphs," in *Proceedings of the ieee conference on computer vision and pattern recognition*, 2016, pp. 5308–5317.
[119] A. Sadeghian, V. Kosaraju, A. Gupta, S. Savarese, and A. Alahi, "Trajnet: Towards a benchmark for human trajectory prediction," *arXiv preprint*, 2018.
[120] H. Yin and C. Berger, "When to use what data set for your self-driving car algorithm: An overview of publicly available driving datasets," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2017, Conference Proceedings, pp. 1–8.
[121] Y. Kang, H. Yin, and C. Berger, "Test your self-driving algorithm: An overview of publicly available driving datasets and virtual testing environments," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 2, pp. 171–185, 2019.
[122] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 261–268.
[123] A. Lerner, Y. Chrysanthou, and D. Lischinski, "Crowds by example," in *Computer graphics forum*, vol. 26. Wiley Online Library, 2007, pp. 655–664.
[124] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
[125] S. Yi, H. Li, and X. Wang, "Understanding pedestrian behaviors from stationary crowd groups," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3488–3496.
[126] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: Human trajectory understanding in crowded scenes," in *European conference on computer vision (ECCV)*. Springer, 2016, pp. 549–565.
[127] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine *et al.*, "Scalability in perception for autonomous driving: Waymo open dataset," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 2446–2454.
[128] G. Awad, A. Butt, K. Curtis, Y. Lee, J. Fiscus, A. Godil, D. Joy, A. Delgado, A. Smeaton, Y. Graham *et al.*, "Trecvid 2018: Benchmarking video activity detection, video captioning and matching, video storytelling linking and video search," in *Proceedings of TRECVID 2018*, 2018.
[129] B. Zhou, X. Wang, and X. Tang, "Random field topic model for semantic region analysis in crowded scenes from tracklets," in *CVPR 2011*. IEEE, 2011, pp. 3441–3448.
[130] J. Shao, C. Change Loy, and X. Wang, "Scene-independent group profiling in crowd," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2219–2226.
[131] B. Benfold and I. Reid, "Stable multi-target tracking in real-time surveillance video," in *CVPR 2011*. IEEE, 2011, pp. 3457–3464.
[132] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang, "The apolloscape dataset for autonomous driving," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 954–960.
[133] W. Zhan, L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kummerle, H. Konigshof, C. Stiller, A. de La Fortelle *et al.*, "Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps," *arXiv preprint arXiv:1910.03088*, 2019.
[134] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 621–11 631.
[135] J. F. P. Kooij, N. Schneider, F. Flohr, and D. M. Gavrila, "Context-based pedestrian path prediction," in *European Conference on Computer Vision*. Springer, 2014, pp. 618–633.
[136] H. Peng, H. Wang, B. Du, M. Z. A. Bhuiyan, H. Ma, J. Liu, L. Wang, Z. Yang, L. Du, S. Wang *et al.*, "Spatial temporal incidence dynamic graph neural networks for traffic flow forecasting," *Information Sciences*, vol. 521, pp. 277–290, 2020.
[137] S. Kreiss, "Deep social force," *arXiv preprint arXiv:2109.12081*, 2021.



Chi Zhang received the B.E. and M.E. degrees in control science and engineering from Zhejiang University, China, in 2014 and 2017, respectively. From 2017 to 2020, she was a research and development engineer at the Intelligence Driving Group, Baidu, Beijing, China, in the area of automated driving perception. Since 2020, she has been pursuing the Ph.D. degree with the department of Computer Science and Engineering, University of Gothenburg, Gothenburg, Sweden. She is a Marie Curie Early Stage Researcher. Her research interests include applying deep learning on pedestrian behavior prediction and learning interactions between vulnerable road users and vehicles.



Christian Berger Dr. Christian Berger is Full Professor at the Department of Computer Science and Engineering at University of Gothenburg, Sweden and received his Ph.D. degree from RWTH Aachen University, Germany in 2010. He coordinated the project for the vehicle "Caroline", which participated in the world's first urban robot race 2007 DARPA Urban Challenge Final. He co-led the Chalmers Truck Team during the 2016 Grand Cooperative Driving Challenge (GCDC), and is one of the two leading architects behind OpenDLV (Open Driverless Vehicle). His research expertise is on architecting complex and distributed realtime software systems, micro-services for cyber-physical and IoT-systems, continuous integration/deployment/experimentation, and data-driven software engineering.

The Journal of Systems & Software 205 (2023) 111810

Contents lists available at ScienceDirect

The Journal of Systems & Software

journal homepage: www.elsevier.com/locate/jss

Human factors in developing automated vehicles: A requirements engineering perspective☆

Amna Pir Muhammad [a,*], Eric Knauss [a], Jonas Bärgman [b]

[a] Department of Computer Science and Engineering, Chalmers University of Gothenburg, Gothenburg, Sweden
[b] Department of Mechanics and Maritime Sciences, Chalmers Technical University, Gothenburg, Sweden

ARTICLE INFO

Article history:
Received 13 September 2022
Received in revised form 13 April 2023
Accepted 24 July 2023
Available online 28 July 2023

Keywords:
Requirements engineering
Automated vehicles
Agile
Human factors

ABSTRACT

Automated Vehicle (AV) technology has evolved significantly both in complexity and impact and is expected to ultimately change urban transportation. Due to this evolution, the development of AVs challenges the current state of automotive engineering practice, as automotive companies increasingly include agile ways of working in their plan-driven systems engineering—or even transition completely to scaled-agile approaches. However, it is unclear how knowledge about human factors (HF) and technological knowledge related to the development of AVs can be brought together in a way that effectively supports today's rapid release cycles and agile development approaches. Based on semi-structured interviews with ten experts from industry and two experts from academia, this qualitative, exploratory case study investigates the relationship between HF and AV development. The study reveals relevant properties of agile system development and HF, as well as the implications of these properties for integrating agile work, HF, and requirements engineering. According to the findings, which were evaluated in a workshop with experts from academia and industry, a culture that values HF knowledge in engineering is key. These results promise to improve the integration of HF knowledge into agile development as well as to facilitate HF research impact and time to market.

1. Introduction

The term automated vehicles (AVs) refers to an emerging technology that increasingly automates driving tasks and decision-making in transportation (Erdal, 2018). The society of automotive engineers (SAE) has defined six levels of automation (0–5) (sae.org, 2021), starting from *no automation* at Level 0. Many automation features of Levels 1 and 2 (providing one or more automated driving assistance systems (ADAS) to the driver of the car) are already available to consumers. Level 3 features such as lane changing (Yu et al., 2018), steering control, and car parking (Wu et al., 2019) are becoming more common. Level 4 is known as high automation, and there are very few companies that have deployed Level 4 vehicles in real traffic (Waymo (Schwall et al., 2020) is one example). However, several companies are promising Level 4 deployment (Anderson, 2020), and prototypes of Level 5 vehicles (full automation that does not require human intervention and can perform driving under all circumstances) are under development.

Thus, the number of vehicles with medium to high levels of automation are increasing; according to Litman, half of all new vehicles will be autonomous (which the author defines as automation Levels 4 and 5) by 2045 (Litman, 2021). As the number of AVs is increasing, so does the number of reported failures. Although fatal crashes of Teslas have been well publicized, Banks et al. (2018), Deaths (2020), Anon (2019), failures of AV technology are not limited to a single brand; for example, a pedestrian was killed by an Uber self-driving car in 2018 (Kohli and Chadha, 2019).

These examples, as well as more recent ones reported in scientific journals (Morando et al., 2021; Inagaki and Itoh, 2013) and the media (Anon, 2021a,b), show how human over-trust in and over-reliance on automated systems can cause fatal failures of AV. Clearly, even if an engineered, automated solution works perfectly in theory, human factors (HF) must be accounted for to ensure perfect functionality on the roads. As a research field, HF considers humans' physical, physiological, social, and cognitive capabilities and limitations while designing a system (Human Factors and Ergonomics Society, 2021). Expanding on this characterization, several definitions of HF are available, depending upon the context (Human Factors and Ergonomics Society, 2021). As part of our study, we extended one of these definitions to enable us to be more precise about HF in relation to AV (see Section 4.1).

☆ Editor: Heiko Koziolek.
∗ Corresponding author.
E-mail addresses: amna-pir.muhammad@cse.gu.se (A.P. Muhammad), eric.knauss@cse.gu.se (E. Knauss), jonas.bargman@chalmers.se (J. Bärgman).

https://doi.org/10.1016/j.jss.2023.111810

Several HF researchers have emphasized the need to consider HF knowledge during AV development (Hancock, 2014, 2017; Lee, 2008; Navarro, 2019). For example, Hancock states that attention must be paid to the proper design of new vehicle automation technologies and warns that with the breakneck speed at which automated and autonomous systems are developing, HF perspectives might be overlooked (Hancock, 2017). According to Lee, HF aspects must be considered in order to increase the safety, trust, and acceptance of automated technology, as well as to avoid its misuse and disuse (Lee, 2008). Currently, companies are trying out different ways to manage the integration of HF knowledge into their research and development (R&D).

In addition to the changes urged by HF researchers, agile development approaches to system engineering are also being introduced to AV R&D organizations (Kasauli et al., 2020). While initially agile approaches were focused on small software development teams (Beck, 1999; Meyer, 2014; Kahkonen, 2004), their success has led to their adoption in the development of large-scale (Dikert et al., 2016; Lagerberg et al., 2013; Salo and Abrahamsson, 2008) and mechatronic systems (Gren and Lenberg, 2020), where non-agile, plan-driven, and stage-gate-based processes have been the norm (Pernstål et al., 2012). While in practice, the integration of agile practices into large-scale systems engineering may look like a hybrid approach (Klünder et al., 2017), our case companies report themselves that they are transitioning or have transitioned to large scale-agile frameworks, such as SAFe (Knaster and Leffingwell, 2017a).

The agile ways of working adopted by these companies are primarily based on the scaled agile framework (SAFe) (Knaster and Leffingwell, 2017a), which promises to provide "proven, integrated principles, practices, and competencies for achieving business agility using Lean, Agile, and DevOps". SAFe suggests distinguishing a number of abstraction levels, including the lowest level teams, a middle layer where different *solution trains* (a group of teams working on a coherent part of the product) are managed, and a portfolio level on top. Due to their iterative nature, agile approaches are suitable for building systems whose requirements may change; further, experience from early versions of a system can impact later versions (Beck, 1999; Meyer, 2014; Gren and Lenberg, 2020). Thus, in theory, agile approaches are well suited to the introduction of stakeholder concerns (such as those provided through HF knowledge) in automation development: Agile often reveals previously unforeseen requirements for a system under development, such as considerations of HF.

To integrate human factors during system development (in general, not specific to agile), researchers advocate incorporating human factors knowledge already into the early stages of development (Calp and Akcayol, 2019; Chua and Feigh, 2011; Håkansson and Bjarnason, 2020). Traditionally, such information is usually included in system requirements, which are defined up front and serve as the basis for any subsequent development work. The process of eliciting, analyzing, describing, and validating requirements is called requirements engineering (RE) (Wiegers and Beatty, 2013). To date, it has been particularly challenging to apply RE to the agile development of systems at scale (Meyer, 2014; Kasauli et al., 2021). Meyer highlights the rejection of upfront analysis as particularly problematic (Meyer, 2014), but other challenges exist, particularly with managing and communicating requirements-related knowledge at scale (Kasauli et al., 2021).

The literature (Hancock, 2014, 2017; Lee, 2008; Navarro, 2019; Beauchamp, 1986) leaves no doubt about the importance of considering HF in AV development. For example, an AV at Level 3 still requires humans to be able to take over control of the vehicle. Especially when it comes to switching control between the human and vehicle, human factors such as reaction time, comfort, fatigue, and understandability must be considered as requirements (Gold et al., 2017). Yet, particularly in the light of well-known challenges for RE in scaled agile system development, it is unclear how to ensure their consideration. There is a lack of empirical research on how to integrate HF aspects of vehicle automation development and communicate such requirements to AV engineers[1] In this study, we distinguish between HF experts and AV engineers[2] in order to clarify how HF experts are currently communicating with AV developers and identify any communication gap, particularly during agile development. This is a relevant research gap with practical implications: Automotive companies are moving towards scaled-agile system development. It is unclear how to introduce HF requirements into agile system development, which is the traditional way of managing knowledge in the development lifecycle. Thus, it is unclear how to ensure that HF knowledge should best be integrated into agile system development, and practitioners struggle with a lack of clear guidelines.

We investigate this research gap in this exploratory qualitative study. Within the general research goal of determining how HF aspects of AV development can be integrated in agile AV development, *this study specifically aims to investigate how HF knowledge as requirements can be integrated in the development process and communicated to AV developers in the context of large-scale agile AV development.* The research goal is operationalized by addressing the following research questions (RQs):

RQ1: How do HF experts and AV engineers characterize HF in relation to AV development?

RQ1 is motivated by the broad spectrum of definitions offered by literature. In order to understand how HF aspects can be communicated, we first need to establish a working definition of HF in terms of AV development. We then explore the relevant properties of HF and agile work in RQ2 to lay the foundation to reach our research aim.

RQ2: Which properties of HF and agile ways of working impact AV development?

In RQ3, we are particularly interested in implications for agile ways of working, HF work, and managing requirements in AV development:

RQ3: What are important implications when aiming to better integrate HF into AV development?

This work answers these research questions by qualitatively analyzing interviews with ten experts (HF experts and AV engineers who work in the automotive industry), complemented by two additional interviews with academic leaders in the field of human factors. Our results indicate that an important property of scaled agile is its way of working, which advocates responsiveness to change by shifting responsibility from managers who

[1] We recognize that many HF experts can also be considered engineers in terms of AV development (the domain of HF engineering).

[2] Note that in this work, HF experts are individuals in an organization that typically have formal training in Human Factors (e.g., often with a background in psychology, behavioral or cognitive sciences) and who have a role in the organization where he or she on a daily basis works with HF aspects (i.e., works with HF related topics; for details of what is meant by that, see the definition of Human Factors in Section 2.1. Further, in this work, an AV engineer is typically a software, electrical, or mechanical engineer, whose work is to develop the AV from a technical perspective, and that does not have an HF background. More precisely, when referring to AV engineers in this paper, we specifically exclude HF engineers (Wickens et al., 2003), i.e., professionals that have a background both in HF and engineering, who, for the purpose of this study are categorized as HF Expert.

plan at the system level to autonomous teams that make local decisions. To support such local decisions, it follows that HF knowledge should be available to the agile teams to raise awareness, enable asking relevant questions, and guide them in the right direction. It also follows that agile AV teams should be able to produce HF knowledge on demand, e.g., by conducting HF experiments within their team's iterative work; further, RE should provide methods for effectively managing the knowledge gained from the experiments. We validated these findings in a workshop setting using a survey questionnaire, as well as in discussions with 28 expert participants from industry and practice. The evaluation study confirms that our findings are very relevant to the industry.

The paper is divided into seven sections. This introduction, Section 1, is followed by Section 2, which provides the background and reviews related work; Section 3 discusses the research methodology. The main findings are presented in Section 4. Section 5 presents the outcome of the survey performed to evaluate the findings of this study. In Section 6, we discuss our findings. Finally, Section 7 concludes the paper.

2. Background and related work

The research presented in this paper is interdisciplinary, targeting both systems and software engineers as well as HF experts. Therefore, this section provides the background on which the argument of the exploratory qualitative analysis is built. This background may seem obvious and basic in parts. However, since the targeted readers belong to many disciplines, some basics need to be explained for completeness: many HF experts are not familiar with the agile way of working or RE, and many AV engineers are not familiar with the domain of HF.

2.1. Human factors in automated vehicle development

Human factors are an integral part of the development of road transport (Wickens et al., 2003). However, as the definitions of HF are many and diverse (Licht et al., 1991; Human Factors and Ergonomics Society, 2021), there may be a problem when people with different definitions are communicating requirements and knowledge (Licht et al., 1991). Taking a scientific perspective of the definition of human factors,

The Journal of the Human Factors and Ergonomics Society describes the science of human factors as pursuing "fundamental knowledge of human capabilities and limitations—and the basic understanding of cognitive, physical, behavioral, physiological, social, developmental, affective, and motivational aspects of human performance" as a means "to yield design principles; enhance training, selection, and communication; and ultimately improve human-system interfaces and sociotechnical systems that lead to safer and more effective outcomes".[3] Although this definition may seem clear and concise, individuals may have different views of what HF entails (Licht et al., 1991), and their views may impact how they consider HF in their profession. Thus it is important, when studying how HF is considered in the workplace, to investigate what their views of HF actually are. This may be particularly important when the subjects in a study have very different backgrounds, such as when studying the role of HF in the development of automated vehicles (as in the current study); HF experts, as well as a range of different engineers, are involved (Wickens et al., 2003). As a consequence, developing a precise definition related to a specific topic (here AV design) is warranted.

Finally, to help readers that are not HF experts get a better grasp of HF, some examples in the field of AV development are listed below. As the HF domain is very broad, also this list is highly diverse and only represent a small fraction of all HF aspects considered in AV design. Its aim is only to provide some insight into AV HF considerations. HF knowledge helps to answer questions on how to design and develop…:

- AVs that are predictable and safe for other road-users
- AVs that users trust (to a reasonable degree)
- AVs that are transparent with their capabilities, avoiding over-reliance
- AVs that drivers like
- human–machine interfaces (HMI) for AV users (e.g., touch screens for adjusting settings) that are safe, user-centered and in-line with the company branding.
- HMIs for other road users (external HMIs) to, for example, communicate state and intent
- AV motions that the users like (e.g., to make them feel comfortable with the AV speed and acceleration, as well as relative speeds and ranges to other road users and infrastructure features)
- AVs that ensure intended effects of the AV functions are reached by considering user's and surrounding road users intent and actions
- AVs' auditory, visual, and haptic information exchange with their users (e.g. as information and warnings) including braking, active vehicle steering and acceleration through actuators
- models of human behavior to use in virtual simulations to assess AV safety

Human factors for AV development include all considerations of the human in the AV design. It does not include the development of hardware and software in general, but many tasks that typically are considered "hard core engineering", such as path planning, has clear HF aspects in them (see list above). Consequently there may be HF requirements on the sensing or actuator system and other AV engineering that may have HF implications (e.g., sensing and actuation needs to provide the path planners the means to navigate in a way that is acceptable to the AV users).

2.1.1. Human factors and its role in AV development

In AV development, HF relates to aspects of both software development and physical AV design. Examples of HF aspects in AV development are many. Note that a common misconception by many non-HF experts is that HF is simply a list of factors, while it is actually a range of aspects that affect humans, or that humans affect (see, e.g., the definition by the Journal of Human Factors). Physical aspects range from seating ergonomics (as AVs are impacting vehicle interiors (Salter et al., 2019)) to the physical design and placement of human–machine interfaces (HMIs). Typically, humans are directly affected by software aspects of HF, including: how and when the (software-based) HMIs display information (Carsten and Martens, 2019), how external road users are to be communicated with (Ackermann et al., 2019a; Faas et al., 2020), how the vehicle stays in the lane (Xu et al., 2017; Miller and Boyle, 2019), how it keeps its distance from a lead vehicle (De Winter et al., 2014; Reagan et al., 2017; Morando et al., 2016), how it overtakes other road users (Abe et al., 2017; Kovaceva et al., 2019), how humans and AVs communicate (Ackermann et al., 2019b), and how AVs can avoid driver over-reliance on the AV performance and ensure that the trust in the AV is properly calibrated (Mirnig et al., 2016; Kraus et al., 2020). These examples highlight the extent to which successful engineering depends on HF knowledge. Yet it remains an open question how engineers gain awareness of HF in their daily work and design decisions.

3 https://journals.sagepub.com/aims-scope/HFS

2.1.2. What HF issues impact AV development?

Kyriankidis et al. highlight that as AV development in the industry keeps moving forward at a fast pace, the gap between research in academia and R&D in the industry continues to grow (Kyriakidis et al., 2019b). They stress the importance of more research on the interconnection of AVs with other road users, human trust in and acceptance of AVs, and how much (and which) information drivers will get and should be getting from AVs. The authors also discuss the need for more experiments to study how humans interact and control transitions between the driver and the AV. Similarly, Noy et al. (2018) argue that the benefits of AVs (such as safety) can only be achieved if they are designed according to standards of human–system integration. The importance of integrating HF into the design and evaluation processes of autonomous vehicles to increase their safety and trust is also highlighted in this position paper (CARTRE, 2018) and in the book by Wickens et al. (2003).

The work by Saffarian et al. (2012) lists six specific issues regarding HF in AV development: overreliance, behavioral adaptation, erratic mental workload, skill degradation, reduced situation awareness, and inadequate mental models of automation functions. The authors proposed a solution for these issues specific to CACC (Cooperative Adaptive Cruise Control), as well a proposing a mechanism of interaction between humans and CACC. However, the solution simply proposed a few different modes to keep the driver in the loop and facilitate cooperation between driver and vehicle.

Chen et al. (2018) describe the importance of transparency between intelligent systems (e.g., robots or AVs) and humans. The authors developed a Situation awareness-based Agent Transparency (SAT) model to ensure an appropriate interplay between AVs and humans. Their study mainly targets human drivers' need for transparency of AV functionality in order to promote better understanding, trust, and interaction.

For each individual HF issue encountered during the AV development process, involved engineers may lack the experience or competence to include the appropriate HF aspects. However, no one can know everything. Communication about HF among stakeholders is therefore crucial. The AV development process must include many stakeholders from different domains, making it interdisciplinary.

2.2. AV development: Processes, approaches, recent developments

In the automotive sector, the R&D required to create cars and trucks and offer related services is a complex affair, involving many disciplines such as mechanics, electrical hardware, and (increasingly) software. Whereas electronics and software in cars were originally introduced simply to optimize engine control, their development now drives 80% to 90% of the innovation in the automotive industry.[4] This subsection provides an overview of AV development in the context of requirements engineering (RE).

2.2.1. Requirements engineering

International standardization and certification bodies provide valuable insights into the fundamental concepts of requirements engineering. The IEEE defines a requirement as either (i) a condition or capability needed by a user to solve a problem or achieve an objective; (ii) a condition or capability that a system or component must meet to satisfy a contract, standard, specification, or other formally imposed documents; or (iii) a documented representation of a condition or capability as in (i) or (ii) (IEEE, 1990). The International Requirements Engineering Board (IREB) describes requirements as representations of the needs and desires of customers and users for new things to be built or old things to be upgraded (International Requirements Engineering Board, 2020). Accordingly, requirements can be of three types: functional (a result or behavior to be provided by a function), quality (a quality concern not covered by functional requirements, such as performance, availability, security, or reliability), and constraint (a further limitation on valid solutions beyond what is necessary to fulfill functional and quality requirements). IREB characterizes Requirements Engineering as specifying and managing "requirements for systems such that the systems implemented and deployed satisfy their stakeholders' desires and needs" (International Requirements Engineering Board, 2020).

Activities of RE typically include elicitation, analysis, specification, validation, and management of requirements (Nuseibeh and Easterbrook, 2000). In addition, requirement prioritization becomes a key RE activity in agile development, supporting elicitation and analysis by identifying the requirements with the highest stakeholder value (Heikkilä et al., 2017). Research emphasizes the interdisciplinary aspects of requirements engineering (Nuseibeh and Easterbrook, 2000); however, we are not aware of any works that explore how HF research can be integrated into requirements engineering activities for agile system development at scale.

2.2.2. Development practices

Traditionally, the automotive environment has been characterized by long lead times (Berger and Eklund, 2015) and stable, sequential engineering practices (Pernstål et al., 2012). Eklund et al. (2014) argue that the industry is currently transitioning from plan-driven, stage-gate processes (Pernstål et al., 2012) to more value-driven, continuous approaches (Knauss et al., 2016; Fagerholm et al., 2017a) (often referred to as agile methods (Meyer, 2014) or agile transformation (Paasivaara et al., 2018)). Gren and Lenberg argue that the main motivation for such a transformation is to be able to respond to changing requirements (Gren and Lenberg, 2020).

Agile methods have traditionally been proposed for small teams (six to eight developers) (Beck, 1999; Schwaber and Beedle, 2001; Meyer, 2014). The core values of agile methods as described in the agile manifesto Beck et al. (2001) are: Focusing on individuals and interactions to develop working software in close collaboration with customers with an emphasis on embracing change while de-emphasizing processes, tools, extensive documentation, contract negotiation, and following plans. In fact, agile methods have been presented as the antithesis of previous plan-driven approaches. In its original form, an agile team would take notes about customer needs in the form of user stories on small index cards. Often, these are described as boilerplate statements: "As a <role> I want <feature> so that < value >" (Leffingwell, 2010). The much more detailed requirements of plan-driven approaches are omitted; instead, agile methods push for a continuous dialogue with customer representatives or product owners and comprehensive sets of tests, which are ideally automated (Meyer, 2014). On the other hand, agile methods have been criticized for limiting requirements engineering to functional requirements described through (exemplary) scenarios and discouraging upfront planning (Meyer, 2014).

2.2.3. Development approaches at scale

Automotive R&D work is typically a collaboration between an OEM (Original Equipment Manufacturer) and suppliers in several tiers. The OEM owns the vehicle brand and orders mechanical, electrical, and software components from suppliers. Thus, the ability to specify requirements for the vehicle and break them down into component specifications is a core competency for an OEM.

[4] According to industry experts: https://tinyurl.com/y9jnoupd.

In order to improve their responsiveness to changing requirements, OEMs have started to bring more development in-house and to identify new collaboration models with suppliers (Hohl et al., 2016; Van Der Valk et al., 2018). As a result, OEMs struggle to maintain effective ways of structuring, documenting, and managing requirements for increasingly complex systems (Liebel et al., 2019; Kasauli et al., 2020). While software teams may have quickly learned to adopt agile methods, company-wide adoption is usually slow, mostly due to skepticism (Lindvall et al., 2004). Thus, new ways of managing requirements must be conceived for OEMs and their supplier value chains (Kasauli et al., 2021).

Moreover, for complex products such as cars, it is important to scale agile methods beyond individual teams, since if the overall plan for the complete vehicle cannot be changed there is limited value in an individual team's ability to respond to change (Gren and Lenberg, 2020). SAFe is the most commonly used framework for scaling agile (Knaster and Leffingwell, 2017b), especially in the automotive domain Kasauli et al. (2021). SAFe describes a requirements information model that groups several user stories into epics. Epics can then describe mid-to-long-term goals for groups of teams. The model also describes non-functional requirements as a way to present quality requirements as constraints for user stories and epics (Leffingwell, 2010).

Previous works have described inadequacies in the SAFe framework (Kasauli et al., 2021) and its requirements information model (Wohlrab et al., 2020). Of particular relevance to this paper is the fact that scaled-agile methods struggle to provide alignment among many software teams (Kasauli et al., 2021; Wohlrab et al., 2020); we need to consider the effects of scaling agility beyond individual software teams since questions about agile ways of working must be part of our exploration of HF. For example, for a given automated driving function, several teams must align on how to address HF. For brevity, hereafter we refer to *scaled agile* or *large scale agile* simply as *agile*.

2.3. Related work: communicating human factors knowledge and requirements to AV engineers

Interdisciplinary communication is often difficult. However, many fields such as aviation, transportation, and medicine, acknowledging the importance of HF knowledge, have worked to integrate HF design principles and techniques into the design and development of products and systems. Vincent et al. (2014) suggested that the communication gap between HF knowledge experts and other developers is due to a lack of common ground; they proposed the use of mediating representations of boundary objects (Star and Griesemer, 1989) for effective communication. Bruseberg (Bruseberg, 2008) introduced a novel methodology that feeds HF knowledge into an architectural framework. However, the author mainly discusses HF from a cognitive perspective. Alternatively, Chua and Feigh (2011) suggest including HF in an early design stage. While HF can provide significant input to improve the communication between HF experts and system engineers, it is unclear exactly how to include HF knowledge in these stages of development. Other authors (Bodenhamer, 2012; Ramos et al., 2012; Orellana and Madni, 2014; Watson et al., 2017) advocate including HF in system design via SysML, using activity, block, and sequence diagrams.

van Maanen et al. (2005) have discussed how HF can be integrated with AI for better human–machine cooperation (HMC). Whereas current customization is limited to static interfaces, improved HMC could provide customized support to users. However, knowledge about both HF and artificial intelligence (AI) and how to integrate them is lacking. To bridge this knowledge gap, van Maanen et al. (2005) have proposed a methodology based on interdisciplinary cognitive engineering (CE+). In CE+, HF experts provide the relevant information (such as the support concepts and rules) and strategies for the specification and evaluation of HMC. The authors concluded that HF and AI must be integrated into the early stages of the development process. In fact, the User Centered Ecological Interface Design (UCEID) (Revell et al., 2018) method proposes a combination of techniques (e.g., data collection and task and cognitive task analysis) to include HF considerations in the early stages of the overall system design processes. The main finding of UCEID is that it is important to meet the dual requirements of demographically diverse clients and technology delivery. It remains unclear how these requirements can be integrated into the (agile) development cycle; however, considering the importance of the issues mentioned above, a way must be found to design AVs with HF in mind (Merat and Lee, 2012; Kyriakidis et al., 2019a).

Adopting this design practice proves to be challenging, not the least because of the adoption of agile development (Mehrfard et al., 2010). Processes have become more iterative, putting more emphasis on a continuous understanding of requirements. It is unclear how the above-mentioned methodologies would work for the communication of HF knowledge in today's large-scale agile AV development. For example, Kashfi shows how difficult it is to align user-centered design and UX in agile development (Kashfi, 2018).

In summary, although communicating HF knowledge to engineering teams is challenging, research provides ample motivation to explore how this challenge can be overcome in practice. To our knowledge, no systematic approach exists that make sure that HF are adequately represented in agile system development.

3. Research method

Our exploration of the role of HF in developing automated vehicles is widely based on the epistemological stance of critical realism (Lawani, 2020), a research philosophy that distinguishes between the 'real' world and the 'observable' world. With respect to this study, we made this distinction by observing and analyzing expert opinions about how HF aspects are addressed in engineering, rather than assuming that we can analyze those aspects directly. Critical realism relies on a common ontology or sociological theory, which we provide through our detailed assumptions about the role of HF, RE, and agile methods based on related work in Section 2. In our study design, however, we were also inspired by the school of pragmatism, focusing on particular causalities of pragmatic relevance (i.e., the implications that follow from particular properties of agile AV development and HF). As discussed by Lawani (2020), pragmatism and critical realism are often associated with each since both advocate the use of mixed methods and on understanding (causal) relationships that are thought to be not directly observable. In fact, we added continuously to our knowledge as we learned new items that did not previously fit into our mental model. Given this mix of epistemological stances, we decided that an exploratory, qualitative inquiry was the most appropriate to address our research questions (Creswell and Creswell, 2017).

Our case consists of a number of automotive companies, including manufacturers and suppliers, collaborating not only within the value chains needed for building automated vehicles but also beyond, to build and maintain excellence in the field. We relied on semi-structured interviews with experts to provide the primary data, since we were specifically interested in applying the personal views of experts in the field (who collaborate within and across value chains and concrete products) to chart the landscape of HF in relation to AV development.

In this section, we describe the collection and analysis of the data.

Table 1
Interviewees' roles and work experience (Experience level: Low = 0–5 years, Medium = 5–10 years, High = More than 10 years).

ID	Role	Experience Level
S1	HF Expert (Specialist)	High
S2	HF Expert (Strategy, Specialist & Research)	High
S3	AV Engineer (Strategy & Architecture)	High
S4	AV Engineer (Requirements & Research)	Medium
S5	HF Expert (Management & Research)	High
S6	HF Expert (Specialist)	High
S7	HF Expert (Specialist & Design)	High
S8	AV Engineer (Safety & Research)	Low
S9	AV Engineer (Strategy & Specialist)	High
S10	HF/AV Engineer	High
*Special	Interviews	
S11	HF Expert (Specialist)	High
S12	HF Expert (Specialist)	High

3.1. Data collection

Our strategy for recruiting interviewees for our study (Palinkas et al., 2015) relied primarily on convenience sampling. That is, we aimed to identify interviewees who possess the relevant expertise and were willing to participate. Our results confirm that such experts are rare among companies, and that it is important to protect their time. In recruiting interviewees, we relied both on the personal and professional networks of the authors, built through years of research with participating companies, and on recommendations from the interviewees themselves. We aimed for a mix of similarity and variation in our sampling in order to cover different perspectives (HF vs. engineering and OEM vs. supplier) in sufficient depth.

As a result of these efforts, we interviewed ten experts from five Swedish companies: four from Volvo Cars, two from Volvo Trucks, two from Zenuity, one from Veoneer, and one from Autoliv. In addition, we conducted two more complementary interviews with international academic leaders in the field (S11 and S12), to get additional perspectives on the definition of HF and emerging themes. All of the industry interviewees have been working with AV companies for years, often more than ten (see Table 1). The experience level of the participants is classified as *low* if they have less than five years of experience, *medium* if they have between five to ten years, and *high* if they have more than ten years. Experience level is the sum of all jobs for each participant, which can span one company or more. However, from talking to them, we can infer that, even though they changed companies, their roles and companies were similar in the same domain, and their experience is well in line with both the needs for their individual job roles and this interview study.

In Table 1, S11 and S12 are separated from the other participants because these interviews were conducted in a slightly different style, and the preliminary results from the other interviewees were kept in mind.

We relied on semi-structured interviews because they are especially suitable for exploratory studies (Creswell and Creswell, 2017): depending on the course of the interview, questions can be adjusted to mitigate the risk of asking the wrong questions, and follow-up questions can be created to satisfy emergent information needs. This approach allowed the participants to articulate their individual and valuable views, concerns, and expectations. Consequently, interviews tended to resemble guided discussions and were engaging both for interviewees and interviewers.

Each interview took between 60 and 80 min. In most interviews, all three authors were present; at least two authors were present in every interview, which allowed us to keep extensive, often verbatim, notes. The second author took notes, and the first author conducted the interview. The second and third authors have extensive experience in the automotive industry, having been involved in various research projects over the years. The second author has particular expertise with RE and Agile and the third author has formal training as an engineer, but has worked between HF and engineering for many years. Given their multidisciplinary background, they was there to ask follow-up questions and provide clarification.

Notes ranged from 700 to 1750 words and contained, on average, 1325 words. We did not record the interviews. We did, however, show our notes to the interviewee during the interview. While we were interested in the perspectives of experts on the role of HF knowledge in AV development, the discussion could have touched on examples of perceived or real shortcomings in processes, which would be very sensitive information. It was thus deemed better not to record the interviews; after a sensitive discussion, any such content was eliminated from the meeting notes or, if necessary, more suitable examples/formulations were substituted. Rutakumwa et al. argue similarly to us that the context can have a negative impact on the quality of answers when recording (Rutakumwa et al., 2020). They also indicated that there is not necessarily a negative impact on the quality of the transcript in relation to its role in the thematic analysis (Rutakumwa et al., 2020).

Before the interviews, we prepared a guide[5] to help us cover the same topics in each interview. Each interview included the nine open-ended questions and detailed follow-up questions contained in the guide. We designed the interview guide with the intention of getting an HF perspective on the design and development of AV technology. The questions in the interview guide are based on our literature review and experience. This includes assumptions of research gaps, as apparent in question three. Studies such Hancock (2019), Wickens et al. (2003), Navarro (2019) clearly indicate that problems of HF are not properly addressed. With respect to our experience, the authors are currently involved in a project (SHAPE-IT, 2023) with many senior human factors experts. Discussions with those experts clearly indicate substantial gaps in the integration of HF in AV development. Although some of the questions posed during the semi-structured may have influenced the interviewers, this is a common risk with this interview format since it involves open-ended conversations. To mitigate this risk, we tried to include three interviewers to capture the conversation and minimize any potential biases that could have arisen during the interviews. Consequently, we believe that the level of risk posed by these potential biases is negligible.

The map between the interview guide questions and the research questions is shown in Table 2.

3.2. Data analysis

In order to analyze the data obtained from the interviews, we relied on the common set of principles (Noble and Smith, 2014) used for qualitative analysis of interview data. Specifically, these principles include: transcribing the interviews, familiarizing ourselves with the data to attain a deep understanding of the phenomena being investigated, coding, generating initial themes, and finalizing the themes and overarching concepts.

[5] We provide the interview guide as well as an overview of our themes in relation to codes and example quotes as data set at Zenodo, DOI: 10.5281/zenodo.5562487.

Table 2
Interview questions, mapped to the research questions.

Interview questions	Research question(s)
1. Background of interviewee (Demographic Data) • What is your role? • What is your experience in that role? • What is your experience with HF/Requirements? Reminder: We will take notes during the interview, which we will send later for confirmation.	Demographics
2. How would you characterize what HF is and how it relates to requirements for AV development (or AI-based systems)?	RQ1
3. In your experience, how does engineering work with or without HF? What is missing?	RQ2
4. How does HF knowledge come to engineers?	RQ3
5. What are the main challenges in conveying requirements from HF to engineers that design automated vehicles (or from engineers to HF experts)? • Follow-up: what about conveying knowledge from HF/behavior as input into the AI-based AV-design process? • Think about comfort zones as an example, safety aspects, software requirements aspects (e.g., AI based control of the vehicle) compared to traditional physical "user experiences" of AV.	RQ2 & RQ3
6. What scenarios related to AV in urban environment are the most difficult (and/or important) to convey requirements to AV-engineers?	Not used
7. Do you have recommendations on how to optimize communication between human factors experts and engineers of AI-based AVs? Any guidelines for incorporating human factors into AI-based AV design guidelines?	RQ2 & RQ3
8. How should the process (or: way of working) for system design look like? Particularly in agile development how we do that?	RQ3
9. Thanks you for the interview, next steps. • Whom else should we interview? • Anything we forgot to ask?	All

The extensive interview notes were a good starting point for further analysis. To familiarize ourselves with the data, we read the interview notes thoroughly while creating memos describing those ideas that the notes inspired (Birks et al., 2008). Then we highlighted parts of the text related to our research interest and assigned them labels (so-called "codes"). In parallel, we continued to create and discuss memos to capture any noteworthy aspects as they surfaced. For these activities, we relied on both generic word processors (MS Word) and specialized qualitative analysis tools (NVivo[6]). Through these steps, we identified the main ideas as well as common perspectives.

[6] https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

After formalizing and coding the data, we further classified all the relevant codes into candidate themes. For example, the following quotes were coded as "validation test" and "test dilemma", respectively.

> *"Perhaps put more emphasis on validation tests, that is, not only automated tests but also test the quality in use."* — S4 - AV Engineer

> *"I have seen people spend three person-years on things they have never tested with real humans. Then they claim they have never had time to do so."* — S2 - HF expert

By analyzing and categorizing these and other relevant quotes, we came up with a theme called "Testing".

The themes and codes were then re-analyzed to check if there was any missing or extra theme with respect to our interview notes or any mismatch in the code classifications. In this way, we refined the set of themes until all authors agreed that it provided complete coverage of all aspects of the data, without redundancy, on a meaningful level of abstraction.

Finally, we renamed our themes to better align with research questions. Section 4 describes the outcomes of our data analysis

4. Findings

This section presents our findings, with each subsection addressing one RQ. We start by defining HF in AV development, based on our interviews and the literature (RQ1). The second research question focuses on the properties of HF and agile ways of working (RQ2). These properties raise important questions (discussed in our interviews) about the interplay of both disciplines. Then, we present the implications that emerged from these discussions in three themes related to research question RQ3: implications for agile ways of working, implications for HF work, and implications for managing requirements.

For each theme, we start our report of results with a box that shows in which interviews we have identified related codes. Table A.3 presents a comprehensive list of all the themes covered in this paper, along with their corresponding codes. Additionally, we provide an overview of our themes and codes per interview as an external resource, here.

4.1. Human factors in relation to AV development (RQ1)

HF Definition based on codes from interviews with: S1-3, S5-6, S11-12

In order to explore the systematic capturing and managing of human factors in AV development, it is important to share a common understanding of the key concepts. Therefore, our first question aimed to understand each interviewee's perspective on human factors and their relation to AV development. Our interviews show a broad and diverse usage of the term 'human factors', which is also reflected in the literature.

For example, the following quote shows a rather broad definition of the term, assigning responsibility for considering human factors to the complete development cycle:

> *"How to safely develop an AD function (without killing humans in the process) so that in case of a crash, people will say that the car was behaving reasonably."* — S1 - HF Expert

In other examples, interviewees had a more technical, outcome-oriented view of the term and how it feeds into other engineering processes:

"Learning the user preferences, should it be race driving, comfort, safety, or speed." — S3 - AV Engineer

" HF was 2WW system ergonomics, then CS brought up HMI. Those have merged since. You have physical interfaces, but also services, but also how users are adopting new functions and whether or not they continue using. HF and HMI are intertwined. Ergonomics is included and overlaps with the cognitive side, e.g., external communication with other road users. Understanding the warnings and so on." — S6 - HF Expert

Human factors knowledge, such as preferences about the level of comfort, safety, and speed, is instrumental for the development of AV. The role of HF in providing input to design and development is also reflected in another interviewee's quote:

"Understanding the interactions between people and all other elements within a system, and designing in light of this understanding." — S5 - HF Expert

However, considering HF requires more than one-way communication with engineers. As the following quote reveals, HF sets limitations on both engineers and users.

"How to communicate the limitations of behavior so that people understand what they are allowed to do and what they are not allowed to do... [This is easy to do with] HF related to safety. [With other] HF [e.g., those] related to a sense of calm or serenity that is a bit more difficult." — S2 - HF Expert

Given the broad use of the term 'human factors', we aimed to integrate different interpretations from practitioners' perspectives into a definition of human factors in AV development.

As part of this process, we relied on the two international experts to provide more insight. They confirmed that a working definition is indeed needed and might need to be compiled from various sources and then matched with comments from our other interviewees:

"So one definition is from the Journal of human factors, which is about knowledge of human capabilities and limitations. I think that would be good, but there is one on this other page... The goal to design safe, comfortable, effective [systems for] human use is almost describing what you are trying to achieve, so I am just wondering whether you could start by saying this is what we believe HF is and then add more with your work." — S12 - HF Expert

In addition, our international experts confirmed that a good working definition must be related to the engineering cycle

"Understand, create and evaluate cycle. HF plays a role in each component & understand is about identifying requirements, human capabilities, limits, needs & describing those in ways that influence the design." — S11 - HF Expert

"AV has physical considerations regarding how you get in a vehicle, make the seats big enough to accommodate the people, there are certain design issues. But it also that people trust the AV to be safe, how do they perceive the important risk, do they feel comfortable with the algorithms, do the algorithms behave as expected, does it enhance end goals: pleasure, satisfaction, aesthetics? " — S11 - HF Expert

This is a cross-cutting theme that is also visible in the other subjects' quotes above.

In summary, we note that multiple definitions of HF exist, even on the homepages of key journals of the field (e.g., The Journal of the Human Factors and Ergonomics Society, 2020), depending on the specific research context. In our research context, it is crucial to link HF to AV design and development, as well as the development cycle. As suggested (by S12 above, for example), we start from a generic established definition of HF (taken from The Journal of the Human Factors and Ergonomics Society, 2020), and relate it to the development cycle. Fig. 1 represents our working definition graphically: added aspects are shown in green, and the most important aspects from our interviews are underlined (both in the Figure and in the quotes above).

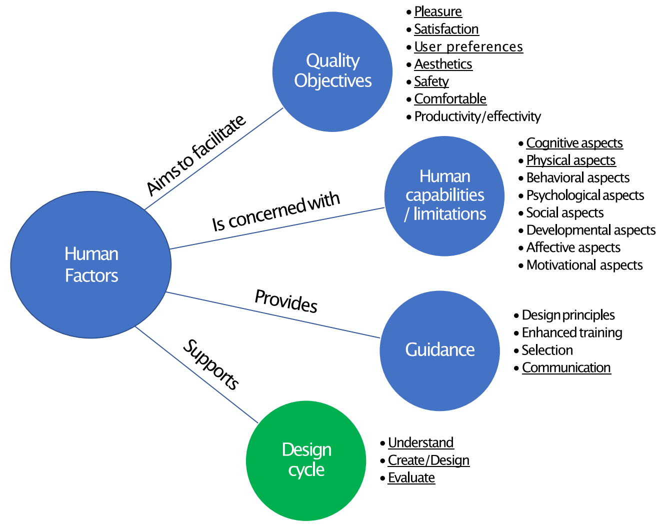


Fig. 1. A mind-map of aspects that define Human Factors in the context of the design and development of automated vehicles.

Definition. The field of *Human Factors in AV Development* aims to inform AV development by providing fundamental knowledge about human capabilities and limitations throughout the design cycle so the product will meet specific quality objectives.

Based on our interviews, we can highlight some critical aspects of this definition that shape the relationship between human factors and agile AV development. Firstly, it is important to relate human factors to *AV development and its product quality objectives*. These objectives usually include an AV design result that is pleasurable, satisfactory, user-preferred, comfortable, aesthetic, effective, and safe for stakeholder interaction (Wickens et al., 2003).

Another component of the definition, *human capabilities and limitations*—which include cognitive, physical, behavioral, psychological, social, affective, and motivational aspects, is the core concern of human factors experts (The Journal of the Human Factors and Ergonomics Society, 2020). It is critical to effectively manage these capabilities and limitations during AV development. Therefore, it is a crucial role of HF in AV development to provide *fundamental knowledge* about human capabilities and limitations and their relation to quality objectives for AV design. Typically, this knowledge is provided in the form of design principles, training, selection, and communication. In this paper, we will focus on the implications of knowledge transfer in the context of agile AV development.

This fundamental knowledge is needed throughout the *design cycle of AVs*. While various design cycles have been proposed, we refer to the phases that Jacobson et al. found to be essential when building software-intense systems (Jacobson et al., 2012): understanding the requirements; shaping, implementing, testing, and evaluating the AV system; and putting the AV system to use. Note that in modern AV development, these phases are iterative and incremental. Relating HF to AV development throughout the design cycle is of paramount importance for discussing the relationship between HF and AV development. Yet, it is missing from many established definitions of HF and therefore highlighted in green in Fig. 1.

Thus, to answer RQ1, we noted that AV development is suffering from the lack of a working definition of HF. From our interviews with industry HF experts, we extracted the core aspects that such a working definition should have and triangulated it with definitions found in the literature. We further validated

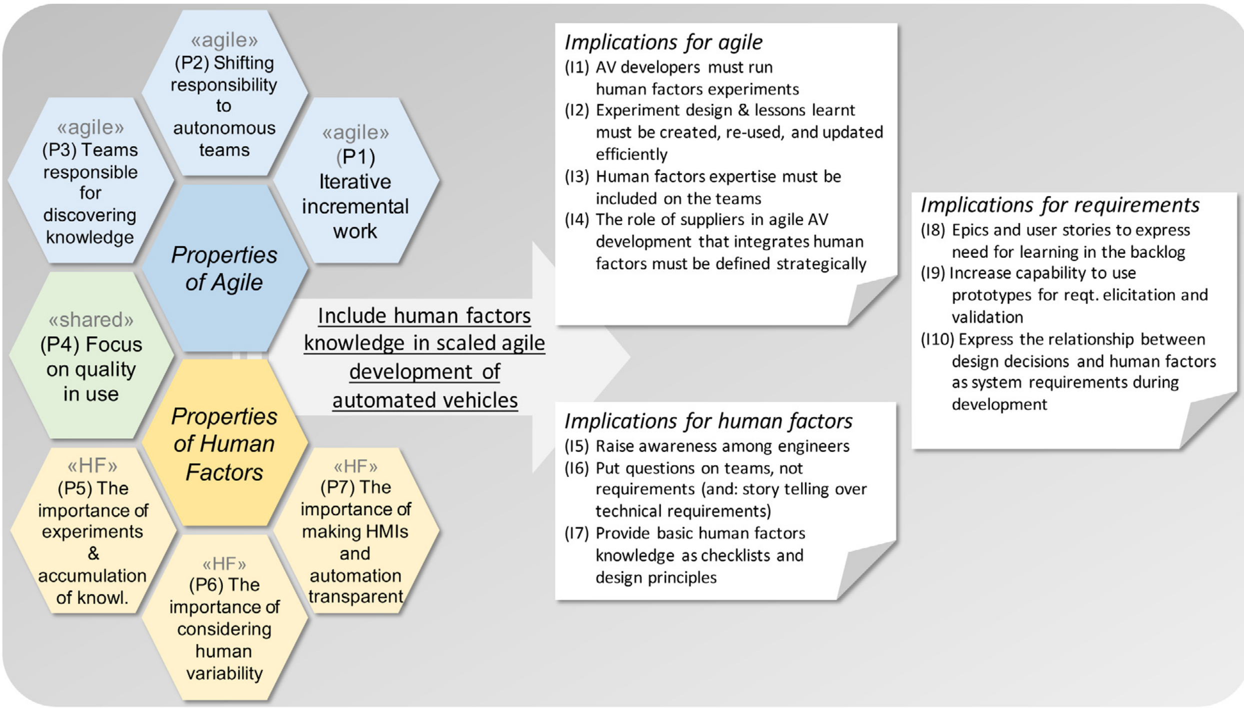


Fig. 2. Taking a requirements engineering perspective, our qualitative study on Human Factors for Automated Vehicles revealed themes relating to properties of human factors and agile system development, as well as implications for human factors and agile system development and requirements.

our suggested definition with interviewees S11 and S12. Thus we have established a common language for addressing RQ2 and RQ3.

4.2. Properties of human factors and agile (RQ2)

In order to lay the foundation for improving the way that HF knowledge and development work are integrated into agile AV development, we first focus on the properties of HF and agile ways of working. We provide an overview of our findings for RQ2 in the left part of Fig. 2.

4.2.1. Properties of agile.

When we started our investigation, we were aware of the role of agile in transforming companies and the challenges this puts on requirements. Initially, we mainly included questions about agility to investigate its influence. However, all interviewees highlighted certain properties of agile that are important when considering the interplay of HF and AV development. In order to mirror the emphasis that our interviewees put on agile methods, we begin by describing the properties of agile that influence the management of HF knowledge most. The following themes emerged from the data analysis of these properties (shown as P1–P4 in Fig. 2).

(P1) iterative incremental work.

P1 based on codes from interviews with: S2, S4, S5, S6, S7

Agile promotes *iterative incremental work*, to help organizations deliver fast and often as well as increase their responsiveness to changing requirements. For example, Subject 4 mentions that a property of agile work is an incomplete specification early on, combined with iterative work:

> *"[…] But we are working in an agile way, so the specification is not complete in the beginning, but we iterate, and changes might come later. "* — S4 - AV Engineer

Subject 2 suggests that this has completely changed how HF are communicated to development teams:

> *"We had requirements, but that has changed with the agile transformation. We now see it mainly as knowledge transfer, how to move HF knowledge to the teams. The game has completely changed. It is much more a social kind of setting."* — S2 - AV Expert

Our interviewees mainly expressed this as a positive change, as expressed by Subject 5:

> *"At least not in the very old way, where high-level aspects are very much disconnected. Waterfall will not be the solution. But better integration and iterative work sound very promising."* — S5 - HF Expert

Yet, it is important to complement the perspective of teams with a full system view and make sure that HF (for example) fit into the big picture, as Subject 7 mentions:

> *"Agile teams tend to get small bits of tasks and work with these for a short period and then leave it because it is not in the backlog anymore. If it was only for the teams to develop, then nobody would take full system view. What kind of language do we use, when to use knobs, touch screens,…,if it was only up to the teams, you would not have that holistic picture. That is our most important part right now."* — S7 - HF Expert

(P2) shifting responsibility to autonomous teams.

P2 based on codes from interviews with: S3, S6, S7

Agile methods aim to achieve fast, incremental delivery and responsiveness to change by *shifting responsibility to autonomous teams*. These teams can then make local decisions quickly. As a result, agile teams dislike static, detailed requirements, which limit the team's autonomy and, therefore, its effectiveness. This property of agile is mentioned by Subject 3 (for example):

> *"[…] they are then responsible for the topic. T-shaped teams."* — S3 - AV Engineer

This property of agile teams has advantages and disadvantages. Subject 6, for example, highlights the transparency that this approach generates.

> *"I like the way we work now with agile trains. Things are very visible; you see all the stories created by the different teams, and you have clear goals. It is in the method that you promote what each team is doing."* — S6 - HF Expert

However, Subject 7 repeats their concern about the potentially missing system level view as a result of increased team responsibilities.

> *"Ideas come up internally that developers and hardware designers should know their requirements by themselves. I feel like that is difficult."* — S7 - HF Expert

(P3) teams responsible for discovering knowledge.

P3 based on codes from interviews with: S1, S3, S6, S7, S9

Instead of receiving detailed requirements, agile teams prefer being *responsible for discovering knowledge* themselves, relying on face-to-face communication rather than on extensive documentation.

This preference is implied by a number of our interviewees. Subject 7, for example, explains how the role of HF experts has changed:

> *"[…] It is less about handing over requirements, and instead being there for discussion or to evaluate the concepts."* — S7 - HF Expert

The responsibility of agile teams to discover knowledge is also evident from how S9 describes the need for agile teams to seek expertise:

> *"Then, as an engineer, you should have enough awareness to know when to seek out that expertise. But it is, of course, only one competence area of many."* — S9 - AV Engineer

Similarly, Subject 1 shares their view on how to guide agile teams to discover knowledge about the right concepts, not by defining requirements but by relating high-level stories that then can be explored:

> *"[…] Do the guerilla requirements. Do not write requirements, but tell interesting stories based on empirical data, getting the right concepts into the brains of engineers (where it then stays because they are so bad at forgetting things)."* — S1 - HF Expert

(P4) focus on quality in use.

P4 (agile) based on codes from interviews with: S3, S4, S6, S10

One of the differences highlighted by our interviewees between agile and traditional approaches is the different concept of quality. The quality of software-based systems is commonly divided into internal quality (structural properties such as maintainability of the software) and external quality (the fulfillment of user requirements—i.e., providing the desired functionality) (Freeman and Pryce, 2009). In contrast, agile approaches suggest that requirements rapidly change and those provided initially may not describe the users' needs by the time the product is finished. Therefore, according to agile approaches, it is not sufficient to fulfill (potentially outdated) requirements to obtain user satisfaction; it is necessary to address the users' actual needs and focus on *quality in use*. Agile practices with this focus include, for example, the on-site customer (Beck, 1999) and sprint demos (Schwaber and Beedle, 2001).

Thus, agile teams take responsibility for regularly demonstrating a working product, putting it to use in the intended context, and enabling feedback by end users and customers.

A good description of this property was given by Subject 3:

> *"[…]Working agile means being able to test what you are doing and improve the quality continuously."* — S3 - AV engineer

It is, however, important to not rely solely on automated tests. Subject 4 highlights the need to push for acceptance tests.

> *"[…] Put more emphasis on validation tests, that is, not only automated tests but also test the quality in use."* — S4 - AV Engineer

This is generally a good fit for HF, as our interviewees mentioned—for example:

> *"Understanding that problem is crucial, as well as getting experience about what users like. How do people want to be addressed?"* — S10 - HF/AV Engineer

> *"[…] If you have a nice mindset and an open point of view, the iterations, increments, and multi-disciplinary work will fix many of these things. User-centered design."* — S6 - HF Expert

There are, however, a number of conceptual mismatches between the HF and agile AV development domains. Examples include agile focusing on delivering a working product and rejecting big up-front analysis and secondary documents (for example, requirements, architectures, or HF studies)—and even removing those documents after implementation is complete (Meyer, 2014). These practices may lead a team to decide on a particular design based on requirements and HF studies, and to maintain only the actual work product. In future iterations, therefore, the rationale for a design decision is no longer available, potentially leading to duplicate or sub-optimal work (since previous requirements and HF knowledge cannot evolve).

4.2.2. Properties of human factors.

In order to represent the relevant properties of human factors, the following themes were derived from certain characteristics referred to by the interviewees.

(P4) focus on quality in use.

P4 (HF) based on codes from interviews with: S2, S9, S10

HF experts also focus on *quality in use*, since they are concerned with deriving knowledge from human interactions with the system (here: the AV). Clearly, with that focus, the internal, structural properties of the software are of little relevance. Even external quality does not sufficiently describe a system's quality from a human factors perspective: A system that fulfills all requirements on paper but is not pleasurable, satisfactory, or safe to use in the real world will fail to win over an end user. As a result, with agile, HF experts and AV engineers are much closer to each other than they were in traditional development approaches (which broke HF quality considerations down into internal and external quality indicators for implementation). This concordance is implied by the following response from S9:

> *"[…]Not sure we are good with agile yet, but ideally, through improved testing, we should get even more improvements. As long as you can include an HF expert, then all should be fine in the larger picture."* — S9 - AV Engineer

Incremental, agile work can actually be ideal for addressing HF. For example, S2 points out that it allows the quick generation of feedback and an understanding of HF in relation to the system under construction.

> *"[…] Could be really interesting to see how an HF requirement changes with time. How and why does it change? You change it because of some feedback. Why did it not work? Because of this test. Then assess the quality of the test (formal or just friends trying it out). Then also heuristic evaluations, defining usability errors. For those, you do not need a lot of subjects. This is not a statistical approach; it can generate a lot of problems at a low cost. But are these the right problems? The key problem is that HF experiments are expensive."* — S2 - HF Expert

(P5) the importance of experiments.

P5 based on codes from interviews with: S2-4, S7, S9, S11

HF experts highlight the importance of experiments and testing the system. In agile development particularly, iterative work demands continuous testing, both to avoid regression problems and to address changing requirements.

HF experts aim to perform experiments with the system under assessment using human subjects who are not on the engineering team developing the product. Thus, HF experts might test how humans react in specific situations, how they get distracted, how they feel about the system, and how the system affects their behavior (e.g., over-reliance), while considering human variability. S2, for example, relates the importance of experiments to the need to identify assumptions:

> *"You need to identify assumptions. …Start from someone's idea and explore it (from engineering), or you can take your own knowledge (HF) and bring it in. And then you create the experiment and the conditions."* — S2- HF Expert

Again, the shift to agile work has significantly changed the work with experiments. As S3 points out, it requires continuously finding ways to test assumptions.

> *"[…] Before it was easier: Just ask this department to come up with requirements from HF perspective, then push it into the development teams. Then, have test methods in place. What we have done…working agile means to be able to test what you are doing and improve the quality continuously. That also well matches with HF."* — S3 - AV Engineer

Even though it might have been *easier* before, as S3 points out, referring to none-agile ways of working, our interviewees confirm that agility promises to be more effective, as stated by S7:

> *"Agile promotes these things; you need to demo regularly. [but are there enough HF people?]."* — S7 - HF Expert

Other interviewees reason that short, quick experiments with quick feedback cycles should be preferred. The short feedback cycle would help to identify challenges and notify the organization while the topic is still hot. This could enable bringing in the right expertise (e.g., HF or control theory) at the right time, and consequently make the team "fluid and agile".

Perhaps experiments to check assumptions could become a continuous source of input to agile development, since assumptions will always come up. S4, for example, speculates about a shared service to provide support for such continuous experimenting:

> *"You could treat this as a shared service for everyone, support to set up such experiments. It should be quick and easy. It is also related to dealing with assumptions in a more structured way than we currently do."* — S4 - AV Engineer

(P6) the importance of considering human variability.

P6 based on codes from interviews with: S1, S4-5, S8, S12

HF play an important role in ensuring that the developed systems are suitable for all humans (with different user characteristics such as age, culture, experience, and visual and cognitive capabilities). Depending on their backgrounds, humans have different capabilities, limitations, and behavior, as for example stated by S4:

> *"Requirements are very different depending on the country and customer company. How does culture change how people think about HF?"* — S4 - AV Engineer

HF knowledge can help design the system to improve its performance, while considering human variability makes the system usable for a diverse set of users. For example, S5 confirms:

> *"Yes. Humans are complex, with strengths and weaknesses that are very different from artificial systems, there is a lot of variability in the performance of a human."* — S5 - HF Expert

This leads to a high level of complexity that must be managed during AV development.

> *"In many cases, the empirical data set is very complex."* — S1 - HF Expert

Bringing the complexity of human aspects into the development of AV also poses technical challenges to engineering, as S8 suggests:

> *"We need the car to handle random walks with these parameters or with those parameters. Can we even model all human traffic in this way?"* — S8 - AV Engineer

The challenge of modeling complex human traffic behavior could also be seen as an argument for the iterative development of AV systems and HF experiments: it would not only allow the incremental verification of assumptions that are relevant for the current development, but it would also allow the accumulation of knowledge about the bigger picture.

(P7) the importance of making HMIs and automation transparent.

P7 based on codes from interviews with: S4, S7, S10-11

It is critical for users of vehicle automation to have a proper understanding of the system's capabilities and limitations (i.e., the decisions the AV makes must be understandable and the user must understand what the system's limits are) in order to respond correctly and avoid misuse or disuse of the system. Yet not all users read the manual or attend training. Therefore, the system's capabilities and limitations must be completely transparent, through HMIs and kinematic cues; the AV's capabilities and limitations should be obvious as a result of proper HF design. S10, for example, frames important HF questions around this theme:

> *"Who even checks the manual? Will you even be able to find the button that activates an assisting system? With new functionality in a car, how do you introduce it to users?"* — S10 - HF/AV Engineer

If a feature is not transparent to users, they might deactivate it (potentially reducing their safety, but even more problematically, resulting in over-reliance and over-trust).

"Try to find ways so that users do not switch off active safety systems. It is about the methods, how you use them, and their purpose, HF and RE." — S4 - AV Engineer

It is through the effective interplay of systems and users that the overall safety goals are reached. Making sure that typical users sufficiently understand new features is, therefore, an integral HF part of developing AV.

"[…]There are certain design issues there, but there is also [the fact that] that people trust the AV to be safe; how they perceive the risk is important. Do they feel comfortable with the algorithms? Do the algorithms behave as expected? Does it operate reliably?" — S11 - HF Expert

Aligning trust and understanding between users and automated systems is of critical importance—but also hard to do. HF expertise is needed, which could, as S7 points out, be obtained from experts on the team or from the results of surveys (or other sources):

" …8/10 people can make sense of the new function in the first attempt. We need either to be there with our expertise or bring in the end users, e.g., in a clinic or survey, have test drivers." — S7 - HF Expert

4.3. Implications (RQ3)

This section presents the implications that emerged from interview notes on the three themes related to research question RQ3 (shown as I1–I10 in Fig. 2). Each theme (implications for agile ways of working, implications for HF work, and implications for managing requirements) is presented in a separate subsection.

4.3.1. Implications for agile

Given the set of properties of agile and HF discussed above, there are certain implications for any organization that aims to take HF knowledge explicitly into consideration during agile AV development. These implications are not currently provided by agile methods, nor are they easily achieved. This section, therefore, highlights the need to adjust agile ways of working and presents, where available, potential approaches indicated by interviewees.

(I1) AV developers must run human factors experiments.

I1 based on codes from interviews with: S2, S5, S7-8, S12

"[…]Holistic view, ideas come up internally that developers and h/w designers should know requirements by themselves. I feel like that is difficult." — S7 - HF Expert

"[…] it is less about handing over requirements, and instead being there for discussion or to evaluate the concepts." — S7 - HF Expert

Thus, when integrating HF knowledge into agile AV development, it follows that agile teams must be able to run HF experiments themselves. This ability is the first implication for agile that we derived from our interview data. For example, S8 clearly states that the engineers are ultimately responsible for the implementation of a function:

"[…] Engineers should make sure that those (requirements) are implemented and tested." — S8 - AV Engineer

This generally includes extensive testing. However, as S12 points out, tests that only focus on technical aspects and ignore HF will not fully cover the actual needs.

"You know engineers will test and retest and retest, but not really with a human in mind…" — S12 - HF Expert

Agile teams know best what specific knowledge is needed at any given time. Yet, those teams usually lack the HF expertise and knowledge, which must then be provided in a different way (see Implication I3).

"For an engineer without HF training, the fundamental thing in HF is to test your assumptions. How do you communicate to engineers that to get HF knowledge, you need to test it with human subjects? Experiments." — S2 - HF Expert

"[…]I do realize that the teams need such HF knowledge." — S5 - HF Expert

Our interview data indicates that a core challenge is that agile frameworks do not offer dedicated support for teams to run HF experiments. Due to the large number of autonomous agile teams and the wide variety of situations in which HF considerations may have to be made, there are often no dedicated HF resources available to take on the role of designing and running HF experiments for the team.

Based on *(P2) shifting responsibility to autonomous teams*, *(P3) teams are responsible for discovering knowledge* and *(P5) the importance of experiments*, we conclude from our data that *AV developers must run human factors experiments.*

(I2) experiment design & lessons learnt must be created, re-used, and updated efficiently.

I2 based on codes from interviews with: S2, S8-9

If agile teams are to take responsibility for running HF experiments (Implication I1), the teams should also be responsible for decisions about which *experiment design & lessons learnt must be created, re-used, updated efficiently*. S8, for example, suggests the need to aim for re-use.

"[…] We must have a generic model for such experiments, that can be reused in different products." — S8 - AV Engineer

In particular, the re-use and updating of designs and lessons require additional attention in agile ways of working. Agile setups must support a single team as it creates HF experiment designs and generates results, which are then re-used by other teams. If a particular change to the system invalidates the results of a study (e.g., by changing how a user interacts with the system), the team must understand the change and, for example, run a new, updated experiment. In short, teams must be able to judge the validity of experimental designs and results and re-run the experiments if needed, as mentioned by S9:

"Create new knowledge on demand but also use the accumulated knowledge from previous projects. Several levels of tests, even with customers." — S9 - AV Engineer

AV development therefore must integrate discovery and reuse of HF knowledge into agile methods, where the focus is on maintaining tests and deploying working versions of the product iteratively. S2 provides thoughts on how this could work in principle:

"With the agile approach, you continuously test. It allows you to fake a finished product. Then you can put an experienced user in the car and see how they react. You can go in both directions: Start from someone's idea and explore it (from engineering), or you can take your own knowledge (HF) and bring it in. And then, you create the experiment and the conditions and then update it." — S2 - HF Expert

Our second implication therefore follows from our data, specifically considering *(P1) iterative incremental work, (P4) focus on quality in use,* and *(P5) the importance of experiments.*

(I3) human factors expertise must be included on the teams.

I3 based on codes from interviews with: S1-2, S6-9

Agile teams should have the expertise that allows them to take ownership and responsibility for identifying HF needs and relevant HF knowledge. Interviewees suggested *including HF expertise in the agile teams* (for example, in the form of T-shaped teams), with each team member having a certain area of expertise.

> *"Not sure we are good with agile yet, but ideally, through improved testing, we should get even more improvements. As long as you can include an HF expert, then all should be fine in the larger picture."* — S9 - AV Engineer

In the experience of our participants, while there is a lack of availability of HF expertise in most companies, there are, different ways of ensuring teams have the necessary expertise.

S8, for example, wonders whether HF experts should be involved in creating abstract, reusable models, or instead be part of the teams which are deriving operational requirements.

> *"[…]But this requires a good model of the HF. We must have a generic model for such experiments that can be reused in different products, or do we need to create those models within the operational requirements specification? In that case, HF experts must be included in the teams."* — S8 - AV Engineer

Similar considerations were also discussed with S6. In typical scaled-agile frameworks, such as SAFe, HF experts could be assigned as a shared resource or within a particular release train. S6 suggests that as a shared resource, HF experts would lack visibility and would thus not be able to have an impact on agile design decisions.

> *"I like the way we work now with agile trains. Things are very visible; you see all the stories created in the different teams, you have clear goals... The problem is, if you are not on the train, you are not able to promote yourself. If you are a shared resource team, you have less visibility. So it will be better to be on the train."* — S6 - HF Expert

For the same reasons, S1 also considers adding HF experts to the release trains; but in line with S8 above, advances the alternative consideration of having HF experts as part of the individual development teams within an agile release train.

> *"You cannot be everywhere. But having your requirements and hand them over and then wait, that is not going to work. Being a part of the team or an agile train to some extent is the way forward."* — S1 - HF Expert

S7 indicates a clear preference that the HF expert should be involved with the teams directly.

> *"[…]The way you communicate your requirements is within the teams. You need to be there. If you are not in the teams, it will be a challenge."* — S7 - HF Expert

In summary, our interviewees indicated that successful AV development relies on HF experts who can guide developers with respect to how to set up an experiment, run it, and interpret its results—as well as judge its credibility (and identify when a change invalidates previous experiment results, requiring another experiment iteration).

While there are clear advantages to including HF experts directly in agile work (i.e., within the teams or in larger release trains that combine a number of teams working on a specific product area), there are also challenges with this setup: for example, lacking HF experts as S2 indicated.

> *"But we are lacking HF people."* — S2 - HF Expert

I3 is established based on *(P2) shifting responsibility to autonomous teams, (P3) teams responsible for discovering knowledge, (P5) the importance of experiments* and *(P6) the importance of considering human variability*.

(I4) the role of suppliers in agile AV development that integrates human factors must be defined strategically.

I4 based on codes from interviews with: S3-6, S10

Given the lack of HF expertise, one has to identify a strategy that ensures that HF are taken into account in agile AV development. Our participants pointed out that the strategy may consist of getting support in certain specialized areas from outside the team or release train, or even from suppliers with expertise in the area. As the automotive value chain is increasingly transformed into agile ways of working and continuous integration and delivery, new collaborative models are emerging that integrate suppliers tightly into incremental work for specific purposes. In fact, large suppliers already do a substantial amount of research on HF related to their current and future product portfolios as, for example, mentioned by S6.

> *"Currently, we are working more on component level. This is even more challenging since it depends on system level engineering decisions, so you should ideally work with an OEM to define the particular requirements for the component and its context."* — S6 - HF Expert

A particular impediment is the access of suppliers to users of a specific AV, which limits the supplier to relying on their more general expertise and specific requirements from the manufacturer, as discussed by S4.

> *"Yes, but we do not often have access to the users, we get the requirements from the OEM, and we rely on them to tell us what is really needed. So perhaps, it is good that things are then indeed separate (HF, RE)."* — S4 - AV Engineer

Still, we conclude from the overall interview data that the role of suppliers is significant for two reasons: (a) they often possess HF expertise that could be valuable to their customers and (b) as agile development includes increasingly large parts of the value chain, our previous reasoning about the need for HF expertise in agile teams also holds for suppliers.

Our final implication for agile is, therefore, to *systematically decide whether and how to include (or get engaged as) a supplier* in the agile development of AVs, including the supplier's HF expertise in the teams when collaboratively designing, developing, and integrating AV components. It is based on *(P2) shifting responsibility to autonomous teams, (P6) the importance of considering human variability* and *(P7) the importance of making HMIs and automation transparent.*

Summary and important questions. The four implications for agile lead to the following important questions for future research in agile AV development:

1. How can developers be encouraged to run HF experiments?
2. How can we efficiently create, re-use, and update HF experiment designs and lessons learnt?
3. How can HF expertise be included in agile teams, given that few experts are available?
4. How can suppliers be involved strategically in working with human factors?

4.3.2. Implications for HF

(I5) raise awareness among AV developers.

I5 based on codes from interviews with: S5-7, S9

Through our interviews, we learned the need to *raise awareness among engineers* about HF and the implications for the final product of not including HF in the development process.

"It is a lot about marketing yourselves internally. For example, we are part of PI planning for different trains, talk to the teams, explain what we need at which point." — S7 - HF Expert

Although conducting extensive experiments and communicating their results are part of agile development, engineers often do not have enough time to acquire the needed information (e.g., due to short, agile development cycles). Moreover, engineering companies may have engineering cultures; generally, engineers prefer gathering information through data rather than HF, which may be considered less important than simply getting the technology working. This culture is implied in the following quote from S5:

"[...]Sometimes, engineering sometimes just seems to think that HF is about putting nice wallpaper on the wall. They do not understand how early [how fundamentally] HF needs to be taken into account." — S5 - HF Expert

S6 points out that for managers, it is often easier to bring a particular expert onto a team than to work on changing the mindset of the engineering department (although it is much less effective):

"They like to bring in a UX engineer rather than work on the mindset. " — S6- HF Expert

A shift of the overall company mindset would be needed so that HF knowledge can be integrated into the AV development more effectively, as S9 hopes for:

"[...] Then, as an engineer, you should have enough awareness to know when to seek out that expertise." — S9 - AV Engineer

I5 is based on *(P2) shifting responsibility to autonomous teams, (P5) the importance of experiments, (P6) the importance of considering human variability* and *(P7) the importance of making HMIs and automation transparent.*

(I6) provide teams with questions, not requirements.

I6 based on codes from interviews with: S2-4, S7, S11

As AV engineers adapt to work in an agile way, communication about HF and its incorporation in the development process must be adjusted as well. One of our interviewees formulated this implication clearly:

"Put questions on teams, not requirements." — S3 - AV Engineer

Agile teams do not like detailed requirements, which are often too detailed and too static, interfering with their autonomy as they seek appropriate solutions and adjust to change as indicated by for example, S2.

"We had requirements, but that has changed with the agile transformation. We now see it mainly as knowledge transfer, how to move HF knowledge to the teams. The game has completely changed. It is much more a social kind of setting." — S2 - HF Expert

It might be better, therefore, to raise important questions and allow the agile team to find answers that fit their current state of development as stated by S7.

"[...] it is less about handing over requirements, and instead being there for discussion or to evaluate the concepts." — S7 - HF Expert

A complementary approach (to raising questions for the team) relies on storytelling. By using stories that highlight the critical concepts while considering questions that point to the critical information needed, agile teams are enabled to take responsibility for HF knowledge. This empowerment is the consequence of *(P3) teams responsible for discovering knowledge*, and *(I3) human factors expertise must be included on the team.*

(I7) provide basic HF knowledge as checklists and design principles.

I7 based on codes from interviews with: S1, S6-7, S12

A key impediment to providing HF expertise to agile teams is the availability of experts, as mentioned by S7:

" We have tried different things. We had one HMI expert in each team, but that did not scale, we do not have enough experts to have one in each team for 100%. Maybe HF experts should provide checklists to engineers." — S7 - HF Expert

We, therefore, add implication *(I7)*: HF experts should *provide basic HF knowledge as checklists and design principles* to development teams. S1, for example, points out that HF experts should work on a higher abstraction level to increase their reach. They should provide guidelines and other reusable knowledge, rather than specific, system-related requirements:

"From an HF perspective, it is important to prioritize the human experience. Better to talk about guidelines than about requirements." — S1 - HF Expert

The availability of such reusable guidelines would be an asset, as S5 confirms:

"Ideally, one would need some guidelines, to coordinate between application projects that must be communicated. Those guidelines can be in PowerPoint or other company standards." — S5 - HF Expert

According to S12, this could be done via checklists:

"I think we need to make engineers aware of the typical HF limitations and capabilities...You know, how is the mental model affected, or, you know, what is the relationship between the system and our mental model, or fatigue, distraction, situation awareness, workload, all of this everyday stuff that we as people suffer from when it comes to interacting with systems. So, you know, it is almost like a checklist...I guess we need to have a certain checklist." — S12 - HF Expert

Several of our interviewees agreed that this could lead to a better utilization of the available HF experts' skills. This implication is supported by *(I1) AV developers must run human factors experiments, (I3) human factors expertise must be included on the teams, (P5) the importance of experiments, (P6) the importance of considering human variability*, and *(P7)the importance of making HMIs and automation transparent.*

Summary and important questions. The implications for HF indicate a strategic, rather than operational, role for HF experts. Instead of designing and running experiments themselves, these experts are increasingly mentoring and supporting agile teams. This raises important questions:

1. How can awareness of HF be raised in agile AV development?
2. How can agile teams be enabled to effectively create and maintain HF knowledge?
3. Which guidelines and design principles can provide basic HF knowledge to agile teams?

4.3.3. Implications for requirements engineering

(I8) use epics and user stories to express a need for learning requirements in the backlog.

I8 based on codes from interviews with: S1, S3, S6

Agile methods provide only a limited view of requirements, focusing mainly on epics and user stories in various backlogs. This shortcoming introduces new challenges for decomposing high-level concerns into different backlog items and distributing them over the different release trains and value streams, as S3 pointed out.

> *"Base product stream, the AD product stream. SAFe will affect the effort to find the right solution very much. Epic on high level, how to divide it into different backlog items. Need to learn it."* — S3 - AV Engineer

While interviewees mention that there is still a lot to learn, advantages and best practices slowly become manifest, as mentioned by S6:

> *" Things are very visible, you see all the stories created in the different teams, you have clear goals... We should likely start documenting them as part of epics in JIRA. We have HF streams, active safety streams, ... The work is cross-functional, so I am both in HF and active safety streams. The recommendations/ functions should be written in a user-friendly way and which value it provides to customer and user."* — S6 - HF Expert

This, in particular, affects strategies to get cross-cutting and interrelated requirements such as those related to HF into the system, as the following practice from S1 suggests:

> *"[...] Do the guerilla requirements. Do not write requirements, but tell interesting stories based on empirical data, getting the right concepts into the brains of engineers (where it then stays because they are so bad at forgetting things)."* — S1 - HF Expert

In the experience of our interviewees, for RE experts in the automotive domain, the change in focus from providing a comprehensive requirements document to managing continuous learning with respect to certain goals is challenging. From our interviews, we conclude that an RE expert should enable teams to approach and document this learning systematically, instead of writing requirements for them. This implication is based on *(P1) iterative incremental work* and *(P4) focus on quality in use.*

(I9) increase capability to use prototypes for requirements elicitation and validation.

I9 based on codes from interviews with: S2, S4, S10-11

Prototyping was suggested by S4 when discussing requirements engineering:

> *"Prototyping for requirements engineering, so one can find specific details about a problem, and use them to discover new requirements."* — S4 - AV Engineer

This is not only a good way for agile teams to discover requirements, but also a necessary way for HF experts to uncover new HF knowledge, as S2 suggests:

> *"Then I like to ask them to help me build a prototype, a Wizard of Oz car. Then I can test it. Because prototyping is a good way for requirements elicitation and validation."* — S2 - HF Expert

Consequently, prototypes are key for aligning HF experts and agile teams as well as facilitating synergies as indicated by S11.

> *"Prototype adds a set of requirements, but also how the requirements are manifest in terms of interaction or physical design. Then HF experts get involved in evaluating that in usability testing and heuristics evaluation."* — S11 - HF Expert

The infrastructure for constructing prototypes has become quite sophisticated, as mentioned by S10—allowing a huge variety of tests to be run and collecting large amounts of data.

> *"What do we need for hardware to succeed in ADAS or AD platform(first question from the system team)? We have this box full of things we can measure in our prototype.–> Which of these tools do we need... It is fun to work with everything. But we need to find the key sensor outputs for good collaboration. If we have new sensor inputs, how can we put a value on those for a collaboration? How can we structure that kind of work?"* — S10 - AV Engineer

We summarize our interview data in this theme as an implication for RE to *increase the capability to use prototypes for requirements elicitation and validation*, based on the identified needs and HF checklists within agile teams. This implication resonates well with *(P3) teams responsible for discovering knowledge* and *(P5) the importance of experiments*, and might offer good support for *(I2) experiment design & lessons learnt must be created, re-used, and updated efficiently*. This is also in line with *(P6) the importance of considering human variability*, as prototype validation must take into account the range of human variability.

(I10) express the relationship between design decisions and human factors as system requirements during development.

I10 based on codes from interviews with: S3, S5, S9, S10

While it makes sense to describe stakeholder requirements as epics or user stories (see Implication I8), it is important to document the desired capabilities of components and subsystems, which follow system requirements; otherwise, it is not sufficiently clear how HF related to essential requirements for automated vehicles can be managed, as implied by S9:

> *"How does this relate to requirements? It is even tricky to define what a safety requirement is. For safety analysis, the human aspects are critical input to system design and testing. That is with safety as a purpose of design. In particular, the person in the car. In trucks, it is mainly for the safety of other road users. But that is very different from functional safety requirements."* — S9 - AV Engineer

However, it is difficult to clearly define these requirements, as well as architectural decisions, in agile projects, as indicated by S3:

> *"Architectural decisions are taken all over the place. The architect must go around and collect them to raise those aspects that should be treated globally. The decisions now are made differently than they were made before. The design decisions should follow system requirements."* — S3 - AV Engineer

Thus, in the experience of our participants, there is a need to document system requirements, which describe how the different parts of the system under construction will address the stakeholder needs. While these requirements are valuable to manage the knowledge about the system with respect to stakeholder needs and HF, they are not suitable input to agile development work. As S10 implies, one needs to closely investigate collaboration in agile system development to identify system requirements.

> *"To be able to create requirements or needs, one has to understand what is the problem with collaboration today."* — S10 - HF/AV Engineer

From the interviews, we derive the implication of using *system requirements to express the relationship of design decisions to HF knowledge*. In the context of the other implications, we suggest that this implies the need to allow teams to *create system requirements together with the system*, i.e. while developing its software and the corresponding tests, and not before. Requirements would be provided during development (in the form of stories) rather than at the beginning. This sequence allows the requirements to remain up-to-date with the current implementation, rendering them useful for informing future system evolution.

Thus, a general approach that fits our interview data is as follows: teams would run experiments during a sprint and then modify the system accordingly for the next release. They would also, at the same time, describe the updated capabilities of the system and *trace system requirements to related existing/future HF experiments, in order to provide rationales for the decisions.*

This implication is based on *(P1) iterative incremental work.*

Summary and important questions.
As with HF, the implications for RE call for a changed role for requirements engineers. A high percentage of requirements will be discovered and managed just-in-time by agile teams. RE experts, therefore, will increasingly provide infrastructure and coaching, which raises important questions:

1. How can epics and user stories be positioned as a means to learn rather than to specify?
2. How can agile teams be enabled to use prototyping to perform HF experiments and discover and manage requirements?
3. How can system requirements be used to efficiently express the relationship between design decisions and HF in continuous development?

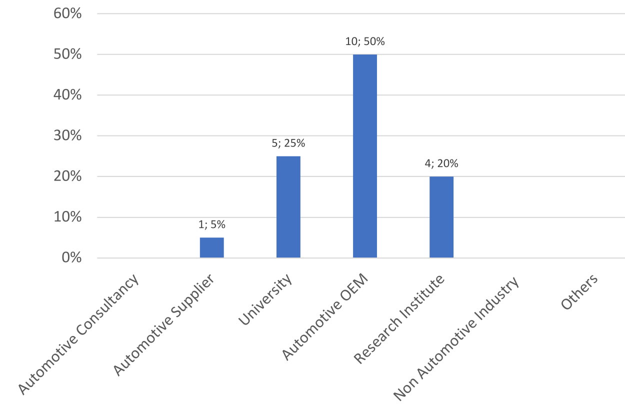
5. Evaluation study

We evaluated the results using a questionnaire-based survey in a workshop setup. By presenting the topic to the audience and directly answering their questions, the workshop format allowed us to ensure that participants understood the topic

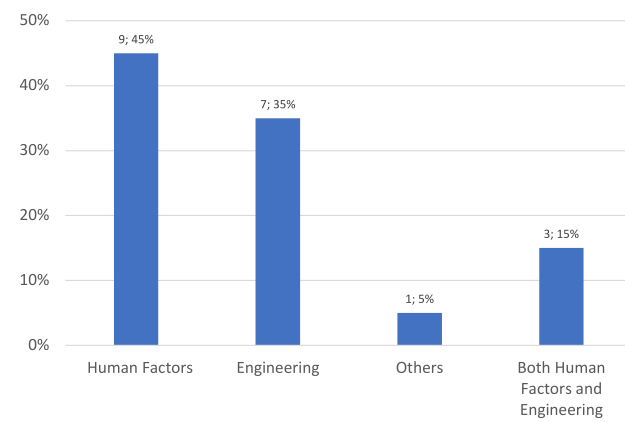
The anonymous questionnaire started with basic demographic questions to assess basic response behavior differences between participants based on their background. We then provided the context, introduced the main topic in the presentation form, and explained the research questions. Next, we explained the research results so that participants could better understand the topic. Keep in mind that the context and description of the outcome of the paper were also provided to participants before the session. Afterward, we asked participants to indicate their level of agreement (on a 5-point Likert scale) with the stated impacts of the properties of agile and HF on AV development. Finally, we asked for the participants' agreement on the implications of the agile way of working, HF, and managing requirements in AV development.

5.1. Participants' demographics

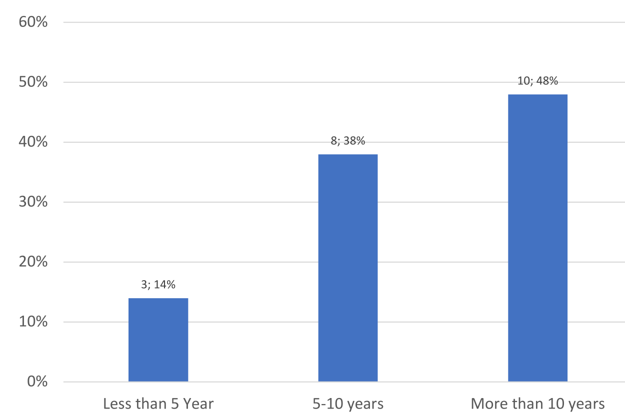
Fig. 3 presents the demographic data of the workshop participants. It displays the absolute number and percentage of respondents for each answer. For this survey, participants were invited from different automotive companies and research institutes, mainly based in Sweden. There were 28 participants in the workshop and we asked three basic demographic questions. We did not include participants from the original interview study, to avoid bias. On average, 17 participants responded to each question, and the rest (on average 11) chose not to answer. Fig. 3(a) depicts the overall results and shows that the majority (50%) of participants work for automotive OEMs, 5% work for automotive



(a) Area of work - (20/28 Respondents)



(b) Work Perspective -(20/28 Respondents)



(c) Work Experience - (22/28 Respondents)

Fig. 3. Demographics. (Semicolon ';' separates absolute numbers and percentages of respondents.)

suppliers, 20% work in research institutes, and the rest were from academia.

Out of the total participants, 20 responded to the second demographic question 3(b). Among these respondents, nine had human factors work perspective, seven had an engineering perspective, and three had experience in both fields. Regarding the third question about work experience, eighteen participants responded, and Fig. 3(c) depicts that 48 percent of participants have more than ten years of work experience.

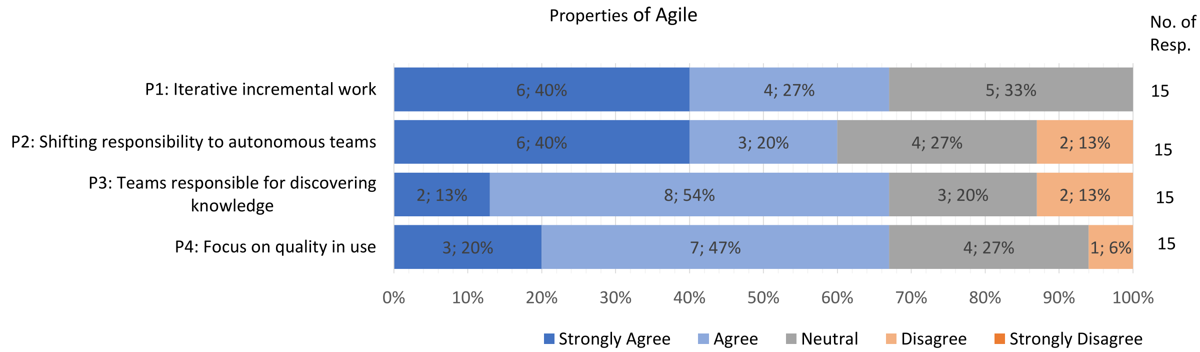


Fig. 4. Level of agreement regarding the impacts of the properties of Agile on AV development. (Semicolon ';' separates numbers and percentages of respondents.)

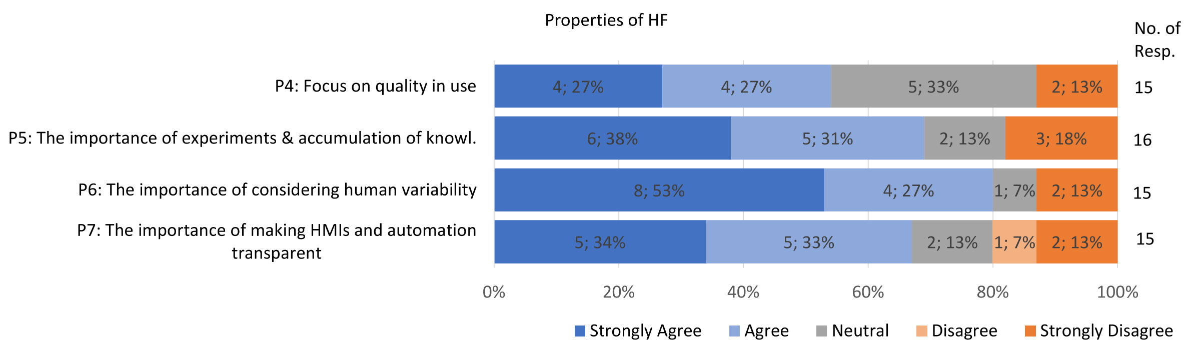


Fig. 5. Level of agreement regarding the impacts of the properties of HF on AV development. (Semicolon ';' separates numbers and percentages of respondents.)

Overall, the fact that all participants were from Sweden limits the generalizability of the results. However, the survey aimed to evaluate our already identified findings (which were obtained using industry experts in Sweden) rather than arriving at a general conclusion or discovering new implications/properties. At the end of the workshop, we asked the participants if we had missed any critical topics.

5.2. Evaluation of properties of agile and HF

On the next questionnaire page, we started with RQ2 and explained the properties of agile and HF which can impact AV development. We then asked the participants to indicate their level of agreement with our interview study findings (on the 5-point Likert scale), in order to assess whether the participants identified the same properties as important.

The survey results are shown in Figs. 4 and 5 for the properties of agile and HF, respectively. The blue bars on the left indicate the percentage of participants who agreed (light blue) or strongly agreed (dark blue) with the findings. The grey bar in the middle shows the percentage of neutral participants, the light orange bar depicts the percentage of participants who showed disagreement, and the dark orange bar on the right shows strong disagreement.

Fig. 4 shows that the majority of participants agreed with *(P1) iterative incremental work*. Five participants were neutral, and nobody disagreed with *(P1)*. For *(P2) shifting responsibility to autonomous teams*, 13% of participants were slightly in disagreement. 40% and 20% of the participants strongly agreed or agreed, respectively, while the rest were neutral. Fifteen participants rated *(P3) teams responsible for discovering knowledge*, majority of participants showed their agreement (54% agreed and 13% strongly agreed) with *(P3)*. For *(P4) focus on quality in use*, 67% of respondents agreed with the statement. One participant strongly disagreed, and the rest were neutral. Fig. 5 presents the properties of human factors. The survey results show that the majority of participants agreed with all the statements, while only a small percentage of respondents disagreed.

The majority of participants either agreed or were neutral with the identified properties for both agile and human factors. Thus we can say that our initial impression that these properties are critical for defining HF requirements in agile AV development is supported by the participants.

5.3. Evaluation of implications

With respect to RQ3, the questionnaire presented Likert scale statements about the implications of combining the relevant properties of HF and scaled agile into the agile way of working, HF, and managing requirements in AV development. The survey results are presented in Figs. 6, 7, and 8, which show the distribution of responses for each implication. In these figures, semicolons ';' are used to separate the numbers and percentages of respondents.

We started with the agile implications, asking the participants to rate their level of agreement for each implication. Fig. 6 shows the findings for each implication of scaled agile on the agile way of working. For both *(I1) AV developers must run human factors experiments* and *(I2) experiment design & lessons learnt must be created, re-used, and updated efficiently*, 50% of participants showed strong agreement, and 72% in total expressed their agreement with the stated implications. The majority (57%) of participants strongly agreed with *(I3) human factors expertise must be included on the teams*, and 64% of participants agreed with *(I4) the role of suppliers in agile AV development that integrates human factors must be defined strategically*.

Generally, more than 50% of respondents agreed with the stated implications of HF and RE (presented in Figs. 7 and 8, respectively). An exception was *(I6) put questions on teams, not requirements*. An equal number of participants agreed and disagreed; however, as 40% of the participants were neutral, there was no clear-cut disagreement. This result suggests that *(I6)*

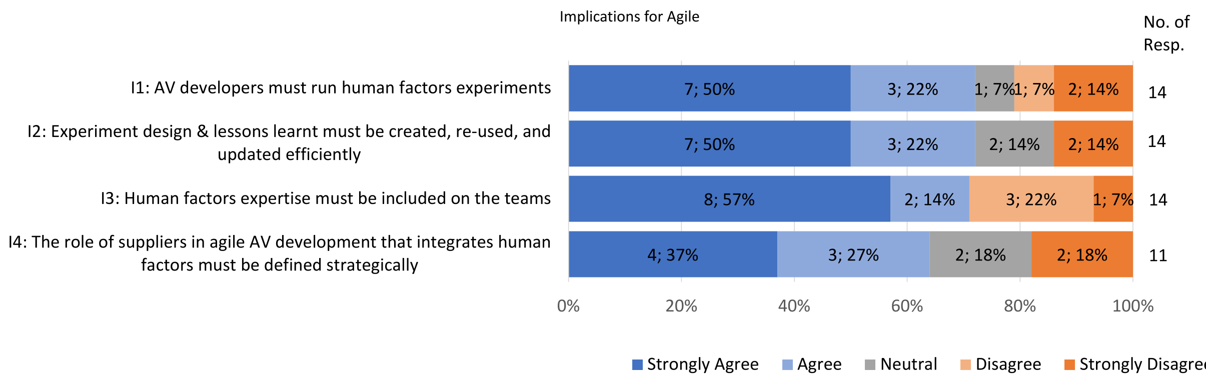


Fig. 6. Level of the agreement for the implications for the agile way of working on AV development.

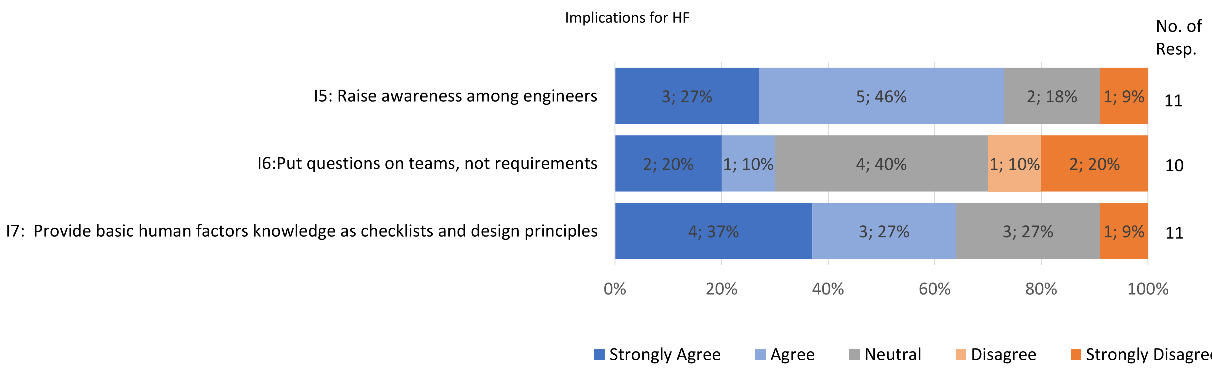


Fig. 7. Level of the agreement for the implications for the HF on AV development.

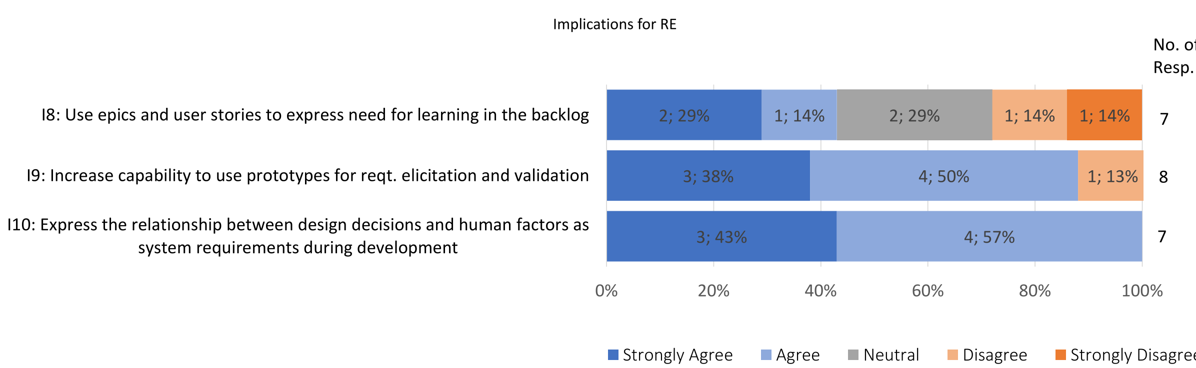


Fig. 8. Level of agreement for the implications for RE on AV development.

should be investigated further. *(I8) use epics and user stories to express a need for learning requirements in the backlog* also showed mixed agreement, indicating the need for extended research on how to represent the need for HF knowledge to AV developers.

For *(I10) Express the relationship between design decisions and human factors as system requirements during development*, all the participants indicated their agreement (57% agreed and 43% strongly agreed).

The results for HF and AV Engineers were similar for most of the questions. However, two HF experts (one with more than ten years of experience and one with less than five) rated the implications for HF very low. On the other hand, all AV engineers rated them highly.

Generally, the majority of participants agreed that all the implications that we derived from the interview notes were relevant and important for bringing HF knowledge into an agile way of working for AV development.

6. Discussion

Based on an exploratory interview study with ten experts from the industry and two experts from academia, this paper charts the landscape of human factors (HF) in relation to the agile development of automated vehicles (AVs). We adopted a Requirements Engineering (RE) perspective, since requirements are traditionally the mechanism for notifying automotive engineers about conditions that should be met by their systems as well as capabilities that the systems should possess (IEEE, 1990; Liebel et al., 2018).

6.1. Implications for practice

We argue that our findings can provide valuable insights for both HF experts and AV engineers in the automotive industry. Particularly, our findings on how to integrate HF and communicate HF requirements during the development should be useful for guiding practitioners. Previous work shows how crucial it is to integrate HF into the RE process. Our results support this finding (e.g. in our themes P5-P7 and I1-I7), but also identify that actually doing so is more difficult with agile development (c.f. themes I1-I10). This is also the case because, the traditional approach to RE has been challenged by the success of agile methods (Meyer, 2014) and their adoption in systems engineering (Liebel et al., 2018). We also acknowledge that areas of AV development that are relatively new, such as AV functionality development based on artificial intelligence (AI, including machine learning) (Nascimento et al., 2019), may require specific focus in the integration of HF. Otherwise, the impact on humans (drivers, occupants, and surrounding traffic) of the (typically highly data-driven Bosch et al., 2018) AI approaches can easily be overlooked.

New roles for HF and RE. Our study took place at a pivotal time in the automotive industry. The automation of driving tasks is proceeding rapidly, adding significant complexity to automotive systems. Automotive companies are transitioning to agile approaches in order to enable shorter development times despite this increased complexity. We were surprised by the strong focus on agile methods in all of our interviews. At the same time, the role of HF knowledge and requirements becomes less clear in the agile setting. Standard RE processes, such as multi-stakeholder analysis, are therefore hard to systematically apply, as discussed by Cheng and Atlee (2009). RE appears to play a smaller role, partially replaced by increment planning and backlog management. Moreover, RE often focuses on specific technical aspects such as functional safety.

This adds to another trend: Automated systems development often prioritizes the technology, without much consideration of HF (Carayon and Hoonakker, 2019). In fact, HF is rarely considered in the early phases (Dul and Neumann, 2009), although our results highlight the importance of doing so. We suggest that this change be enacted through RE, which may help to identify a role for HF in organizations that seek "to inform AV development by providing fundamental knowledge about human capabilities and limitations throughout the design cycle so the product will meet specific quality objectives".

We also suggest further refining the role of RE so that it can better adjust to the needs of agile development, while also improving the support required to integrate HF knowledge into agile development. We envision a role that is less prescriptive and focused on setting requirements for developers, and instead more supportive: enabling developers to explore, document, and re-use requirements-related knowledge. This role will be particularly useful for identifying HF knowledge (e.g., results from experiments) that is no longer valid due to system/software changes—thus, calling for new experiments.

Finally, our findings likely also have implications outside of the actual development of AVs. For example, it may have implication on how computer sciences, agile methods, requirement engineering, and human factors are taught at university level. Exactly how teachers should integrate it in their teaching is out of scope of this study, but it may be relevant to at least talk about the engineer/human factors communication gap, as well as how experts from different domains can contribute to, for example, experiments that include both technology and HF.

Testing and experiments. The field of HF highly prioritizes experimenting and testing. With agile's fast, iterative way of working, there is a need to test regularly and quickly while keeping accumulated knowledge in mind. In contrast with fields such as software testing, in which tests are very formalized and mature, HF has a few substantial challenges:

1. In the context of AV development, formalization of (most) experiments is not mature enough.;
2. Humans are adaptive and unpredictable, making the formalization of experimental protocols and passing thresholds difficult.

Thus, we encourage future research to improve the integration of tests and experiments from an HF perspective into AV development, keeping accumulated knowledge and ensuring that HF experts are part of the experimental setup.

6.2. Implications for research

Common understanding of terms. Definitions provided by our interviewees differed substantially, not only between the HF experts and the AV engineers, but also among the HF experts. This ambiguity identifies a critical communication gap (Bruseberg, 2008). In this work, we propose a slightly refined definition of HF, geared towards the development of AVs (see Definition in Section 4.1) and relating specifically to the essential phases of system engineering. Our results, however, call for future research to achieve an aligned understanding of HF and related concepts through all the systems engineering disciplines involved in AV development.

Raise awareness and develop mindset in agile engineering. It is important to raise awareness and develop an HF-friendly mindset in development teams, in order to improve the communication of HF requirements and their incorporation in the development process. A suitable mindset would consider not just the user experience or HMI, but all aspects of human interactions with a system. Many HF experts agree with this assessment (Wickens et al., 2003; Salvendy, 2012; Flemisch et al., 2008); however, to our knowledge, there is little awareness in systems and software engineering, areas where research is highly encouraged. Awareness could be raised by training engineers in interdisciplinary work so that it becomes easier to integrate HF experts in agile teams (as in I3). In addition, research is needed to determine how to increase the ability of agile teams to manage open questions (see I6) as well as their experimentation infrastructure (see I2) (Fagerholm et al., 2017b; Schermann et al., 2018; Fabijan et al., 2017).

Need to develop and empirically evaluate new approaches to manage HF knowledge. This qualitative study presents several implications which human factors experts, AV agile teams, and requirement engineers can adopt to integrate the knowledge of HF during the agile AV development process.

> *"Or should the team explore the HF? But then we would need a really good model that the team can explore and a lot of expertise that the team can assess. On the high level, we may only have a very crude understanding."* — S8 - AV Engineer

In particular, the need to have AV developers participate in (or even run) HF experiments *(I1)* requires the attention of researchers. In continuous software development, there is a trend towards data-driven decision making and experimentation (Fabijan et al., 2017; Schermann et al., 2018; Meyer, 2015; Kohavi et al., 2009; Kevic et al., 2017).

It could be exciting to compare such experiments on variants of software with HF experiments and investigate possible synergies, which might provide insights into how HF experiments can be integrated into the fast-paced agile development environment.

In summary. We believe that our exploratory study provides a foundation for future research that could improve RE in AV development, as well as refining communication about the HF perspective within the agile way of working. Both HF and RE experts should re-interpret their roles, enabling and facilitating agile teams seeking knowledge—instead of providing comprehensive and detailed knowledge themselves. We anticipate that future research in agile work will formalize ways to manage HF experiments, as well as their results, efficiently. Being able to keep knowledge across design cycles will contribute to a mature synergy between these formerly disparate ways of working.

6.3. Discussion of quality

Our particular epistemological stance (critical realism with influences from pragmatism and constructivism) and choice of method for data collection and analysis also influence the discussion of research quality. In particular, the predominant positivist approach to validity – in terms of construct validity, external validity, internal validity, and generalizability – fits this study poorly. Instead, for this qualitative inquiry, we followed advice from Leung (Leung, 2015), discussing validity, reliability, and generalizability in terms that are a better fit in the context of this study.

6.3.1. Credibility

The credibility of our study is supported by the diverse background of the researchers and our ability to interview the leading experts in the domain. This was the first joint interview study of the authors, which allowed us to bring in complementary perspectives by recruiting interviewees from each author's personal network. Further, we asked each interviewee to suggest additional candidates to mitigate a potential selection bias. By inviting such a diverse group of interviewees and collecting their potentially contradicting perspectives on the topic, we had to challenge and overcome pre-conceptions. We described our background assumptions in detail in Section 2 and challenged them throughout the analysis of our data. This approach has led us to construct our mental model about HF in agile system development. For example: We learned that most participants had only recently moved to agile development approaches. We understand that in such approaches, teams are expected to generate knowledge as needed to implement features. We learned that high-quality HF knowledge stems from experiments and relies on a high level of HF expertise. Thus, we conclude that HF expertise must be included in agile teams to facilitate agile system development that includes HF knowledge, an implication that fits the data from our interviews well and resonated also with participants in the evaluation workshop. The credibility of this study also relies on the quality of answers that we received, both during the interviews and the evaluation activities. Most of our interviewees have to solve the challenges described in this paper as part of their daily job. Significant events can of course influence the answers we receive and we assume that potential challenges encountered at individual case companies with the ongoing agile transformations may fall into this category. We also understand that the context of each expert matters to a degree, therefore, a different sampling might have caused variation in our findings. We mitigated such effects to the best of our ability through an in-depth analysis and construction of what we believe to be the underlying causal relationships, as dictated by critical realism. It is our estimate that such variations would mainly affect ideas about suitable solutions, and to some degree the implications that we derived from our interviews, while the definition and the properties would have been affected to a lesser degree.

6.3.2. Resonance

Through our data collection and analysis, we aimed to establish resonance, e.g., by asking for clarification when we felt that our assumptions were challenged. One such example occurred when we learned that something that was described as relatively easy to accomplish by one participant was described as very difficult by another participant. We learned that in the first case (driver monitoring), a rich set of models, checklists, and design principles exists, which was missing for the second (monitoring of cyclists). The lack of these resources made communication and incorporation of HF considerations considerably more difficult in the second case. Thus the apparent contradiction was explained, providing us with a richer understanding of possible challenges. Implication I7 (provide basic HF knowledge as checklists and design principles) and our definition of HF in the context of the design and development of automated vehicles both reflect the lesson learned.

6.3.3. Usefulness

We believe that our study, albeit a preliminary exploration, is significantly useful. Integrating the design cycle into our working definition of HF in the context of the design and development of automated vehicles is one example of its utility, since the new definition makes it possible to specify where in the design cycle HF knowledge becomes useful. In addition, we believe that our implications provide useful knowledge to those who are tasked with the design of methods and tools for development, as well as to HF experts who aim to increase their impact on AV development. We derive confirmation of these conclusions through feedback received after presenting the study results to the participating companies.

6.3.4. Transferability

Case studies aim to investigate a phenomenon in depth within its natural context. They do not generally aim for generalizable findings in the same way, as for example, an experiment would. Instead, as qualitative research, case studies should lead to theoretical generalizability: concepts that are transferable in principle. Wieringa and Daneva, for example, highlight the ability to provide a causal or structural (architectural) explanation as a theoretical generalization (Wieringa and Daneva, 2015), which then can be transferred to other contexts.

In our study, we provide such explanations through the properties of agile and HF, and the implications for agile, HF, and requirements. Fig. 2 provides an overview of these findings in a qualitative model, specifically relating the concepts (implications) to assumptions (properties) that we have identified through our interviews with experts. In this way, we provide both causal explanations (properties of agile and HF generate implications) and structural explanations (integrating HF into large-scale agile system development will benefit from addressing implications in the area of agile, HF, and RE). This knowledge is transferable, allowing experts from different domains to judge how our concepts apply to them.

Our results stem from the automotive industry, including considerations of automated cars and trucks, and should be applicable to other cases in that domain. We further believe that our concepts are transferable, not only beyond the national AV hotspot where we recruited most of our interviewees, but also to other automated vehicles such as aircraft or ships. It would be harder to transfer beyond the realm of automated vehicles, and even more so if no physical product is created. For example, we would assume that a web application will have very different constraints on prototyping and testing.

7. Conclusion

In this paper, we present an exploratory, qualitative inquiry into how to manage HF knowledge during AV development. Our investigation revealed the fundamental role that large-scale agile development plays in the automotive sector. From our data, we

Table A.3
Overview of themes in relation to codes.

Themes	Example codes
Human factors in relation to AV development	HF relates to safety; HF relates to limitations and capabilities; HF relates to user preferences; Designing in light of understanding; HF is human–machine interaction; Understand, Create and evaluate cycle; HF journal definition
Properties	
(P1) Iterative incremental work	Agile transformation; Agile way of working; Iterative work; Mindset-Agile way of working; Iterative work
(P2) Shifting responsibility to autonomous teams	Teams responsible for topic; Work with agile teams; Teams' autonomy
(P3) Teams responsible for discovering knowledge	Guerrilla requirements; Teams responsible for topic; Knowledge by discussion; Teams' autonomy; Knowledge discovery by Eng.
(P4) Focus on quality in use - Agile	Testing and quality; Quality in use; User centered design; User-centric development
(P4) Focus on quality in use - HF	Quality assessment with HF knowledge; quality improvement with HF knowledge; Assist people
(P5) The importance of experiments	Assumptions and experiments; Continues test and HF experiments; Test assumptions; Agile support; Criticality of human aspects; Agile Experiments
(P6) The importance of considering human variability	Empirical data; Demographic and culture; Human variability; Consideration of different parameters; HF expert evaluate differently
(P7) The importance of making HMIs and automation transparent	Clear HMI; User understandability; Transparency; User trust and comfort
Implications	
(I1) AV developers must run human factors experiments	HF experiments by engineers; Teams & HF knowledge; Engineers & HF requirements; Engineers & HF knowledge; Test with human subject
(I2) Experiment design & lessons learnt must be created, re-used, and updated efficiently	Experiments design; Experiment Model; Use of accumulated knowledge
(I3) Human factors expertise must be included on the teams	HF in teams; Lacking HF people; HF in teams; HF in the teams; Include HF in teams; Improvement with HF in teams
(I4) The role of suppliers in agile AV development that integrates human factors must be defined strategically	HF knowledge by Suppliers; HF requirements by OEM; Work with OEM; OEM's thoughts; OEM' s role
(I5) Raise awareness among AV developers	HF awareness; Mindset; HF marketing; Awareness by engineers
(I6) Provide teams with questions, not requirements	Knowledge transfer; Ask questions, not requirements; HF on crucial questions; No requirements, only discussion; HF on teams
(I7) Provide basic HF knowledge as checklists and design principles	Human experience via guidelines; Provide HF data; checklists by HF; HF Req as checklist
(I8) Use epics and user stories to express a need for learning requirements in the backlog	Tell stories; SAFe & backlog; Stories & Epics
(I9) Increase capability to use prototypes for requirements elicitation and validation	Prototyping for Req elicitation & validation; Prototyping for requirements; Use of prototyping; Prototyping for requirements and evaluation
(I10) Express the relationship between design decisions and human factors as system requirements during development	Design Decisions; Decision and requirements; Purpose of design; Identify problem with collaboration

derived a working definition of Human Factors for AV development, discovered the relevant properties of agile and HF, and defined implications towards agile ways of working, managing HF knowledge, and managing requirements.

Experiments and experience are integral parts of HF. It is a challenge to fit HF knowledge (and the corresponding requirements) into the agile way of working that the automotive industry is moving towards, with its fast pace of change.

As our properties and implications reveal (e.g., P3 and I3), there is an increased need to bring HF expertise to the development teams, caused by the team-based approach and team responsibilities inherent in agile AV development. The paucity of HF experts and the intermittent need for HF expertise in many agile AV development teams makes the inclusion of HF expertise in teams a challenge. In addition, fast, iterative increments do not typically allow time for the rigorous experiments that HF experts may need in order to ensure user-centered quality. In general, reflections from this study and responses from (especially but not exclusively) the HF experts indicate that it is important to push for an HF culture in companies, in the same way that many automotive companies have a safety-first culture. Why not safety and human factors first? Our exploratory study, admittedly limited in scope, relies on 12 interviewees, mainly recruited from a national hotspot of AV development. We believe that our study demonstrates the relevance of this research topic, as well as the value that additional interviews (beyond the scope of this study) could provide.

While further research is still required, our study indicates the potential benefits of integrating HF into the agile way of working. This integration may include protocols where the process can support an environment suitable for iterative HF experiments and user studies based on accumulated knowledge, epics, user stories, and HF checklists. Over time, these protocols will enable developers to create knowledge and data with good reliability. Hence, future work will have to provide a conceptual framework which HF experts and AV engineers can use to support iterative experiments and to accumulate HF knowledge over time. Implementation of this framework would help the automotive industry and individual agile teams alike.

CRediT authorship contribution statement

Amna Pir Muhammad: Conceptualization, Methodology, Writing – original draft, Investigation, Formal analysis, Data curation. **Eric Knauss:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Investigation, Validation, Visualization. **Jonas Bärgman:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Investigation, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have linked interview guide and code book in the manuscript.

Acknowledgments

The authors would like to thank Alessia Knauss for her guidance and support throughout the paper's writing. They would also thank all the interviewees for their time and valuable thoughts, and Tina Mayberry for her language review. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 860410. https://www.shape-it.eu

Appendix. List of themes and related codes

Table A.3 provides all the themes covered in this paper and their corresponding codes. Details (e.g., which code was provided by which interviewer) can be found at the following link 10.5281/zenodo.5562487.

References

Abe, G., Sato, K., Itoh, M., 2017. Driver trust in automated driving systems: The case of overtaking and passing. IEEE Trans. Hum.-Mach. Syst. 48 (1), 85–94.
Ackermann, C., Beggiato, M., Schubert, S., Krems, J.F., 2019a. An experimental study to investigate design and assessment criteria: What is important for communication between pedestrians and automated vehicles? Appl. Ergon. 75, 272–282.
Ackermann, C., Beggiato, M., Schubert, S., Krems, J.F., 2019b. An experimental study to investigate design and assessment criteria: What is important for communication between pedestrians and automated vehicles? Appl. Ergon. 75, 272–282.
Anderson, M., 2020. The road ahead for self-driving cars: The AV industry has had to reset expectations, as it shifts its focus to level 4 autonomy-[News]. IEEE Spectr. 57 (5), 8–9.
Anon, 2019. Preliminary Report Highway HWY19FH008. ntsb.gov, Published online (Accessed 11 October 2021).
Anon, 2021a. Tesla Autopilot Crash Investigation Expands as Ford, BMW and 10 Other Carmakers Asked for Data. Euronews, published online on 2021-09-16, (Accessed 11 October 2021).
Anon, 2021b. A Tesla Mystery: Why Didn't Auto-Braking Stop These Crashes?. Los Angeles News, published online on 2021-10-07, (Accessed 11 October 2021).
Banks, V.A., Plant, K.L., Stanton, N.A., 2018. Driver error or designer error: Using the Perceptual Cycle Model to explore the circumstances surrounding the fatal Tesla crash on 7th May 2016. Saf. Sci. 108, 278–285.
Beauchamp, G., 1986. Technology in the dystopian novel. Mod. Fiction Stud. 32 (1), 53–63.
Beck, K., 1999. Extreme Programming Explained: Embrace Change. Addison-Wesley.
Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., et al., 2001. Manifesto for agile software development. URL https://agilemanifesto.org/.
Berger, C., Eklund, U., 2015. Expectations and challenges from scaling agile in mechatronics-driven companies–a comparative case study. In: International Conference on Agile Software Development. Springer, pp. 15–26.
Birks, M., Chapman, Y., Francis, K., 2008. Memoing in qualitative research: Probing data and processes. J. Res. Nurs. 13 (1), 68–75.
Bodenhamer, A., 2012. Adaptations in the US army MANPRINT process to utilize HSI-inclusive system architectures. Procedia Comput. Sci. 8, 249–254.
Bosch, J., Olsson, H.H., Crnkovic, I., 2018. It takes three to tango: Requirement, outcome/data, and AI driven development.. In: SiBW. pp. 177–192.
Bruseberg, A., 2008. Human views for MODAF as a bridge between human factors integration and systems engineering. J. Cogn. Eng. Decis. Mak. 2 (3), 220–248.
Calp, M.H., Akcayol, M.A., 2019. The importance of human computer interaction in the development process of software projects. arXiv preprint arXiv:1902.02757.
Carayon, P., Hoonakker, P., 2019. Human factors and usability for health information technology: old and new challenges. Yearb. Med. Inform. 28 (01), 071–077.
Carsten, O., Martens, M.H., 2019. How can humans understand their automated cars? HMI principles, problems and solutions. Cogn. Technol. Work 21 (1), 3–20.
CARTRE, 2018. Position Paper on Human Factors. Tech. Rep., Coordination of Automated Road Transport Deployment for Europe, URL https://knowledge-base.connectedautomateddriving.eu/wp-content/uploads/2019/08/CARTRE-Human-Factors-Theme-PositionPaper_180904_FINAL.pdf.
Chen, J.Y., Lakhmani, S.G., Stowers, K., Selkowitz, A.R., Wright, J.L., Barnes, M., 2018. Situation awareness-based agent transparency and human-autonomy teaming effectiveness. Theor. Issues Ergon. Sci. 19 (3), 259–282.
Cheng, B.H., Atlee, J.M., 2009. Current and future research directions in requirements engineering. In: Design Requirements Engineering: A Ten-Year Perspective: Design Requirements Workshop, Cleveland, OH, USA, June 3–6, 2007, Revised and Invited Papers. Springer, pp. 11–43.
Chua, Z.K., Feigh, K.M., 2011. Integrating human factors principles into systems engineering. In: 2011 IEEE/AIAA 30th Digital Avionics Systems Conference. IEEE, pp. 6A1–1.
Creswell, J.W., Creswell, J.D., 2017. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Sage publications.
De Winter, J.C., Happee, R., Martens, M.H., Stanton, N.A., 2014. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. Transp. Res. F 27, 196–217.
Deaths, T., 2020. Tesla Deaths Is a Record of Tesla Accidents That Involved a Driver, Occupant, Cyclist, Motorcyclist, or Pedestrian Death, Whether or Not the Tesla or Its Driver Were at Fault. Tech. Rep., URL https://www.tesladeaths.com. (Accessed 17 February 2021).
Dikert, K., Paasivaara, M., Lassenius, C., 2016. Challenges and success factors for large-scale agile transformations: A systematic literature review. J. Syst. Softw. 119, 87–108.
Dul, J., Neumann, W.P., 2009. Ergonomics contributions to company strategies. Appl. Ergon. 40 (4), 745–752.
Eklund, U., Holmström Olsson, H., Strøm, N.J., 2014. Industrial challenges of scaling agile in mass-produced embedded systems. In: International Conference on Agile Software Development. Springer, pp. 30–42.
Erdal, Y., 2018. Increasing the Safety and Reliability of Autonomous Vehicles. Tech. Rep., Integrated Transport Research Lab, KTH, URL https://www.itrl.kth.se/polopoly_fs/1.863811.1600689663!/YasinErdal_IRPs.pdf. Online (Assessed 2 October 2020).
Faas, S.M., Mathis, L.-A., Baumann, M., 2020. External HMI for self-driving vehicles: which information shall be displayed? Transp. Res. F 68, 171–186.
Fabijan, A., Dmitriev, P., Olsson, H.H., Bosch, J., 2017. The evolution of continuous experimentation in software product development: from data to a data-driven organization at scale. In: 2017 IEEE/ACM 39th International Conference on Software Engineering. ICSE, IEEE, pp. 770–780.
Fagerholm, F., Guinea, A.S., Mäenpää, H., Münch, J., 2017a. The RIGHT model for continuous experimentation. Syst. Softw. 123, 292–305. http://dx.doi.org/10.1016/j.jss.2016.03.034.
Fagerholm, F., Guinea, A.S., Mäenpää, H., Münch, J., 2017b. The RIGHT model for continuous experimentation. J. Syst. Softw. 123, 292–305.
Flemisch, F., Schieben, A., Kelsch, J., Löper, C., 2008. Automation spectrum, inner/outer compatibility and other potentially useful human factors concepts for assistance and automation. Hum. Factors Assist. Autom.
Freeman, S., Pryce, N., 2009. Growing Object-Oriented Software, Guided By Tests. Pearson Education.
Gold, C., Naujoks, F., Radlmayr, J., Bellem, H., Jarosch, O., 2017. Testing scenarios for human factors research in level 3 automated vehicles. In: International Conference on Applied Human Factors and Ergonomics. Springer, pp. 551–559.
Gren, L., Lenberg, P., 2020. Agility is responsiveness to change: An essential definition. In: Proceedings of the Evaluation and Assessment in Software Engineering. pp. 348–353.
Håkansson, E., Bjarnason, E., 2020. Including human factors and ergonomics in requirements engineering for digital work environments. In: 2020 IEEE First International Workshop on Requirements Engineering for Well-Being, Aging, and Health. REWBAH, IEEE, pp. 57–66.
Hancock, P.A., 2014. Automation: how much is too much? Ergonomics 57 (3), 449–454.
Hancock, P.A., 2017. Imposing limits on autonomous systems. Ergonomics 60 (2), 284–291.
Hancock, P., 2019. Some pitfalls in the promises of automated and autonomous vehicles. Ergonomics 1, http://dx.doi.org/10.1080/00140139.2018.1498136.

Heikkilä, V.T., Paasivaara, M., Lasssenius, C., Damian, D., Engblom, C., 2017. Managing the requirements flow from strategy to release in large-scale agile development: a case study at ericsson. Empir. Softw. Eng. 1–45.
Hohl, P., Münch, J., Schneider, K., Stupperich, M., 2016. Forces that prevent agile adoption in the automotive domain. In: International Conference on Product-Focused Software Process Improvement. Springer, pp. 468–476.
Human Factors and Ergonomics Society, 2021. Definitions of Human Factors and Ergonomics. Federated Society of the International Ergonomics Association, URL https://www.hfes.org/About-HFES/What-is-Human-Factors-and-Ergonomics. (Accessed 17 February 2021).
IEEE, 1990. IEEE Standard Glossary of Software Engineering Terminology. IEEE Std 610.12-1990, pp. 1–84. http://dx.doi.org/10.1109/IEEESTD.1990.101064.
Inagaki, T., Itoh, M., 2013. Human's overtrust in and overreliance on advanced driver assistance systems: a theoretical framework. Int. J. Veh. Technol. 2013.
International Requirements Engineering Board, 2020. IREB Certified Professional for Requirements Engineering: Foundation Level. Tech. Rep. Version 3.0.1, IREB e.V., URL https://www.ireb.org/en/downloads/#cpre-foundation-level-syllabus-3-0.
Jacobson, I., Ng, P.-W., McMahon, P.E., Spence, I., Lidman, S., 2012. The essence of software engineering: the SEMAT kernel. Commun. ACM 55 (12), 42–49.
Kahkonen, T., 2004. Agile methods for large organizations-building communities of practice. In: Agile Development Conference. IEEE, pp. 2–10.
Kasauli, R., Knauss, E., Horkoff, J., Liebel, G., de Oliveira Neto, F.G., 2021. Requirements engineering challenges and practices in large-scale agile system development. J. Syst. Softw. 172, 110851.
Kasauli, R., Knauss, E., Nakatumba-Nabende, J., Kanagwa, B., 2020. Agile islands in a waterfall environment: Requirements engineering challenges and strategies in automotive. In: Proc. of International Conference on Evaluation and Assessment in Software Engineering. EASE, Trondheim, Norway, pp. 31–40. http://dx.doi.org/10.1145/3383219.3383223.
Kashfi, P.H., 2018. Integrating User EXperience Principles and Practices Into Software Development Organizations: An Empirical Investigation. Chalmers Tekniska Hogskola (Sweden).
Kevic, K., Murphy, B., Williams, L., Beckmann, J., 2017. Characterizing experimentation in continuous deployment: a case study on bing. In: 2017 IEEE/ACM 39th International Conference on Software Engineering: Software Engineering in Practice Track. ICSE-SEIP, IEEE, pp. 123–132.
Klünder, J., Hohl, P., Fazal-Baqaie, M., Krusche, S., Küpper, S., Linssen, O., Prause, C.R., 2017. HELENA study: Reasons for combining agile and traditional software development approaches in German companies. In: Product-Focused Software Process Improvement: 18th International Conference, PROFES 2017, Innsbruck, Austria, November 29–December 1, 2017, Proceedings 18. Springer, pp. 428–434.
Knaster, R., Leffingwell, D., 2017a. SAFe 4.0 Distilled: Applying the Scaled Agile Framework for Lean Software and Systems Engineering. Addison-Wesley Professional.
Knaster, R., Leffingwell, D., 2017b. SAFe 4.0 Distilled: Applying the Scaled Agile Framework for Lean Software and Systems Engineering. Addison-Wesley Professional.
Knauss, E., Pelliccione, P., Heldal, R., Ågren, M., Hellman, S., Maniette, D., 2016. Continuous integration beyond the team: a tooling perspective on challenges in the automotive industry. In: Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement pp. 1–6.
Kohavi, R., Longbotham, R., Sommerfield, D., Henne, R.M., 2009. Controlled experiments on the web: survey and practical guide. Data Min. Knowl. Discov. 18 (1), 140–181.
Kohli, P., Chadha, A., 2019. Enabling pedestrian safety using computer vision techniques: A case study of the 2018 uber inc. self-driving car crash. In: Future of Information and Communication Conference. Springer, pp. 261–279.
Kovaceva, J., Nero, G., Bärgman, J., Dozza, M., 2019. Drivers overtaking cyclists in the real-world: Evidence from a naturalistic driving study. Saf. Sci. 119, 199–206.
Kraus, J., Scholz, D., Stiegemeier, D., Baumann, M., 2020. The more you know: trust dynamics and calibration in highly automated driving and the effects of take-overs, system malfunction, and system transparency. Hum. Factors 62 (5), 718–736.
Kyriakidis, M., de Winter, J.C., Stanton, N., Bellet, T., van Arem, B., Brookhuis, K., Martens, M.H., Bengler, K., Andersson, J., Merat, N., et al., 2019a. A human factors perspective on automated driving. Theor. Issues Ergon. Sci. 20 (3), 223–249.
Kyriakidis, M., de Winter, J.C., Stanton, N., Bellet, T., van Arem, B., Brookhuis, K., Martens, M.H., Bengler, K., Andersson, J., Merat, N., et al., 2019b. A human factors perspective on automated driving. Theor. Issues Ergon. Sci. 20 (3), 223–249.
Lagerberg, L., Skude, T., Emanuelsson, P., Sandahl, K., Ståhl, D., 2013. The impact of agile principles and practices on large-scale software development projects: A multiple-case study of two projects at ericsson. In: 2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. IEEE, pp. 348–356.
Lawani, A., 2020. Critical realism: what you should know and how to apply it. Qual. Res. 21 (3), 320–333. http://dx.doi.org/10.1108/QRJ-08-2020-0101.
Lee, J.D., 2008. Review of a pivotal human factors article:"humans and automation: use, misuse, disuse, abuse". Hum. Factors 50 (3), 404–410.
Leffingwell, D., 2010. Agile Software Requirements: Lean Requirements Practices for Teams, Programs, and the Enterprise. Addison-Wesley Professional.
Leung, L., 2015. Validity, reliability, and generalizability in qualitative research. Fam. Med. Prim. Care 4, 324–327. http://dx.doi.org/10.4103/2249-4863.161306.
Licht, D., Polzella, D., Boff, K., 1991. Human Factors, Ergonomics, and Human Factors Engineering: Analysis of Definitions. Tech. Rep. 89-01, Wright-Patterson Air Force Base OH: CSERIAC Program Office.
Liebel, G., Tichy, M., Knauss, E., 2019. Use, potential, and showstoppers of models in automotive requirements engineering. Softw. Syst. Model. 18 (4), 2587–2607.
Liebel, G., Tichy, M., Knauss, E., Ljungkrantz, O., Stieglbauer, G., 2018. Organisation and communication problems in automotive requirements engineering. Requir. Eng. 23 (1), 145–167.
Lindvall, M., Muthig, D., Dagnino, A., Wallin, C., Stupperich, M., Kiefer, D., May, J., Kahkonen, T., 2004. Agile software development in large organizations. Computer 37 (12), 26–34. http://dx.doi.org/10.1109/MC.2004.231.
Litman, T., 2021. Autonomous Vehicle Implementation Predictions Implications for Transport Planning. Tech. Rep., Victoria Transport Policy Institute, URL https://www.vtpi.org/avip.pdf.
Mehrfard, H., Pirzadeh, H., Hamou-Lhadj, A., 2010. Investigating the capability of agile processes to support life-science regulations: the case of XP and FDA regulations with a focus on human factor requirements. In: Software Engineering Research, Management and Applications 2010. Springer, pp. 241–255.
Merat, N., Lee, J.D., 2012. Preface to the special section on human factors and automation in vehicles: Designing highly automated vehicles with the driver in mind. Hum. Factors 54 (5), 681–686.
Meyer, B., 2014. Agile! Bertrand Meyerthe Good, the Hype and the Ugl. Springer.
Meyer, M.N., 2015. Two cheers for corporate experimentation: The A/B illusion and the virtues of data-driven innovation. Colo. Tech. LJ 13, 273.
Miller, E.E., Boyle, L.N., 2019. Behavioral adaptations to lane keeping systems: Effects of exposure and withdrawal. Hum. Factors 61 (1), 152–164.
Mirnig, A.G., Wintersberger, P., Sutter, C., Ziegler, J., 2016. A framework for analyzing and calibrating trust in automated vehicles. In: Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. pp. 33–38.
Morando, A., Gershon, P., Mehler, B., Reimer, B., 2021. A model for naturalistic glance behavior around tesla autopilot disengagements. Accid. Anal. Prev. 161, 106348.
Morando, A., Victor, T., Dozza, M., 2016. Drivers anticipate lead-vehicle conflicts during automated longitudinal control: Sensory cues capture driver attention and promote appropriate and timely responses. Accid. Anal. Prev. 97, 206–219.
Nascimento, A.M., Vismari, L.F., Molina, C.B.S.T., Cugnasca, P.S., Camargo, J.B., de Almeida, J.R., Inam, R., Fersman, E., Marquezini, M.V., Hata, A.Y., 2019. A systematic literature review about the impact of artificial intelligence on autonomous vehicle safety. IEEE Trans. Intell. Transp. Syst. 21 (12), 4928–4946.
Navarro, J., 2019. A state of science on highly automated driving. Theor. Issues Ergon. Sci. 20 (3), 366–396.
Noble, H., Smith, J., 2014. Qualitative data analysis: a practical example. Evid.-Based Nurs. 17 (1), 2–3.
Noy, I.Y., Shinar, D., Horrey, W.J., 2018. Automated driving: Safety blind spots. Saf. Sci. 102, 68–78.
Nuseibeh, B., Easterbrook, S., 2000. Requirements engineering: a roadmap. In: Proceedings of the Conference on the Future of Software Engineering pp. 35–46.
Orellana, D.W., Madni, A.M., 2014. Human system integration ontology: enhancing model based systems engineering to evaluate human-system performance. Procedia Comput. Sci. 28, 19–25.
Paasivaara, M., Behm, B., Lassenius, C., Hallikainen, M., 2018. Large-scale agile transformation at ericsson: a case study. Empir. Softw. Eng. 23 (5), 2550–2596.
Palinkas, L.A., Horwitz, S.M., Green, C.A., Wisdom, J.P., Duan, N., Hoagwood, K., 2015. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm. Policy Ment. Health Ment. Health Serv. Res. 42 (5), 533–544. http://dx.doi.org/10.1007/s10488-013-0528-y.
Pernstål, J., Magazinius, A., Gorschek, T., 2012. A study investigating challenges in the interface between product development and manufacturing in the development of software-intensive automotive systems. Int. J. Softw. Eng. Knowl. Eng. 22 (07), 965–1004.
Ramos, A.L., Ferreira, J.V., Barceló, J., 2012. LITHE: an agile methodology for human-centric model-based systems engineering. IEEE Trans. Syst. Man Cybern. 43 (3), 504–521.

Reagan, I.J., Kidd, D.G., Cicchino, J.B., 2017. Driver acceptance of adaptive cruise control and active lane keeping in five production vehicles. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 61. SAGE Publications Sage CA: Los Angeles, CA, pp. 1949–1953.
Revell, K., Langdon, P., Bradley, M., Politis, I., Brown, J., Stanton, N., 2018. User centered ecological interface design (uceid): a novel method applied to the problem of safe and user-friendly interaction between drivers and autonomous vehicles. In: International Conference on Intelligent Human Systems Integration. Springer, pp. 495–501.
Rutakumwa, R., Mugisha, J.O., Bernays, S., Kabunga, E., Tumwekwase, G., Mbonye, M., Seeley, J., 2020. Conducting in-depth interviews with and without voice recorders: a comparative analysis. Qual. Res. 20 (5), 565–581. http://dx.doi.org/10.1177/1468794119884806.
sae.org, 2021. SAE Levels of Driving Automation™ Refined for Clarity and International Audience. Tech. Rep., SAE International, URL https://www.sae.org/blog/sae-j3016-update. (Accessed 23 August 2021).
Saffarian, M., De Winter, J.C., Happee, R., 2012. Automated driving: human-factors issues and design solutions. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 56. Sage Publications Sage CA, Los Angeles, CA, pp. 2296–2300.
Salo, O., Abrahamsson, P., 2008. Agile methods in European embedded software development organisations: a survey on the actual use and usefulness of extreme programming and scrum. IET Softw. 2 (1), 58–64.
Salter, S., Diels, C., Herriotts, P., Kanarachos, S., Thake, D., 2019. Motion sickness in automated vehicles with forward and rearward facing seating orientations. Appl. Ergon. 78, 54–61.
Salvendy, G., 2012. Handbook of Human Factors and Ergonomics. John Wiley & Sons.
Schermann, G., Cito, J., Leitner, P., 2018. Continuous experimentation: challenges, implementation techniques, and current research. IEEE Softw. 35 (2), 26–31.
Schwaber, K., Beedle, M., 2001. Agile software development with Scrum Prentice Hall PTR Upper Saddle River. NJ, USA.
Schwall, M., Daniel, T., Victor, T., Favaro, F., Hohnhold, H., 2020. Waymo public road safety performance data. arXiv preprint arXiv:2011.00038.
SHAPE-IT, 2023. Marie Skłodowska-Curie Action (MSCA) Innovative Training Network (ITN) project. [link] URL https://www.shape-it.eu. (Accessed 06 March 2023).
Star, S.L., Griesemer, J.R., 1989. Institutional ecology,translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. Soc. Stud. Sci. 19 (3), 387–420.
The Journal of the Human Factors and Ergonomics Society, 2020. [link] URL https://journals.sagepub.com/description/hfs. Online (Assessed 16 November 2020).
Van Der Valk, R., Pelliccione, P., Lago, P., Heldal, R., Knauss, E., Juul, J., 2018. Transparency and contracts: continuous integration and delivery in the automotive ecosystem. In: 2018 IEEE/ACM 40th International Conference on Software Engineering: Software Engineering in Practice Track (ICSE-SEIP). IEEE, pp. 23–32.
van Maanen, P.-P., Lindenberg, J., Neerincx, M.A., 2005. Integrating human factors and artificial intelligence in the development of human-machine cooperation. In: Proc. of the 2005 International Conference on Artificial Intelligence. ICAI'05, Citeseer.
Vincent, C.J., Li, Y., Blandford, A., 2014. Integration of human factors and ergonomics during medical device design and development: It's all about communication. Appl. Ergon. 45 (3), 413–419.
Watson, M.E., Rusnock, C.F., Colombi, J.M., Miller, M.E., 2017. Human-centered design using system modeling language. J. Cogn. Eng. Decis. Mak. 11 (3), 252–269.
Wickens, C., Lee, J., Liu, Y., Gordon-Becker, S., 2003. Designing for People: An Introduction to Human Factors Engineering. Pearson.
Wiegers, K., Beatty, J., 2013. Software Requirements. Pearson Education.
Wieringa, R., Daneva, M., 2015. Six strategies for generalizing software engineering theories. Sci. Comput. Program. 101, 136–152. http://dx.doi.org/10.1016/j.scico.2014.11.013.
Wohlrab, R., Knauss, E., Pelliccione, P., 2020. Why and how to balance alignment and diversity of requirements engineering practices in automotive. J. Syst. Softw. 162, 110516.
Wu, T.-F., Tsai, P.-S., Hu, N.-T., Chen, J.-Y., Jheng, J.-J., 2019. Implementation of auto parking system based on ultrasonic sensors. J. Internet Technol. 20 (2), 471–482.
Xu, Y., Chen, B., Shan, X., Jia, W., Lu, Z., Xu, G., 2017. Model predictive control for lane keeping system in autonomous vehicle. In: 2017 7th International Conference on Power Electronics Systems and Applications-Smart Mobility, Power Transfer & Security. PESA, IEEE, pp. 1–5.
Yu, H., Tseng, H.E., Langari, R., 2018. A human-like game theory-based controller for automatic lane changing. Transp. Res. C 88, 140–158.