# A Thematic Analysis on Trust-Building Mechanisms in Autonomous Vehicles

Master Thesis

by

Lars Benedict Martin

Degree Course: Information Systems M.Sc.
Matriculation Number: 2267670

Institute of Applied Informatics and Formal Description Methods (AIFB)

KIT Department of Economics and Management

| | |
|---|---|
| Advisor: | Prof. Dr. Ali Sunyaev |
| Second Advisor: | Prof. Dr. Andreas Oberweis |
| Supervisor: | M.Sc. Maximilian Renner |
| Submitted: | July 16, 2022 |

# Abstract

Autonomous vehicles (AVs) have a wide range of potential advantages, and they open up new transportation options. However, these advantages can only be realized if the general public accepts AVs. One of the requirements preceding acceptance is trust in such an autonomous technology, which is critical to its success. But currently, a significant part of the population does not seem to trust AVs.

To address this issue, research is trying to approach the problem from different perspectives. First, from a psychological perspective, which specifically examines human factors, and second, from a technical perspective, which looks at technical solutions. This work attempts to combine and report from both perspectives. A literature search was conducted for this purpose, which led to 55 relevant scientific papers. Based on the literature base, a Thematic Analysis (TA) was conducted to find trust influencing themes and sub-themes. Altogether, 24 trust-influential sub-themes (characteristics) were identified, which were classified into 6 distinct themes: (1) *Information Exchange*, (2) *User Perception and Comprehension*, (3) *Perceived Intelligence* - from a technological perspective, and (4) *Trust in the Manufacturer*, (5) *Trust in the Technology*, and (6) *Trust in the Legislative* - from a psychological perspective.

Based on our results, we argue that trust in AVs should not be seen as a binary problem, and cannot be significantly increased by covering single, selected characteristics alone. Our results show that trust in AVs can even decrease when individual characteristics are unduly addressed. As a result, if trust in AVs is to be increased, specific characteristics' interdependencies must be evaluated and treated accordingly.

# Contents

**List of Figures** **iv**

**List of Tables** **v**

**1 Introduction** **1**
- 1.1 Motivation . . . 2
- 1.2 Objectives . . . 3
- 1.3 Outline . . . 4

**2 Background** **5**
- 2.1 Artificial Intelligence . . . 5
- 2.2 Autonomous Vehicles . . . 5
- 2.3 Trust . . . 7
- 2.4 Trust in Automation and Autonomous Vehicles . . . 7

**3 Methods** **10**
- 3.1 Data Collection . . . 10
- 3.2 Data Analysis . . . 11
- 3.3 Recognition of Characteristic Interdependencies . . . 12

**4 Results** **13**
- 4.1 Psychological Perspective . . . 17
  - 4.1.1 Trust in the Manufacturer . . . 17
  - 4.1.2 Trust in the Technology . . . 18
  - 4.1.3 Trust in the Legislative . . . 22
- 4.2 Technological Perspective . . . 23
  - 4.2.1 Information Exchange . . . 24
  - 4.2.2 User Perception and Comprehension . . . 26
  - 4.2.3 Perceived Intelligence . . . 27

**5 Discussion** **30**
- 5.1 Principle Findings . . . 30
- 5.2 Intra-Perspective Interdependencies . . . 31
- 5.3 Inter-Perspective Interdependencies . . . 33
- 5.4 Research and Practical Implications . . . 34
- 5.5 Limitations and Future Research . . . 35

**6 Conclusion** **37**

**A Appendix** **38**

A.1 List of Reviewed Literature . . . . . 38
A.2 Overview of Themes . . . . . 45

**References** **52**

# List of Figures

Figure 1 – *Process of the literature selection. The DBs and the retrieved articles can be seen in the first box.* . . . . . 13
Figure 2 – *Distribution of the number of articles over the years. The left distribution shows the number of articles from the technological perspective, while on the right, the distribution of the articles from the psychological perspective is shown.* . . . . . 14
Figure 3 – *Distribution of the number of articles combined in dark green, from a technological perspective in blue, and from a psychological perspective, in light green.* . . . . . 15
Figure 4 – *Number of types of articles. Different types of simulation approaches (virtual reality, projectors, or screens) were grouped together as driving simulators. Some articles used mixed approaches.* . . . . . 15
Figure 5 – *Holistic picture of trust in AVs in form a of a topic map. The red box shows the main goal, green indicates the perspectives, blue represents the generated main themes, and white the characteristics.* . . . . . 16

# List of Tables


1 *SAE-Levels of Automation* . . . . . 6
2 *Luminary Articles* . . . . . 11
3 *Example of the Coding Process* . . . . . 12
4 *List of Reviewed Literature* . . . . . 44
5 *Overview of Characteristics and Keywords of the Theme Information Exchange* . . . . . 45
6 *Overview of Characteristics and Keywords of the Theme User Perception and Comprehension* . . . . . 46
7 *Overview of Characteristics and Keywords of the Theme Perceived Intelligence* . . . . . 47
8 *Overview of Characteristics and Keywords of the Theme Trust in the Technology* . . . . . 49
9 *Overview of Characteristics and Keywords of the Theme Trust in the Legislative* . . . . . 50
10 *Overview of Characteristics and Keywords of the Theme Trust in the Manufacturer* . . . . . 51

# 1 Introduction

Self-driving cars, commonly referred to as AVs, are one of the most significant innovations of the last decade (Adnan et al., 2018, p. 820). Tens of companies invested around $120 billion in the years between 2017 and 2019, according to McKinsey (Holland-Letz et al., 2019), with a clear vision: to revolutionize intelligent transportation. The expected benefits of autonomous vehicles are far-reaching, including improving road safety and reducing road fatalities. For example, roughly 94% of all accidents in the USA in 2018 were humanly caused (Singh, 2018, p. 1). Other benefits include the generation of less traffic and less environmental pollution (Taiebat et al., 2018, p. 11450), and helping elderly and mobility-impaired people to maintain an active lifestyle (Choi & Ji, 2015, p. 692). Many of the disadvantages of driving a car may be turned into benefits, such as using travel time to work on projects rather than being tied to tracks, as is the situation with trains. An ideal AV can pick up passengers and transport them without diversions to their destination.

These advantages are made possible by the rapid development of AV technology, which is based on advances in artificial intelligence (AI) research and the merging of AI with common vehicle technology into intelligent automation (Hengstler et al., 2016, p.105). Although current cars already use AI to support human users in accomplishing their tasks and to support them while steering and braking, the human driver remains in full control of the car (Koester & Salge, 2020, p. 1). One of the AV industry's long-term goals is to develop vehicles that adapt to drivers' needs and provide an authentic driving experience (Rödel et al., 2014, p. 8) in fully automated driving scenarios. To realize these long-term goals, a paradigm shift from augmentation (e.g., lane departure warning systems) to automation must take place (Koester & Salge, 2020, p. 1) while considering the desires of society.

A survey in the USA, UK, and Australia revealed that 92 % of the participants are highly concerned about the safety of AVs, and 22% of the participants could not imagine riding an AV at all (Schoettle & Sivak, 2014, p. 24). Another study based on the same data revealed that 84 % of drivers do prefer no self-driving or partially self-driving (Schoettle & Sivak, 2015, p. 12). Several studies do support these findings (e.g., Kyriakidis et al. 2015, p. 132; Rödel et al. 2014, p. 3; Bansal et al. 2016, p. 6; Zmud & Sener 2017, p. 2508). As fatal accidents in which AVs were involved show, skepticism towards AVs is not unjustified (e.g., Wakabayashi 2018; Pietsch 2021). These accidents lead to uncertainty and vulnerability when interacting with an AV. Hence, it is not surprising that potential customers will look closely at AVs before entrusting them with their lives (Köster & Salge, 2021, p. 2).

## 1.1 Motivation

It is critical to consider the population's skepticism, which is to be derived from the named studies, because almost no one understands nor has experienced the driving functions of AVs yet, making judgment nearly impossible (Häuslschmid et al., 2017, p. 1). Named reasons for skepticism include surrendering control (Sun et al., 2020, p. 1170), a lack of understanding of technology in general (Q. Zhang et al., 2020, p. 260), as well as a lack of trust in AVs (Wintersberger & Riener, 2016, p. 298). However, exploiting the full benefits of AVs can only be achieved when the technology has been adopted on a large scale (T. Zhang et al., 2019, p. 208). As a result, human users must accept safety-critical systems and entrust their lives to computers (Wintersberger et al., 2019, p. 2) to make the vision of AVs a reality. Although it will be some time before highly AVs can be used commercially, it is important to find out to what extent the public's attitude towards the vehicles can be changed in order to create acceptance (Choi & Ji, 2015, p. 692).

Whether technology will be successful or not depends on several factors. A critical one is user acceptance (Davis, 1993, p. 475). According to the Technology Acceptance Model (TAM) postulated by (Davis, 1993, p. 476), technology acceptance is dependent on the perceived usefulness of the system and the perceived ease of use, which are the cognitive responses to an external stimulus, for example, system design features. In other words, the TAM is used to predict the intention of using a technological system (Buckley et al., 2018a, p. 203). That means, transferred to AVs, people will perhaps accept AVs when they see their benefits in terms of driving performance and transportation (Perceived Usefulness) (Buckley et al., 2018a, p. 203), and when they believe the AV reduces their effort of driving (Perceived Ease of Use)(De Angelis et al., 2017, p. 245). Over the years, the TAM has been extended by additional factors (e.g., TAM2 Venkatesh & Davis 2000, p. 188; TAM3 Venkatesh et al. 2003, p. 447). In addition to that, the importance of trust has been determined as critical for the acceptance of technology (J. D. Lee & See, 2004, p. 58). Trust is a necessary precondition of acceptance and is a key determinant for the adoption of new technologies (Körber et al., 2018, p. 306). The first bridge between trust and acceptance was built by the introduction of the Trust TAM (Belanche et al., 2012, p. 53), which emphasizes the importance of trust in the acceptance of technology in general. A deeper connection between trust and acceptance of automation is abstracted in the automation acceptance model (AAM) proposed by Ghazizadeh et al. (2012, p. 43), which is centered on trust and compatibility constructs (Ghazizadeh et al., 2012, p. 46). Consequently, trust is not only an important determinant in interpersonal relationships between humans but also in relationships between humans and machines (Choi & Ji, 2015, p. 692), as the acceptance models show. Hence, before acceptance of AVs may arise, trust in AVs must be fulfilled (Körber et al., 2018, p. 306).

However, fostering trust in AVs is more difficult than fostering trust in other technologies.

People are physically involved with their own lives during the interaction with AVs, which leads to a high perception of risk and, hence, potentially to a low level of trust. While much research on trust in automation has been conducted (e.g., J. D. Lee & See 2004, p. 52; Hoff & Bashir 2015, p. 407), the field of AV research is relatively young, and not all facets of trust are known yet. In the research of trust in AVs, two perspectives can be distinguished: 1.) a psychological perspective (e.g., Hengstler et al. 2016, p. 105; Koester & Salge 2020, p. 1) and 2.) a technological perspective (e.g., Ekman et al. 2018, p. 95; Morra et al. 2019, p. 9438). The psychological perspective emphasizes impressions, feelings, and fears of society and their attitude toward trust in AVs. Questions such as *"what must the AV have for you to trust?"*, *"why do you feel this way?"* arise when researching specific trust-building characteristics. From a technical perspective, the technology itself is critical, especially in the domain of human-computer interaction. At the center of research are design features and algorithmic approaches, where human factors play a subordinate role (Sun & Zhang, 2021, p. 28214) and are frequently overlooked. Technical solutions have already been applied on a temporary basis, and the study participants' reactions to these solutions are of interest. Accordingly, to draw a holistic picture considering both perspectives, it is important to understand not only technological arguments but also psychological influences to increase trust in AVs.

## 1.2 Objectives

In order to understand which characteristics influence trust and how these characteristics can be used to increase the acceptance of AVs in the population, the following research question (RQ) guides this thesis: *What are the key characteristics that influence trust in autonomous vehicles?* To answer the RQ, the following sub-goals are pursued in this thesis:

- Generation of trust influencing themes from a psychological and technological perspective.
- Generation of trust influencing characteristics (sub-themes) within the themes from a psychological and technological perspective.
- Derivation of a holistic picture of trust influencing characteristics (sub-themes) of both perspectives.
- Examination of interdependencies between the themes.
- Examination of interdependencies between the trust influencing characteristics (sub-themes).

## 1.3 Outline

This thesis is structured as follows: In section 2, the foundations for understanding artificial intelligence, AVs, and trust are laid. Section 3 introduces the methodology of the literature review and the process of a TA, including the steps and coding examples. The results of the literature review and the TA are described in section 4. The results are discussed in section 5, including research and practical implications as well as the future research and limitations of this thesis. The thesis is concluded in section 6.

# 2 Background

## 2.1 Artificial Intelligence

The term "artificial intelligence" (AI) is widely known; however, a commonly accepted definition has not been proposed yet. One prominent definition of AI is *"[...] the science and engineering of making intelligent machines, especially intelligent computer programs"* (McCarthy, 2004, p. 2), defined by one of the founding fathers of the research field of AI. However, considering the possible application areas and different perspectives, it is likely that a commonly accepted definition will not ever exist. Independent of an exact definition, two notations of AI are critical. (1) Strong AI, also known as artificial general intelligence, refers to computer systems with a broad scope and strong generalization capabilities (e.g., human-level or higher) in all facets and domains (Goertzel, 2014, p. 3), and is currently a work of fiction (Hengstler et al., 2016, p. 105). (2) Weak AI or narrow artificial intelligence implies systems with domain-specific intelligence, for example, object recognition or pattern recognition, which exceeds human capabilities (Hengstler et al. 2016, p. 105; Pandl et al. 2020, p. 57076). Currently, only weak AI systems are of commercial interest. With the increasing use of AI and the merging of many technologies, trust also plays a major role in the field of AI. For this reason, characteristics that contribute to trustworthy AI were presented. These are known as FATE: **F**airness, **A**ccountability, **T**ransparency and **E**xplainability (Shin, 2020, p. 2). Fairness refers to the fairness of algorithmic decisions made by the AI (e.g., unbiased and non-discriminatory). Accountability refers to who is accountable for the decisions and actions an AI makes. Transparency describes the requirements of AI to make transparent actions. In other words, a user must be able to understand how a decision was made when the AI is transparent. Explainability refers to the ability of the AI to explain why decisions are made.

## 2.2 Autonomous Vehicles

An AV is a vehicle equipped with autonomous technology (California Legislative Counsel, 2012). A vehicle equipped with autonomous technology can drive without physical control or monitoring by a human operator (California Legislative Counsel, 2012). The latest advancements in autonomous technology and AVs are based on AI technology, machine learning , as well as sensing and computing technology (Paden et al., 2016, p. 33). Automation is defined as technology that actively selects data, transforms information, makes decisions or controls processes (J. D. Lee & See, 2004, p. 50) whereas the term "autonomous" inherently includes a decision-making process (Hengstler et al., 2016, p. 105). To distinguish the capability of AVs, various classification systems have been developed, including the National Highway Traffic Safety Administration (NHTSA) standard and

the Society of Automotive Engineers (SAE) (SAE International, 2021) standard (Kaur & Rampersad, 2018, p. 87). The capability of AVs is commonly divided into six levels, starting at level 0 (no driving automation) and ending at level 5 (full driving automation) SAE International (2021). SAE levels 0, 1, and 2 incorporate advanced driving assistance systems (e.g., adaptive cruise control). However, these still rely on human drivers and constant monitoring. With increasing SAE levels, 3, 4, and 5, the vehicle requires less human monitoring up until level 5, where the car can drive autonomously (see Table 1).

***Table 1:*** *SAE-Levels of Automation*

| | |
|---|---|
| Level 0 - No Driving Automation | The vehicle is manually controlled, with no automation |
| level 1 - Driver Assistance | The vehicle has driving assisting functions (e.g., cruise control) |
| level 2 - Partial Driving Automation | The vehicle has partial control while driving (e.g., braking and accelerating). The driver must be ready to take control at any given time. |
| level 3 - Conditional Driving Automation | The vehicle has conditional control (e.g.n highways). The driver must be ready to override options |
| level 4 - High Driving Automation | The vehicle has conditional control (e.g., highways(. The driver does not have to be ready, the vehicle can intervene on its own in the case of system failures |
| level 5 - Full Driving Automation | The vehicle can drive fully automated on its own and a driver is not needed. |

In this regard, users can find themselves in three different roles when using level 3 and above AVs (Saleh et al., 2017, p. 2). (1) The user is the driver or passenger of an AV. (2) The user is classified as a vulnerable road user. These can be passers-by who have nothing to do with the actual driving event or even cyclists. (3) The user is a driver of a vehicle that shares the road with AVs. The majority of research has been done in the first perspective, and this perspective is the only relevant one in this report.

## 2.3 Trust

The abstract concept of trust has been researched for many years and can be viewed from different perspectives (Mayer et al. 1995, p. 712; J. D. Lee & See 2004, p. 50). Defining the concept of trust is a difficult task and a single definition is not expedient considering the multidimensional facet of trust and the dependency on context. One facet of trust is the psychological perspective (in terms of interpersonal relationships) and the following, most adopted definition by Mayer, Davis, & Schoorman (1995, p. 712)

> *"[...] the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (p. 712).*

In this definition, the party who trusts (trustor) expects a positive outcome or a positive future behavior or intention of the party who has to be trusted (trustee) (Kaur & Rampersad, 2018, p. 90). Hence, building trust is always associated with risks, uncertainty, and possible failure of collaboration (Mcknight & Chervany 2000, p. 827; Mishra 1996, p. 25). Mayer, Davis, & Schoorman (1995) discovered that in order to form trust, a trustee must have three personality traits: *Ability*, *benevolence*, and *integrity* (p. 717). Ability refers to the group of skills, competencies, and the personality that influence trust within a specific domain. Benevolence refers to the trustee's good intentions toward the trustor. The conformity of the trustor's behavior toward the trustee and his principles is referred to as integrity. Mcknight & Chervany (2000) added predictability to this view. Predictability refers to the consistency with which the trustee's actions are carried out. The concept of trust has been recognized as an important field of research in technological problems and automation, which is the other facet of trust, the technological perspective.

## 2.4 Trust in Automation and Autonomous Vehicles

Although some people intuitively think of trust in terms of interpersonal relationships, trust is a critical factor in human-computer interaction such as automation technology. The most common definition of trust in automation is the *"[...] attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability"* (J. D. Lee & See, 2004, p. 54). In the evolution of the study of trust in automation, several studies are influential. First, Rempel, Holmes, & Zanna (1985, p. 111) proposed a model of the dynamics of trust between people. Their model introduces three main components: *Predictability*, *dependability* and *faith* by which different perspectives of trust are reflected. Predictability describes the consistency of recurrent behavior and the ability to forecast individuals' behavior by the trustor (Rempel et al.,

1985, p. 96). Dependability refers to the characteristics and qualities of the trustee and not to certain behavioral events that foster trust (Rempel et al., 1985, p. 96). Faith is the essence of trust that comes in when no anticipation of prior experience has been encountered(Rempel et al., 1985, p. 96). B. M. Muir (1994, p. 1912) used the previous dynamic interpretation of trust and extended the model from interpersonal relationships to human-machine relationships. B. M. Muir (1994, p. 1917) proposed that trust in the relationship between humans and machines evolves alongside different bases of trust that change through experience with systems over time due to changes in their mental model. Based on the work by J. Lee & Moray (1992, p. 1247) and the three bases of trust in automation, namely *Performance*, *process* and *purpose*, J. D. Lee & See (2004, p. 59) propose a general theory and the most influential theory for trust in automation. Performance describes what an automation system does to achieve a goal and is based on the system's skill and knowledge, while process describes how the system was created and is connected to its algorithms, and purpose describes why the system was created (J. Lee et al., 2016, p. 204). Using the three bases as a foundation, it was discovered that context is a critical component of trust in automation (J. D. Lee & See, 2004, p. 70). To summarize the key aspect of trust, whether from the psychological or technical perspective: Trust is given by a trustee and corresponds to an expectation of something or someone (Sun et al., 2020, p. 1171). The underlying idea remains the same and is mostly linked to terms such as "firm belief," "decision making," and "behavior." (Cioroaica et al., 2020, p. 22).

Aside from the previously mentioned trust in automation, wherein humans might not be in life-threatening situations, the field of AVs is unique due to various critical factors like risk, uncertainty, and vulnerability (Ekman et al., 2018, p. 96). In other words, *"AVs (1) involve very high physical risk, (2) automate a highly complex human task, and (3) reflect a step-change from augmentation to automation"* (Koester & Salge, 2020, p. 4). Unlike technology that does not create life-threatening situations, AVs require a high level of trust before they can be used, the so-called initial trust. Hoff & Bashir (2015, p. 421) propose a model that is widely used in AV trust research (e.g., T. Zhang et al. 2019, p. 210; Ekman et al. 2018, p. 95; J. D. Lee & Kolodge 2020, p. 261) which differentiates trust before and during interaction with automation technology (Raats et al., 2020, p. 2). For instance, showing users the AV's capabilities before letting them use it themselves can boost initial trust. While it is evident that machines are superior in many aspects of life, this is not necessarily the case with AVs because driving a car is extremely intuitive and needs a lengthy and methodical process of acquiring driving expertise (Koester & Salge, 2020, p. 4). This implies that before using and trusting an AV, consumers must be totally convinced of its capabilities. As a result, understanding trust-influential characteristics is critical in order to develop trust.

However, it must be noted that increasing trust is not necessarily the right approach.

Different AVs are not equally competent nor safe nor are AVs in general completely save. Therefore, users have to be cautious when trusting (B. M. Muir, 1987, p. 535). Users should base their understanding of AV trust on the system's function, usability, and operation (Sun & Zhang, 2021, p. 28214). Consequently, users have to trust AVs appropriately and use them properly (Wintersberger et al., 2019, p. 4). An inappropriate level of trust means the user is overestimating the AV's capabilities, which leads to over-reliance and is called *overtrust* (Yokoi & Nakayachi, 2021, p. 1466). Another inappropriate level of trust is *undertrust* which means users underestimate the AVs' capabilities, which leads to underreliance (Yokoi & Nakayachi, 2021, p. 1466). To cope with overtrust, users must be taught the capabilities and restrictions so that trust may be reduced. Undertrust can be increased by considering trust-increasing methods. The domain of trust research calls this process *calibration of trust*. The trust-calibration approach tries to calibrate users' trust based on the capabilities of the technology (B. M. Muir, 1987, p. 536). J. D. Lee & See (2004, p. 55) described the process of trust calibration as: *"calibration refers to the correspondence between a person's trust in the automation and the automation's capabilities"* (p. 55).

# 3 Methods

We conducted a literature review by following the established guidelines for literature reviews by Webster & Watson (2002), and Kitchenham & Charters (2007). We divided the literature review into three distinct stages: *plan*, *conduct* and *report*. During the *plan* stage, we determined the need for a literature review and created a research protocol based on which the TA was conducted. We then integrated the established TA guidelines by Braun & Clarke (2006) during the *conduct* stage in order to analyze the literature and generate themes and sub-themes by which trust can be influenced. A TA according to Braun & Clarke (2006) is a six-step process consisting of the following steps: 1.) Familiarization with the data 2.) Generation of initial codes 3.) Generation of themes 4.) Review of themes 5.) Theme definition and naming 6.) Writing up the report. We removed the final stage of the TA due to the overlap with the literature review in order to cover the breadth of the issue and effectively answer the RQ. Lastly, we report our findings by addressing the RQ by using the information derived from the themes we actively generated based on the data by using the methods of the TA in an analytic and interpretive way.

## 3.1 Data Collection

We collected our data based on the predefined protocol of the literature review by gathering RQ-related keywords by analyzing highly relevant research articles in the domain of trust in automation, as shown in table 2 by initially searching Google Scholar. We derived relevant databases (DBs) and keywords accordingly and added additional established DBs from computer science, information systems, and related fields to gather relevant literature.

We searched the following established DBs for computer science and related fields: ACM DL (A) , IEEE Xplore (B), Science Direct (C), AISeLibrary (D), Web Of Science (E), and arXiv (F). Using arXiv (F) as DB allows us to gain more insight into present knowledge and work in process literature. However, the fact that arXiv has no quality control was carefully considered. Sage Publications and Taylor & Francis Online were considered by cross-references by Web of Science (E).

**Selection of search strings:** In January 2022, we queried relevant DBs to identify potential literature. Hence, the following search string is used: *Trust AND (Automated OR Autonomous OR Self-driving OR Driver-less) AND (Car OR Vehicle*)*. Potential problems regarding different spellings, for example, *self driving* or *self-driving* were considered.

**Formal selection criteria:** (1) Articles must be written in the English language (2)

Only peer-reviewed articles are included (3) Carefully considered preprint articles from arXiv are included from 2019 - 2022 (4) Articles must be legally available

**Thematic selection criteria:** (1) Only articles with a focus on AVs in the use-case of non-rail-bound vehicles (with a focus on cars) are included. (2) The main focus of the articles is trust in AVs, not trust in automation, trust in general or acceptance of AVs. (3) The main focus of an article can be on one specific characteristic of trust. (4) The results must be applicable to AVs at an SAE level of 3 or above.

***Table 2:*** *Luminary Articles*

| **Reference** | **Citations** | **Databases** |
|---|---|---|
| Hengstler et al. (2016) | 398 | ScienceDirect |
| Waytz et al. (2014) | 759 | ScienceDirect |
| J. D. Lee & See (2004) | 3720 | SAGE Publications |
| Hoff & Bashir (2015) | 1191 | SAGE Publications |
| Choi & Ji (2015) | 588 | Taylor & Francis Online |

## 3.2 Data Analysis

The data analysis consisted of two parts: 1.) Including or excluding literature according to the predefined formal and thematic selection criteria by applying them to the title, keywords and abstracts of the articles, found by the DB search. The selection criteria were then applied to the remaining articles in full-text, excluding further articles from the literature review. Further articles were included based on forward and backward reference searching. Forward searching finds additional literature by looking up which articles have cited the specific paper of interest. By backwards searching, additional literature can be found by examining the references citing a specific article. 2.) Analyzing the articles based on the approach of the TA by Braun & Clarke (2006) by following the steps.

**Analysis and Classification Process:** In order for a TA to work, a sound DB on the desired topic must be available so that the topics can be derived inductively. We gathered the data by reading all of the studies and coding each text passage of interest that mentioned trust or trust-influencing characteristics. The initial coding was an iterative process, and no annotation scheme was used. An example of initial codes, final codes and the respective theme can be seen in Table 3. In addition to that, we tracked the meta-characteristics of the articles. These meta-characteristics include the perspective of

the study (psychological or technical), the kind of study (driving simulator, interviews, surveys, etc.), and how trust was measured.

***Table 3:*** *Example of the Coding Process*

| Content | Initial Code | Sub-Theme | Theme |
|---|---|---|---|
| For this purpose, their recommendation was to make sure that the feedback is relayed in a simplified manner that is understandable by all of those involved(Filip et al., 2017, p. 137) | Feedback must be understandable | Explainability | Trust in the Technology |
| The results indicated that the combination of a human-like appearance with high autonomy led to higher ratings of perceived trust, safety, and intelligence (Niu et al., 2018, p. 2) | Appearance must be human-like | Anthropomorphism | Perceived Intelligence |

The TA's sub-themes were produced by looking for repeating patterns in the DB's codes. This step was then performed iteratively to replace the initial codes with appropriate names and to create a hierarchy with the individual final themes. Note that the sub-themes are referred to as (trust-influencing) characteristics in the following to answer the RQ accordingly. Additional Material can be be found in the appendix in Table 4 - Table 10.

## 3.3 Recognition of Characteristic Interdependencies

As a rule, interdependencies between individual characteristics influence user trust in AVs. This is because driving with AVs has many influencing factors, such as one's safety, experience with technology, or even demographic factors. To find these interdependencies, codes or characteristics and themes frequently appearing together in the literature are reviewed and validated against the literature base. These interdependencies are dicussed in section 5.2

# 4 Results

In the following section, we discuss the descriptive results by describing the analysis of meta data of the found articles, followed by addressing the RQ from the technological and psychological perspective, including a comprehensive visual in the form of a topic map to summarize our findings.

**Descriptive results:** Our DB search has found 1631 articles spread over all DBs. Web of Science yielded the most articles (548), followed by IEEE Xplore (539), ACM DL (280), ScienceDirect (117), and arXiv (75). Contrary to planning, AISeLibrary could not be used for DB search due to licensing problems. After eliminating 204 duplicates, we were left with the remaining 1427 articles. In the next step, we applied the formal and thematic selection criteria to titles, keywords, and abstracts to further remove 1298 articles. The reasons for exclusion were: trust in different systems, such as recommender systems or trust as a technological component (confidence), trust was only mentioned, trust research in the wrong domain (rail-bound or unmanned aerial vehicles) or false positives. The remaining 129 articles were then analyzed in full text, of which 80 articles were excluded from the literature review. In the final step, a forward and backward search was conducted, which yielded 5 more articles, for a total of 55 as shown in Figure 1.

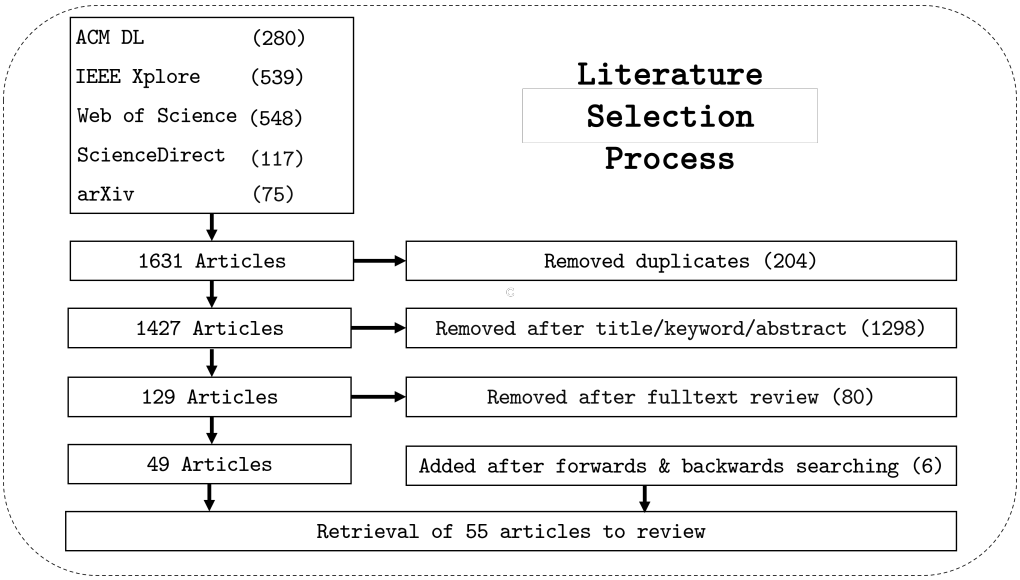



***Figure 1:*** *Process of the literature selection. The DBs and the retrieved articles can be seen in the first box.*

Of those 55 articles, 19 (35%) are from a psychological perspective and 36 (65%) are from a technological perspective. The distribution of the number of articles over the years from both perspectives is shown in Figure 2. Despite the fact that no time boundary was applied for the publication date, the first article we included was published in 2013. Considering the fact that trust research in AVs is still a growing field and assuming

research results to be applicable to SAE-level 3 AVs, these results are plausible. The rapid increase in the number of publications confirms our assumption of a young and rapidly growing field of research. A combined distribution of the number of articles is shown in Figure 3. The most common study is the driving simulator (34 articles), where

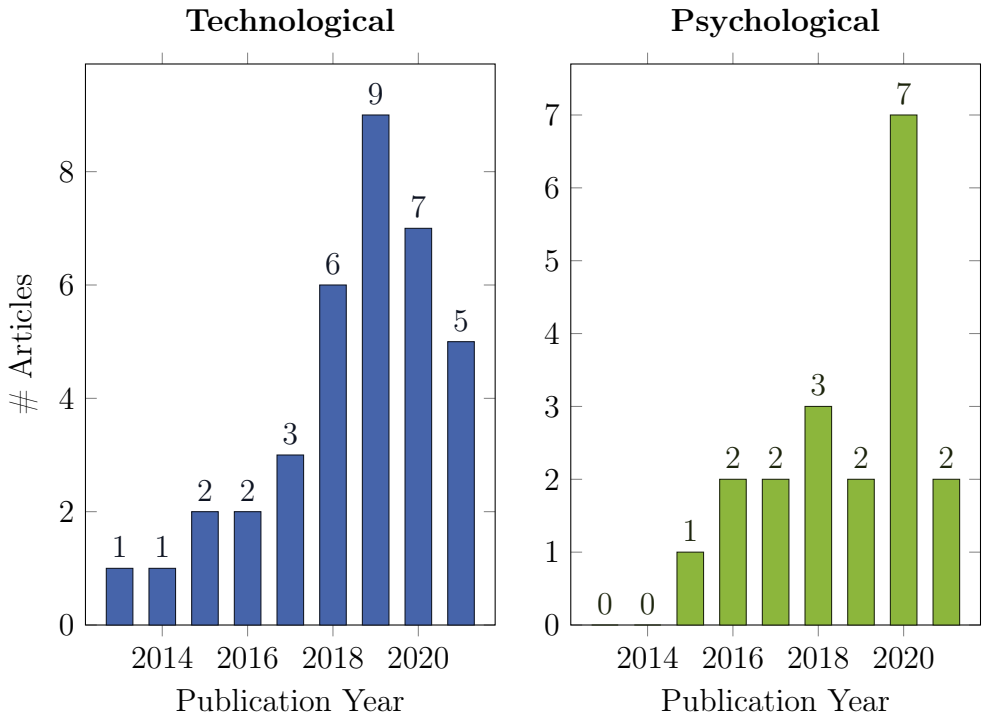



***Figure 2:*** *Distribution of the number of articles over the years. The left distribution shows the number of articles from the technological perspective, while on the right, the distribution of the articles from the psychological perspective is shown.*

different methods were used, for example, virtual reality, using projectors, or different types of screens. The driving simulator has been exclusively used from a technological perspective. Other methods were surveys (8), framework building (4), Wizard-of-Oz (3), interviews (1), thematic analysis (1), inductive case studies (1), questionnaires (1), and a conjoint analysis (1), as shown in Figure 4. In order to analyze changes in trust, all articles measured trust by using a combination of approaches or on their own. The only often cited method is the trust checklist introduced by Jian, Bisantz, & Drury (2000) with 16 mentions across all articles. They measure subjective trust by asking twelve questions on a 7-point scale (not at all = 1, extremely = 7) about participants' feelings regarding the system. For example, the system is deceptive; I am wary of the system; the system is reliable. Other trust measure methods include measuring physiological data (electrodermal activity, heart rate, gaze behavior), collecting feedback from interviews (structured, semi-structured, and unstructured), think-aloud procedures, observations, and surveys.

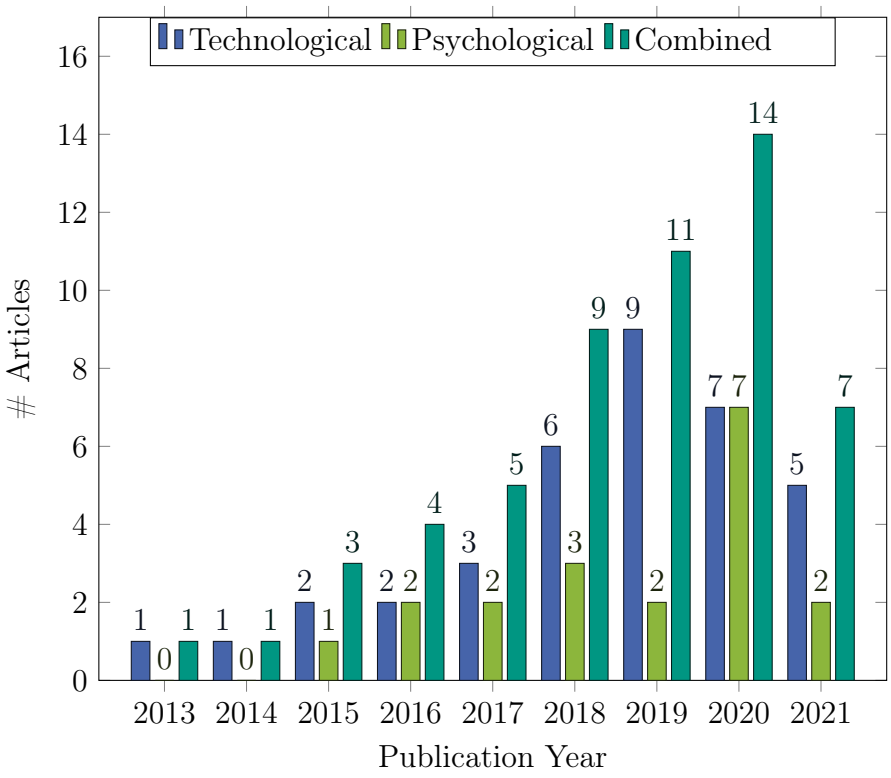



***Figure 3:*** *Distribution of the number of articles combined in dark green, from a technological perspective in blue, and from a psychological perspective, in light green.*

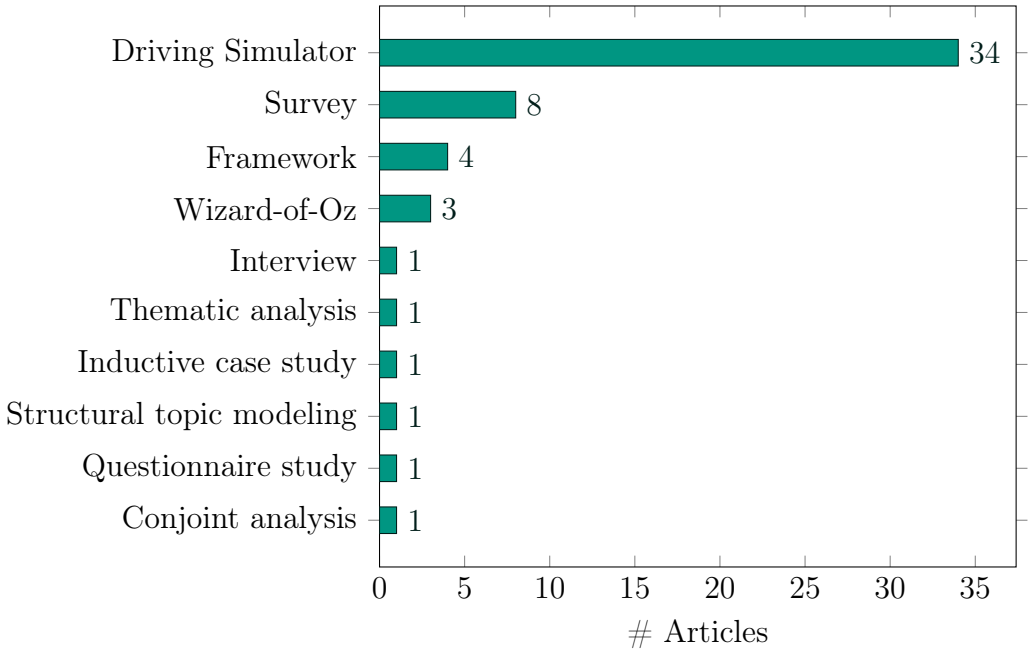



***Figure 4:*** *Number of types of articles. Different types of simulation approaches (virtual reality, projectors, or screens) were grouped together as driving simulators. Some articles used mixed approaches.*

As stated in the beginning, the goal of this thesis is to create a holistic picture of trust influencing themes and characteristics from a psychological and technological perspective that directly impact the acceptance of AVs. The holistic picture can be seen in Figure 5 in the form of a topic map. During our analysis, six core themes emerged. Three of them from a psychological perspective and three from a technological perspective. These are: *Trust in the Technology*, *Trust in the Manufacturer*, and *Trust in the Legislative* (psychological perspective), *Information Exchange*, *User Perception and Comprehension*, and *Perceived Intelligence* (technological perspective). We tried our best to ensure diverse and distinctive themes and characteristics with a minimal degree of overlap. However, some characteristics overlap to some degree.

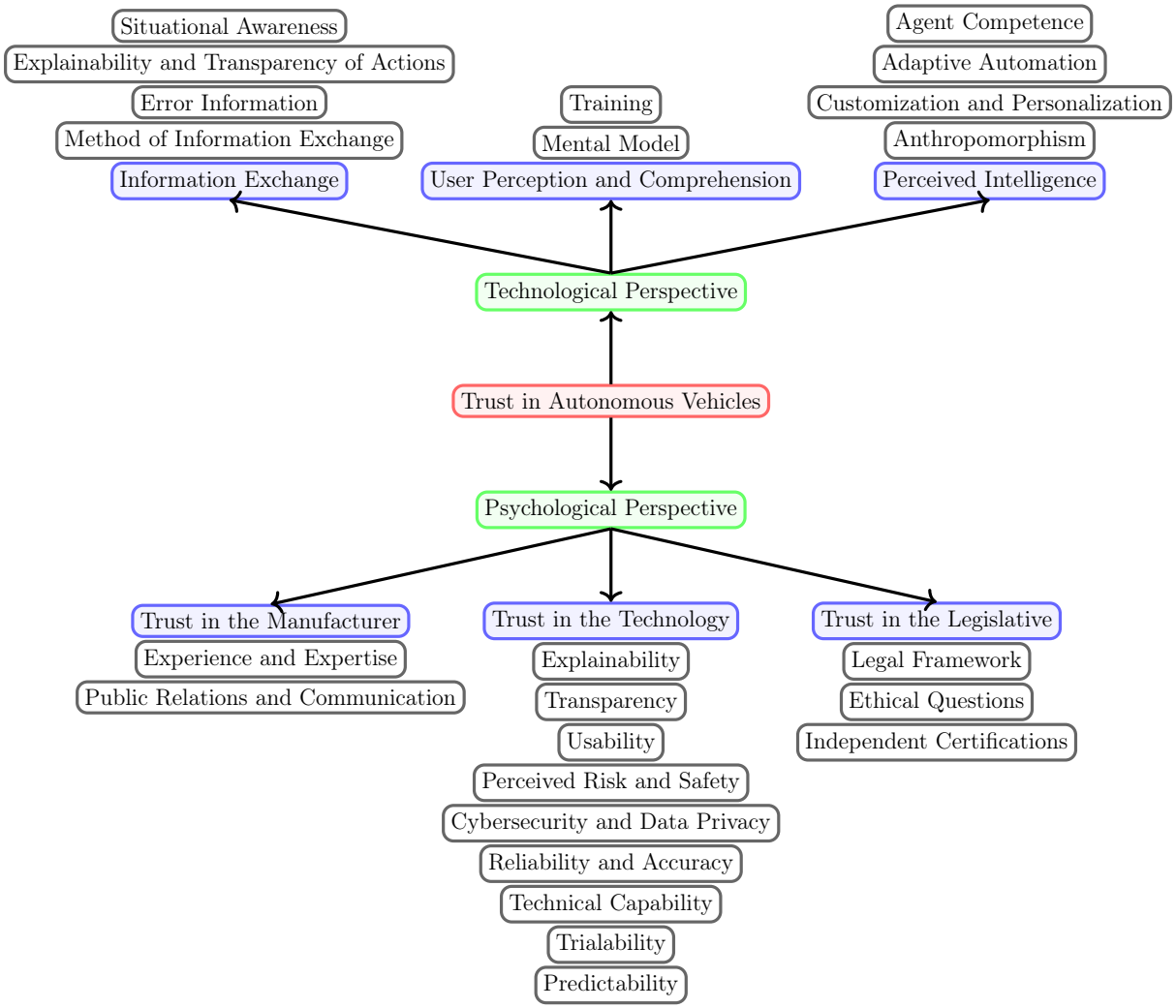



***Figure 5:*** *Holistic picture of trust in AVs in form a of a topic map. The red box shows the main goal, green indicates the perspectives, blue represents the generated main themes, and white the characteristics.*

## 4.1 Psychological Perspective

The psychological perspective is argued in the form of beliefs, thoughts, and fears of people when interacting with AVs and is collected via (semi-or unstructured) interviews or surveys, in which people can freely talk about their perception of AVs. For example, in the study of Buckley et al. (2018b, p. 172) a participant talked about trust in AVs and reported: *"I would completely trust it in the scenario that just played out, but I think that in, in a real time environment, I think that there are just too many of those, too many of those variables for it to be just a perfectly spotless system, you know, so it always seems that there's always that degree of risk"*. In this case, the participant was talking about perceived risk when directly interacting with an AV, which yields a sub-characteristic of "perceived risk". Hence, to cover all points of influence of trust from a psychological perspective, three themes were generated: *Trust in the Manufacturer*, *Trust in the Technology*, and *Trust in the Legislative*.

### 4.1.1 Trust in the Manufacturer

The second first theme, *Trust in the Manufacturer* describes the public perception of the AV manufacturer, which he has built up over the years. This includes the experience and competence of the manufacturer in the eyes of the population, as well as the communication strategy (e.g., the influence of brands) and the intention behind the AV manufacturing project. Companies often think that the quality of their products is the only criterion for their success, while marketing and commercialization are overlooked (Slater & Mohr 2006, p. 26; Hengstler et al. 2016, p. 106). For this reason, it is important to look not only at technical factors but also at factors such as the image or the nature of public communication of companies to build trust. The better the reputation of the car manufacturer is, the more likely it is that people trust the AV (Carlson et al., 2014, p. 24)

**Experience & Expertise:** *Experience and expertise* refers to the quality and recognition of the manufacturer's technology already established on the market can be seen as a benchmark for future technology. For example, well-known car manufacturers such as Daimler may be trusted more than newcomers to the market because their vehicles are well-known throughout the world. For example, a study by Carlson et al. (2014, p. 26) has shown that people trusted an AV more because Google manufactured it and people associated Google with good experiences. Hence, trust can be built upon predecessor technology that is already in use, such as automated driving assistance systems, showing their expertise in these domains (Koester & Salge 2020, p. 7; Köster & Salge 2021, p. 5).

**Public Relations & Communication:** In addition to experience in the development of technology, *public Relations and communication* between the company and the public

are also of important significance. Even if manufacturer awareness (e.g., Daimler) is a factor, the brand itself is typically more well known and recognizable and is associated with certain expectations (e.g., Mercedes-Benz). The Performance and quality of products are often closely associated with brands and are linked to how reliable and successful products are for their specific purpose (Celmer et al., 2018, p. 1762). The expectation of performance is based on the attributes such as *intent, methods, competence, and history* of a brand, which are closely linked to trust-building mechanisms in AVs (Celmer et al., 2018, p. 1763). Hence, manufacturers whose brands are associated with good quality must recognize their position and fulfill the positive expectations that are linked to them, otherwise people may not trust the AV. This can be achieved by communicating the development process through news or other channels and addressing risks and opportunities to strengthen the companies' credibility (Hengstler et al., 2016, p. 111). However, it is also necessary to pay attention to the choice of words (e.g., instead of AI choosing autonomous as a word) and that the communication process begins early enough (Hengstler et al., 2016, p. 112).

### 4.1.2 Trust in the Technology

The second main-theme *Trust in the Technology"* describes the relationship between people and their perception of the technology of AVs. The technical components of an AV are diverse and extremely complex. In addition to the classic safety aspects of conventional vehicles such as airbags and seat belts, hardware such as computing units and numerous sensors must function perfectly in even the most severe situations and guarantee the safety of all involved (Koester & Salge, 2020, p. 5). In order to achieve the highest safety standards, hardware and software must synergize flawlessly.

**Explainability:** *Explainability* refers to the information given to the users by the AV about internal decisions taken and actions made. With explanations, feelings of fear and uncertainty can be tackled by providing information about why the AV has taken a specific action or decision by explaining it. Furthermore, explainability enables users to gain a deeper understanding of the system and its internal procedures (Linardatos et al., 2020, p. 3) which increases their confidence and fosters trust (Omeiza et al., 2021, p. 194). However, it is important to choose which information is conveyed. Not all users are equally educated and perhaps lack the required knowledge to understand all explanations given by the AV (Wintersberger et al., 2020, p. 253). In addition to that, not all driving scenarios allow equal explanations of situations. When hazardous situations occur, the AV must quickly provide explanations that are precise and easy to understand, while normal driving situations allow more detailed explanations. Hence, different types of explanations for specific driving situations are used. First, simple explanations provide information about which actions were executed (Ha et al., 2020, p. 275). Explanations are

not formulated in full but only give the most important information. Second, attributional explanations allow to convey more information than necessary and may increase trust due to the preferences of users (Ha et al., 2020, p. 275). Ultimately, the AV must be able to decide, depending on the context, which information needs to be explained and how, in order to clarify situations that arise in the best possible way.

**Transparency:** While explainability provides information about why actions and decisions were made, *transparency* provides information about the inner workings behind them. Explainability and transparency are highly connected. Some scholars do use explainability and transparency equally in terms of providing transparent information about the decisions of a system. Hence, trust can be increased when users can understand how the inner workings of a system come to decisions (Choi & Ji, 2015, p. 699). Providing information about the inner workings of AI and machine learning algorithms remains problematic since the black box properties make it difficult, even for developers, to understand how decisions are made. While laypersons may not understand how decisions are made, transparency of how the AV was trained, which data has been used, which model has been tested, etc. fosters trust due to the transparent development process and transparent information about the inner workings (Shin, 2020, p. 554). However, legal frameworks must be observed as well as business secrets and privacy must be protected.

**Usability:** The *usability* of an AV generates trust through well-designed interfaces and is often only perceived subconsciously. Interfaces must be designed in such a way that users can understand the technology easily and intuitively without having to overcome obstacles (Hengstler et al., 2016, p. 110). Attributes of usability include learnability, efficiency, as well as error handling and satisfaction (Morgan et al., 2017, p. 328), when interacting with an AV. Definitions include: *"[...] the technology does what the users expect it to do"* and *"absence of frustration in using something"* (Stadler et al., 2019, p. 205). When designing interfaces, attributes such as the context of use and the age structure of the users must be carefully considered (Hengstler et al., 2016, p. 110). Older users might be overwhelmed by the complexity of the AV, especially when individuals have sensory and/or physical impairments (Morgan et al., 2018, p. 592). This can lead to cognitive overload and may decrease trust in AVs. Accordingly, the accessibility for each population group must be set up to ensure usability.

**Perceived Risk & Safety:** *Perceived risk & safety* cannot be decreased by one specifically due to the many factors that influence it. However, lowering the perceived risk is essential to building a vehicle that users can trust. Perceived risk is an often-underestimated characteristic, especially during users' initial interaction with an AV (Li et al., 2019, p. 177) due to uncertainty and risks of physical harm. It can be seen as an evaluation by the user regarding the probability of negative events and the seriousness of negative events while interacting with an AV (Li et al., 2019, p. 178). Trust increases when perceived

risk decreases (Pavlou, 2003, p. 112), and if the AV behaves as the user expects it to, the perceived risk will decrease (Choi & Ji, 2015, p. 694). Perceived risk is high when, for example, the visibility is low and users cannot predict what happens in the next situation due to their experience in these situations (Ekman et al., 2020, p. 67). For these reasons, it is critical to pay sufficient attention to the perceived risk characteristic, even if only indirect influence can be exerted, for example, by increasing the explainability or transparency of the AV. In addition, the perceived risk can be reduced by safety measures. It must be ensured that physical integrity is maintained through cutting-edge technology. These safety measures include not only the materials of the AV, etc., but also the algorithmic and decision-making processes, which must be robust under all circumstances.

**Cybersecurity & Data Privacy:** *Cybersecurity* as a characteristic to build trust describes the increasing vulnerability of AVs due to the complexity and interaction of hardware and software components (Linkov et al., 2019, p. 1). These issues include *data privacy* and security (e.g., hacking, data security) related concerns as well (Buckley et al. 2018b, p. 173 ; Hengstler et al. 2016, p. 108). Privacy concerns are linked to handling personal data (Large et al., 2019, p. 50) and the potential misuse of data by public or private individuals (Sun et al., 2020, p. 1182). Privacy concerns may be avoided by anonymizing and encrypting data streams. Security concerns include intentional attacks on the AV's system, the takeover of the AV itself, or information about damaged components (e.g., sensors, software bugs, etc.) of the AV. For example, the AV as an AI-enhanced system has a large attack surface, especially when data streams are manipulated and the algorithms make different decisions with this data and potentially endanger the system's environment. Technical solutions to prevent this type of attack are to use critical data exclusively from the AV itself, without data flowing through other systems, or to encrypt data using the latest encryption technologies. In addition to technical issues, human factors as one of the biggest security gaps in AVs must be considered carefully, due to the lack of knowledge, awareness of potential critical situations, or their overtrust in the car (Filip et al. 2017, p. 2, Parkinson et al. 2017, p. 9). The AV must have a certain robustness to resist attacks. Thus, trust can be increased by developing security standards and by protecting the safety paths and sensor systems of the AV (Hengstler et al., 2016, p. 108). Another way trust can be increased is by explaining how data is collected and for what purpose it is used (Sun et al., 2020, p. 1182). This can ensure that users understand the data flow and may be able to make their own decisions regarding the use of their data.

**Reliability & Accuracy:** The belief that the performance of the AV is consistent and precise is referred as *reliability* (Kaur & Rampersad, 2018, p. 92). In other words, when the AV is reliable, the likelihood of failure is low (Zhao et al., 2019, p. 13). Users seem to trust AVs when the performance of the AV is perceived as at least equal to their ability to drive a vehicle (Merritt et al., 2013, p. 524). Serious errors that call the vehicle's

reliability into question are fatal to building trust. The reliability of AVs is questioned when unpredictable situations occur and technology may reach its limits in so-called edge cases (really rare events or events that have not been seen before), for example, spotting people during a thunderstorm when the weather is foggy (Omeiza et al., 2021, p. 197). In other words, the AV is reliable when it makes the right decisions with a high degree of *accuracy*. It is critical that both reliability and accuracy can always be considered stable, which leads to an increase in trust (B. Muir & Moray, 1996, p. 441). This can be achieved by knowingly implementing redundancy in the form of backup systems and testing the AV thoroughly (Koester & Salge, 2020, p. 6, 11). Reliability and accuracy may suffer if unrepresentative data sets are used to train the AV. For example, AVs that were trained by using German road data might not perform well in the UK or Australia due to different traffic rules or different landscapes.

**Technical Capability:** The characteristic *technical capability* summarizes the capabilities of the AV in terms of the features required and the competence to solve the desired problem (Choi & Ji, 2015, p. 694). In other words, the functionalities an AV has implemented to be able to solve problems are the technical capabilities of the AV. These can be capabilities such as "behave as expected", "drive to speed", or "behave as a driver", which show that the AV has the technical capabilities to deal with all possible situations and may even mirror human behavior (Buckley et al., 2018b, p. 170). Other capabilities include real-time object detection and path planning obeying the rules of the road (Khastgir et al., 2018, p. 295). To build an appropriate level of trust, the AV must be able to communicate its capabilities to prevent people from trusting the AV not enough (undertrust) or completely (overtrust) (Khastgir et al., 2018, p. 301). Trust can be built by incorporating features that convey security and competence and communicating these clearly to create a basis of trust between the AV and the user.

**Trialability:** *Trialability* refers to the opportunity to have access to AVs and experience their capabilities and functionalities personally (Siau & Wang 2018, p. 51; Lee et al. 2011, p. 126). Before trusting AVs, users want to see them work first (Buckley et al., 2018b, p. 170), whereas the understanding of AVs can also be increased. For example, the public can engage with AVs through test drives, demonstrations, training, or presentations when companies offer these opportunities (Yuen et al., 2020, p. 10). If companies create these risk-free environments, interested users can deepen their understanding and experience AVs personally, which can lead to an increase in trust in AVs.

**Predictability:** *Predictability* refers to how well the AVs' actions can be predicted from moment to moment (J. Lee & Moray, 1992, p. 1246). Even if the predictability of the actions of AVs is not always obvious, since the respective situation plays a major role (e.g., driving on the highway, driving in the rain, etc.), the actions must be deterministic. That means that the reaction of the AV to situations should always be the same. Hence, even if

dangerous situations or errors occur, the user can understand and predict the action of an AV in each situation, which in turn can increase trust in the AV (Khastgir et al., 2018, p. 291). However, predictability alone might not ensure appropriate trust, the information given by the AV has to be consistent with the cognitive processes of the user during the interaction (Ekman et al., 2019, p. 277). Conversely, if the AV behaves unpredictably (e.g., speeding, tailgating, or sudden lane changes), trust will almost certainly decrease (Filip et al., 2017, p. 518).

### 4.1.3 Trust in the Legislative

While the other two main themes of the psychological perspective deal with either the technology of the AV or the manufacturer behind the AV, the theme *Trust in the Legislative* deals with the legal framework conditions that must be in place for AVs to be trusted. In this context, questions of accountability in the event of accidents are of interest, as are legal framework conditions relating to data security and privacy. In order to clarify the question of accountability, independent control mechanisms must exist that make the AV appear trustworthy. This can be solved by independent certifications. Moreover, trust can be built when ethical and moral issues are clarified.

**Legal Framework:** *Legal framework* describes the laws, standards, and rules all involved parties in the building process of AVs have to fulfill in order to build safe AVs. Even though AVs will be very safe in the future and there may be fewer accidents, it is inevitable that accidents will happen. This does not necessarily have to be due to the capabilities of the vehicle, but can also be due to the carelessness of pedestrians or other road users. When the user is no longer in control of the AV (often called Human-on-the-loop), how does the legal framework handle these situations when all decisions are made by the AV (Kyriakidis et al., 2015, p. 6)? In these cases, for example, the manufacturer or the developer of the software can be held responsible, since the control of the vehicle is not the responsibility of the user. Since there are so many decisions, each one is independently determined by the AV and cannot reasonably be double-checked manually. The liability, accountability, and regulation barrier is one of the most difficult barriers to define when adopting AVs (Penmetsa et al., 2019, p. 10). As of now, no clear legal framework exists which apportions the liability between third parties (manufacturers, programmers, suppliers etc.) in the case of accidents (Taeihagh & Lim, 2019, p. 8). To make AVs a sustainable mobility alternative that users can trust, legal frameworks and regulations need to be internationally established.

**Ethical Questions:** In addition to possible technical legal provisions, *ethical questions* and moral issues must also be clarified in order for people to trust the behavior of the AV. For example, if accidents are inevitable, programmers might have to design "crash

algorithms" which determine who might be harmed in these situations (Taeihagh & Lim, 2019, p. 8). An prominent example of moral question is *"the trolley problem"* (Thomson, 1985, p. 1395-1415), which describes a situation in which a decision has to be made about whether a group of people or a single person is killed, when killing is in inevitable. When decisions are made in a dynamic environment, the AVs logic behind such algorithmic decision-making in uncertainty must be understood clearly. One troubling example is that corporations might prioritize the safety of those who are now in the AV over the safety of other road users. Hence, AV decisions must be unbiased, nondiscriminatory, and have to be in line with human rights. Discrimination does not have to be intentional, but it may be apparent in the data sets obtained. Data sets must be created with extensive statistical understanding and care to avoid discrimination based on acquired data. Individuals, for example, may be denied AV services in the future because they have an uncommon name that has drawn negative attention in the past, and this is reflected in the data. Consequently, establishing clear legal frameworks for technical questions as well as moral and ethical questions can increase trust in AVs (Bruckes et al., 2019, p. 4).

**Independent Certifications:** Another way to increase trust in AVs is through *independent certifiers* (Koester & Salge, 2020, p. 7). Independent certifiers, for example, look at the safety and capability of the AV's technology or confirm that AVs have already traveled a certain distance without an accident (Fagnant & Kockelman, 2015, p. 176). When demanding certification processes are completed and the AVs are considered certified, this has a strong external impact and may increase trust. These certifications demonstrate that official standards have been successfully implemented and that user safety is ensured (Raj et al., 2020, p. 131). To ensure that certifications clearly express the AV's strengths and any potential shortcomings, certain sub-areas like explainability, transparency, cybersecurity, etc. should also be certified in addition to the AV as a whole. These certifications can be made through various audits. The external impact can be further strengthened if the certifiers are given a direct insight into the technology to understand the internal methodologies and algorithms (Koester & Salge, 2020, p. 7) or are part of the early development process (Köster & Salge, 2021, p. 9).

## 4.2 Technological Perspective

While the psychological perspective is very broad and looks at many areas, the technological perspective is mostly driven by design features in the area of human-machine interfaces and is not as broad. These trust-building characteristics go deeper into technological possibilities and are less about capturing the broad spectrum of users' thoughts or fears compared to the psychological perspective. Thus, of particular interest are the implementations of the respective trust-building characteristics and how users react to

them instead of capturing general thoughts about AVs. These can be special displays and interface designs that are used for communication or intelligent agents that serve as driver substitutes. For this perspective, three main themes were generated: *Information exchange.*, *User Perception and Comprehension*, and *Perceived Intelligence.*

### 4.2.1 Information Exchange

The first main theme of the technological perspective describes the importance of providing various types of information during an interaction between an AV and a user and how different ways of information exchange can build trust.

**Method of Information Exchange:** The *method of information exchange* is one of the most critical characteristics for building trust due to the importance of how information is conveyed instead of only considering which information is relevant (Morra et al., 2019, p. 9439). In general, two methods are widely used to convey information: visually and audibly. Other, currently rarely researched methods include: haptically, for example, through specific vibration patterns when using touch displays, different kinds of visual cues such as light patterns, or augmented reality enhanced windshields (Morra et al. 2019, p. 9441; von Sawitzky et al. 2019, p. 1). Besides the individual methods, it is also important how each of the individual methods exchanges information. The visual method mostly uses visual interfaces such as heads-up-displays or common displays. Heads-up-displays enable information to be projected directly into the user's field of view, such as the route being traveled or detected obstacles and objects. Displays, on the other hand, are widely used. Different methods include the birds-eye perspective (showing the car and its surroundings from above), sensor indicators, or ring concepts which indicate the distance to objects (Ekman et al. 2016, p. 2; von Sawitzky et al. 2019, p. 3). Each of these methods is often complemented by an audio component by giving speech output and creating an audiovisual system. However, it is critical to consider that users have preferences and may suffer from a high cognitive load when too many methods are used simultaneously (Morra et al., 2019, p. 9439). Consequently, to build trust, users must not be viewed as one homogeneous group but rather as individuals with beliefs and preferences. Using the wrong method of information exchange may result in a decrease in trust.

**Explainability & Transparency of Actions:** *Explainability of actions* is the technological equivalent of "explainability" from a psychological perspective and refers to the car's ability to give explanations (why) of its actions, whereas the *transparency* refers to how the AV acts. In both cases, a combination of explanations about "how" and "why" seems to increase trust the most (Wintersberger et al. 2020, p. 253 ; Ha et al. 2020, p. 279). With "how," information about actions is provided (e.g., "The car is braking"), whereas "why" provides information about the reasons for an action (e.g., "Obstacle

ahead") (Koo et al., 2015, p. 269). A combined message provides combined information on why and how actions were made (e.g., "The car is braking due to an obstacle ahead") (Ruijten et al., 2018, p. 4). But not only the content of the explanation is critical, but also the timing of the explanation (Stanton & Young, 2000, p. 317). Several studies found that providing information before the AVs action does increase trust (e.g., Forster et al. 2017, p. 370; Koo et al. 2015, p. 270), whereas providing explanations after the AV acted did not increase trust in the AV (Körber et al., 2018, p. 13), which indicates that simply providing explanations is not enough to increase trust in AVs (Du et al., 2019, p. 437). However, even when explanations are perfectly timed, too many or too complex explanations may lead to cognitive overload and can decrease trust in the AV (Koo et al., 2015, p. 273). How explanations can be provided is diverse and solutions of the characteristic "Method of Information Exchange" can be used, for example, conversational or visual user interfaces. In addition to explaining why the AV acted, the steps on how to act (transparency) must also be explained before the action is performed and not while it is being performed. This reduces the uncertainty of subsequent actions and thus lowers the perceived risk, which in turn can lead to more trust (Du et al., 2019, p. 437).

**Error Information:** Besides explaining actions and how they are performed, information about *errors* is critical. Errors during autonomous driving will always happen. To ensure that trust in the AV does not decline due to errors, errors or incidents that were unplanned must be justified immediately afterwards (Dzindolet et al., 2003, p. 697). When an error remains unexplained, the trust between user and AV is violated and might decrease to a lower level (Dzindolet et al., 2003, p. 698), depending on the level of trust before the error occurred. When users are familiarizing themselves with an AV and have not had much experience yet, errors and incidents during the phase of familiarization might immediately decrease trust (Ekman et al., 2018, p. 100). For such situations, a feedback system should be implemented so that users can give their feedback to the developer when they do not understand errors or explanations instead of only an automated feedback system. The disparity between the real situation and the intended situation when critical situations arise, however, is of special concern for the developer (Stanton & Young, 2000, p. 318). To determine whether a hardware malfunction or software error has occurred, internal procedures and algorithms are critical. In this case, it is critical to implement a system that can evaluate algorithmic parameters in the event of errors and store all algorithmic parameters at all times.

**Situational Awareness:** *Situational awareness* in this context refers to the car's ability to sense, position, process, and act according to situations (Filip et al., 2016, p. 1). When the AV can understand its immediate environment, for example, spatial information (e.g., traffic objects) (Sonoda & Wada, 2017, p. 191), trust can be increased when the information is shared with the user. Spatial information provided by the AV allows users

to perceive some part of the decision process of the AV (Sonoda & Wada, 2017, p. 191). Understanding some parts of the decision process behind an action can increase trust. Furthermore, the latest technological advances make it possible to accomplish tasks that humans cannot. One example of this is the vehicle-to-everything or V2X communication, which enables "superhuman" driving (Wintersberger et al. 2019, p. 5; Hobert et al. 2015, p. 64). In an ideal world, vehicle-to-everything enables communication between vehicles, infrastructure, and other networks (smartphones, etc.), which supports the AV to be aware of the situation it is in. For example, when the weather is foggy and the AV is in ambiguous situations, vehicle-to-everything communication can be used to communicate with other road users, even when the human eye cannot see them (Wintersberger et al., 2019, p. 5). Consequently, when the AV can communicate clearly that it is aware of the situation, trust can be increased.

### 4.2.2 User Perception and Comprehension

The main theme *User Perception and Comprehension* outlines, from a technical standpoint, what is required for users to comprehend what an AV is and how their internal perspectives may be leveraged to develop mental models of the AV, which are essential for building trust (Morra et al., 2019, p. 9438).

**Mental Model:** A *mental model* can be understood as the *cognitive compatibility* of a user and can be defined as: *"mapping of the relevant information in the situation onto a mental representation of that information"* (Rousseau et al., 2004, p. 4). A mental model in the context of AVs is the representation of the user's thoughts about the AVs' capabilities and functionalities and contains thoughts about why and how the AV works (Beggiato, Pereira, et al., 2015, p. 77). Even when the AV works effectively, trust may be reduced if there is a discrepancy between the user's mental model and the AV's behavior (J. D. Lee & See, 2004, p. 72). By expanding the users' knowledge of the AV and giving details regarding activities, such as by using visual or audio outputs to explain events, an adequate mental model may be created (Morra et al., 2019, p. 9441). The goal of this process is to calibrate the user's mental model and the capabilities of the AV accordingly to meet the user's expectations as well as internalize the limitations of the AV so that potential hazardous situations can be understood and averted preemptively. This procedure should occur during the learning phase, when users are learning about the AV, in order to develop an unbiased mental model. If the user's mental model of the AV is correct, his knowledge and expectations will be closer to reality, which might foster greater trust between the user and the AV.

**Training:** *Training* might become relevant due to the shift of the role of users from manual to a supervisory role (Saffarian et al., 2012, p. 2298). In contrast to the trialability

characteristic of the psychological perspective, in which the physical exchange between the AV and the interested user is in the foreground, the understanding of the user regarding the AV's capabilities and limits is trained during the training. The training method should take into account characteristics such as the personality and demography of the groups so that the training is an individual process because training is essential to calibrate their trust (Kraus et al., 2021, p. 1099). Training methods include introductory videos and texts, as well as introductory drives (Körber et al., 2018, p. 9), which influence the perceived risk and can increase trust when the perceived risk decreases (Li et al., 2019, p. 177). An appropriate level of trust can be generated when the user's knowledge of the AV is improved not only before the first usage but also after the first usage (Ekman et al., 2018, p. 98). Accordingly, the user's knowledge must be continuously expanded and skills trained so that they can intervene in emergencies and critical situations.

### 4.2.3 Perceived Intelligence

The last main theme *Perceived Intelligence* describes the importance of making the AV more "human-like," which includes properties such as competence and expert behavior. The majority of users might not like robot-like or machine-like movements, especially when the AV's behavior is not natural (e.g., automatic movement of the steering wheel) (J. Lee et al., 2016, p. 205). If the AV can simulate human behavior and mental processes, trust can be developed and maintained.

**Anthropomorphism:** *Anthropomorphism* can be defined as *"[...] the tendency to imbue the real or imagined behavior of nonhuman agents with humanlike characteristics, motivations, intentions, or emotions"* (Epley et al., 2007, p. 864). Anthropomorphism attempts to grasp the tendency of humans to attribute human characteristics to non-humans, especially regarding rational thoughts and conscious feelings (Waytz et al. 2014, p. 113; Gray et al. 2007, p. 619). These attributes are not limited to superficial aspects of people, such as appearance, voice, or body, but include essential characteristics of people, such as behavior, human thinking, and feelings (Waytz et al., 2014, p. 113). In the case of AVs, this is achieved by giving the AV a name, a voice, a gender, and possibly an appearance to make it seem more human. To accomplish this, anthropomorphic human-machine interfaces are used to carry out the essential human characteristics (Forster et al., 2017, p. 366). In most cases, an intelligent agent is created whose appearance and behavior are adapted to the respective situation and the respective user to match the main task. For example, anthropomorphism should be avoided in emergencies so that information is concise (Niu et al., 2018, p. 6). User behavior reflects that anthropomorphism can increase trust in the AV when design features are implemented correctly (Waytz et al. 2014, p. 116; Large et al. 2019, p. 57; Forster et al. 2017, p. 371).

**Agent Competence:** *Agent competence* refers to the AV's ability to demonstrate competence and expertise in the form of an intelligent agent, as people tend to trust expert systems more (Choi & Ji, 2015, p. 420). The system's purpose and performance must align to be viewed as competent, which also relies on the user's expectation regarding the capabilities of the system before the first usage (Ekman et al., 2018, p. 98). When the AV does not perform to the user's expectations, users judge it as incompetent and not intelligent enough yet (J. Lee et al., 2016, p. 204). For example, this may be the case if the AV cannot keep on track properly or if speed limits are ignored, even if they are obvious. Another way to radiate competence is through conversational interfaces, which can be achieved through linguistic skills. A special possibility is integrating politeness, which is an innate human ability to form cooperation and relationships (J.-G. Lee et al., 2019, p. 3). When the AV can convey information politely through conversational interfaces, methods of polite speech can be used. Polite speech increases the perception of social presence and increases trust in the AV (J. Lee & Lee, 2022, p. 107015).

**Adaptive Automation:** *Adaptive automation* describes the AV's capacity to adjust situationally to the given conditions. In the domain of autonomous driving, it is critical to ensure the safety of users by using adaptive systems that can cope with faults and reach safe states (Zimmermann & Wotawa, 2020, p. 1189) and adapts to users to increase trust. For example, users can specify their preferred driving style (e.g., defensive, aggressive, or customized). Defensive driving styles are characterized by using the highest gear possible, avoiding standstills, early indications, and higher distances to objects, while an aggressive driving style is characterized as the opposite of the defensive driving style, where possible (Ekman et al., 2019, p. 270). A customized driving style can be individualized by the users. While an aggressive driving style should be avoided in most cases (Shahrdar et al., 2019, p. 520). However, some users prefer a more aggressive driving style (Ekman et al., 2019, p. 270). Another important point of adaptive automation is that the shared driving goals of users and AV (e.g., eco-friendly, time-saving, safety, comfort, etc.) must not be violated (Verberne et al., 2012, p. 808). To make this work, the user must explicitly discuss this intention with the AV. Hence, adaptive automation can generate trust, especially if the AV can ensure safety and can adapt according to user goals (Verberne et al., 2012, p. 809)

**Customization & Personalization**: In contrast to adaptive automation, *customization and personalization* refers to changing the non-critical systems according to their preferences. Customization refers to user-made choices whereas personalization refers to data-reliant adaptations. Each person's personality is different, and the experience with an AV must be tailored accordingly to generate trust (Kraus et al., 2021, p. 1099). For example, anthropomorphic avatars whose gender, appearance, and manner of speech can be customized according to the user's wishes (Ekman et al., 2018, p. 99). An interactive

learning curve between the user and the system is made possible via customized interfaces. Users will typically require more information in early phases of experiencing AVs to establish trust than they would if they were already familiar with AVs. In addition, interfaces should be adaptable to capitalize on human nature's individuality, as manufacturers already do with customizable vehicles available for ordering. A feature that could potentially be personalized (within legal boundaries) is the driving style (Mühl et al., 2020, p. 1335). Users could create their driving profiles by demonstrating to the AV how they drive and the AV can learn the driving style and adapt its driving style (Kraus et al., 2021, p. 1099). Accordingly, psychological mechanisms such as extraversion, introversion, neuroticism, agreeableness, self-esteem, etc. must be understood in order to build user-tailored solutions that any user can trust (Kraus et al., 2021, p. 1100).

# 5 Discussion

## 5.1 Principle Findings

The results suggest that trust in AVs can be built if both the psychological characteristics and the technological characteristics are considered and given equal weight. The psychological perspective stands out primarily for its breadth and different characteristics of trust in AVs, with the technological perspective going deep into technical details. Furthermore, the results indicate that trust in AVs is not a binary question, i.e. trusting or not trusting the AV, but that different levels of trust in AVs exist. The benefits of AVs can only be realized if they are deployed across the board, and that is only possible through acceptance, which in turn depends on trust. Even if building trust in AVs is the main focus of those involved, trust must be increased with care. Some people trust technology blindly, while others do not trust the same technology at all. Accordingly, trust must be calibrated to an appropriate level. People who do not trust AVs at all need to be educated and trained differently than people who trust AVs blindly, because both can lead to critical situations. Manufacturers, certifiers and legislators are in a tight spot to find a suitable solution. Manufacturers must build trustworthy vehicles, independent certifiers must confirm this trustworthiness, and all parties must adhere to the legal framework. This means that not only the technology is important to increase trust, but also the actions of manufacturers, certifiers and legislation in order for AVs to be accepted.

Furthermore, AVs are technology enhanced with AI and only work because of it. A closer look at the individual characteristics reveals that all the FATE characteristics of the trustworthy AI domain introduced in section 2 can also be found in the subdomain in the trust in AVs, only with slightly different wording due to the TA, although this was not explicitly researched. This finding implies that research results from the field of trustworthy AI may also have application in the sub-field of AVs.

The results further indicate that the individual characteristics should not be considered independently of each other and are not equally decisive to build trust. For example, the characteristics explainability and transparency of the psychological perspective complement each other. Explainability describes why the AV makes an action whereas transparency explains how the action will be taken. In the technological perspective, these two characteristics are always mentioned together and explainability is often a subcharacteristic of transparency. These interdependencies arise not only within (intra) each perspective but also between (inter) each perspective. Furthermore, not all characteristics are equally important for developing trust. That is a result of human nature and the variety of personalities. For example, some people come to trust in the AV after only one training session whereas some others still do not trust AVs at all. Particularly critical

are the aspects of AV-safety (reliability, accuracy, and perceived risk) for building trust due to the possibility of physical harm. Other characteristics such as public relations and communication may not generate any trust in certain situations.

## 5.2 Intra-Perspective Interdependencies

Overall, it can be stated that the consideration of interdependencies from different perspectives is critical and frequently complements each other. However, there are certain dependencies between almost every theme and characteristic, which is why only the dependencies between the most impactful are discussed below.

**Psychological Perspective**

The results suggest that there are interdependencies between the characteristics of explainability and transparency. However, explainability appears to be more valuable, since users place a larger value on the AV's explanations of actions than on making the underlying workings of these activities transparent. In other words, transparency seems to be less important than explainability, and can increase trust mostly in combination with explainability (Ekman et al. 2018, p. 100; Koo et al. 2015, p. 273), even though explainability is often mentioned as a sub-characteristic of transparency. In all cases, the context of the situation is important. Transparency seems to be only important when safety-critical situations occur (Ekman et al., 2018, p. 100). In addition, explanations of the AV must be adapted to the respective situation, so that in safety-critical situations, for example, only important information is conveyed. Hence, the type of explanation is critical and situation dependent. Otherwise might too much information decrease trust due to a cognitive overload (Filip et al. 2016, p. 5; Morra et al. 2019, p. 9443; Ha et al. 2020, p. 279). Furthermore, the characteristics of explainability and transparency are important for the introduction of independent certifications and official standards, whereby the legal framework is also decisive here. If explainability and transparency are given, and certifiers are possibly directly integrated into the development process, then audits can take place. Obviously, certifiers must have the necessary competence in order to be able to carry out this work.

Another important connection is between the characteristics explainability, transparency, and predictability. Both characteristics, explainability and transparency, help users gain a deeper understanding of the AV, which in turn increases the predictability of the AV. The more experience users gain when driving AVs, the less uncertainty users experience about the actions of AVs, which in turn can lead to increased trust in AVs. That means, when the AV is able to communicate the upcoming situation clearly, situational uncertainty can be reduced and trust can be increased (Sonoda & Wada, 2017, p. 191).

One of the most important characteristics is *perceived risk*, which captures the user's fears refers. In general, almost all characteristics can influence the perceived risk. As an example, when the AV's actions are explainable and transparent, the user can understand the AV's actions. The AV's internal education system can be used to improve the user's understanding and to better classify the perceived risks when using an AV. When the AV is reliable and the technology is accurate, has been tested, and certified appropriately according to standards, the perceived risk decreases, which can lead to an increase in trust (Li et al., 2019, p. 184).

**Technological Perspective**

An apparent link in the technological perspective is that between the method of information exchange and the other characteristics. If the AV is not able to communicate its current status, trust cannot be built due to uncertainty. Uncertainty motivates people to actively search for information to reduce uncertainty (Jayaraman et al., 2019, p. 4). Explanations, transparency, and errors of action can reduce the feeling of uncertainty when the method of information exchange is appropriate. Explanations of planned actions of the AV can best be communicated through visual channels. How actions will be performed, i.e., the transparency behind them can be communicated via voice channels. Potentially hazardous situations can be conveyed through haptic feedback, and errors can be clarified through the combination of visual and linguistic channels. One key concept to building trust in this regard is providing a feedback loop of the relevant information (Häuslschmid et al., 2017, p. 2)

By using multiple channels, the agent can be perceived as competent and aware of the situation, especially if the agent can communicate through speech (Forster et al., 2017, p. 365). Communication via language appears human and conveys the essential ability to think and feel, which increases the anthropomorphic features of the AV (Waytz et al., 2014, p. 113). On the other hand, an agent that has a lot of anthropomorphic properties is considered competent, so the efforts to give the AV a mind seem to be worthwhile in order to build trust (Waytz et al., 2014, p. 116). Hence, the competence of an agent is linked to how anthropomorphic the agent seems to be as well as how well the agent can convey information through different channels of information exchange.

How anthropomorphic an agent is seen can be controlled in part by adaptive technology and personalization as well as customization, which shows the dependency of these connections. Through customizable interfaces and methods of information exchange, users are able to create avatars that match their imaginations and convey competence. For example, some users link the attractiveness and aesthetics of avatars to competence (Häuslschmid et al., 2017, p. 8). This also includes the ability to choose individual methods of how the AV explains actions, and how the information is conveyed so that trust can be built on a user-by-user basis. Expressing information human-like and as user-tailored solu-

tions might increase trust in AVs due to individual cognitive affirmation processes (Niu et al., 2018, p. 2), because the style in which information is presented is highly individual (Stockert et al., 2015, p. 2895). In addition to a user-tailored method of information exchange, the information which is presented must also be customizable, i.e., distance driven, current and planned maneuvers, current speed, and the speed limit to increase trust in AVs (Beggiato, Hartwich, et al., 2015, p. 5).

## 5.3 Inter-Perspective Interdependencies

The legal framework and all the other characteristics that it affects represent perhaps the most significant interdependence. In Germany, the first legal frameworks have already been developed, and German car manufacturers have already recognized them as a basis for the acceptance of AVs (Mercedes-Benz Group 2017; Audi AG 2022). The legal framework in combination with official standards could, for example, regulate which data sets are used for training so that population groups are not discriminated against due to different algorithms; which data safety measures must be taken; or technical security measures must be implemented in AVs. All parties involved are forced to act within a certain framework so that users of AVs can be sure of their own safety due to a legal framework so that AVs can be trusted. This trust could be built on the reduced perceived risk as the legal regulations provide certainty, especially about technical obligations that have to be implemented. The function of checking whether the technology has been implemented following standards and the law should be left to independent certifiers in the form of audits. Most users will not have in-depth knowledge of AVs, so independent third parties are critical. As a result of the regulations that must be created, legal frameworks serve as the foundation for numerous interdependencies. For example, it might be argued that drivers would be protected instead of vulnerable road users if regulations did not intervene in the development processes. This, for example, raises ethical questions.

Another interdependency exists between public relations and communication of the technological perspective and the characteristic mental model of the technological perspective. Building an accurate mental model is important for users' cognitive processing of the capability of AVs (Large et al., 2019, p. 54). It is therefore important that car manufacturers in particular openly seek exchange with society so that no distorted mental models are created in people's minds and AVs are therefore not accepted. This communication includes clear company policies, such as the internal handling of collected data or ethical issues, as well as attempts to explain the technology of the AVs individually. In the near future, the exchange between car manufacturers and the public will have to be much closer because, at present, campaigns tend to look more like advertisements without any attempt at explanation. This may lead to an interest in the technology but at the same

time to skepticism due to a lack of knowledge about AVs and therefore to low trust.

In conclusion, the most important interdependencies are those that have a direct impact on possible safety-related issues when running an AV. These include: perceived risk; cybersecurity; reliability and accuracy; technical capability; predictability; situational awareness; as well as questions of liability in general. This is due to the fact that the physical and financial consequences are at the forefront of these characteristics, as they must be addressed in all cases prior to the large-scale deployment of AVs. Otherwise, trust in AVs will most likely never be built and an appropriate level of acceptance will most likely never be achieved in order to reap the benefits of AVs.

## 5.4 Research and Practical Implications

The results can be used to consider the most important aspects of the two perspectives on trust in AVs not in isolation but to combine both the human aspects and the technological aspects in future studies. The findings suggest that building rapport requires more than just AV technology. Instead, greater focus should be placed on human aspects, such as their behavior and concerns with regard to AVs. This is because an isolated view may distort results in situations of mutual characteristic interdependencies. The proposed trust-building characteristics in this work can be used by the research community to build a framework of trust in AVs by further confirming or refuting the findings in future studies from interdisciplinary perspectives.

From a practical perspective, the results can be used by manufacturers to build trust in AVs by considering not only technological aspects but also aspects such as early interactions with their potential customers so that appropriate mental models can be built. The importance of cooperation between certifiers and manufacturers should not be neglected either, as certificates can have a strong, trustworthy external effect. Certificates can be used to prove which technological aspects are integrated into the AV, which weaknesses it still has, as well as enable the comparability of different AVs from different manufacturers.

The results also highlight the fact that no one characteristic can ever be used to continually increase trust. Some factors, like the AV's safety, require greater consideration, but others, like personalization or customization, are most likely less important. Manufacturers need to understand the importance and dynamics of trust for AV adoption. Trust builds slowly and can be destroyed quickly if they do not show sensitivity. The trust of potential customers moves on an axis between overtrust and undertrust and has to be appropriately balanced to build up trust accordingly. If manufacturers fail to recognize the importance of trust, AVs will fail to obtain the necessary threshold of acceptance, resulting in commercial failure and the loss of all potential benefits.

## 5.5 Limitations and Future Research

We are aware of the problems that might occur, which include a potential bias in the search string or a biased analysis of the articles. Conducting a TA is a highly individual activity in which the interpretation abilities of the researcher are important. In order to decrease the possibility of biased results, we worked according to established methods with the clear goal of replicability and transparency. Furthermore, since this thesis is based on already existing literature, the synthesized knowledge is a snapshot of currently available results from studies. This means that the information published by the manufacturers themselves has not been taken into account in this work. However, some studies have been conducted in collaboration between researchers and manufacturers, covering the technological perspective. Another point to consider is that the users of AVs are not the only ones whose trust is needed, even if this work is reported from their perspective. Future research must also consider vulnerable road users in order to obtain generalizable results.

The results suggest that the psychological perspective, as well as the technological perspective, cannot and must not be strictly separated. Trust must not be seen as binary and must be built up appropriately. To this end, future research must focus more on the importance of trust calibration in AVs so that trust is neither too strong nor too low. Future research could add trust to the TAM for driverless technology(Koul & Eydgahi, 2018, p. 38), as trust is widely recognized as a core component of vehicle acceptance. This means that further models need to be developed that take into account various influencing factors, such as how trust in AVs can be measured appropriately as present trust measuring models may not be totally applicable to AVs.

The primary limitation of the reviewed literature is AV technology. The majority of the technological perspective's results were obtained through multiple simulations that could potentially not accurately represent driving with an AV on real roads. It can be argued that simulations cannot entirely reflect human behavior and human thinking. Whatever happens in the simulations, the feeling of physical integrity may play a role in the subconscious, and the outcomes of the experiments may not be totally transferable to reality. The feeling that dangerous scenarios could arise at any time and that this uncertainty cannot be resolved by one's own actions can lead to a variety of outcomes when driving in the real world. Furthermore, so far, only scientific results have been examined, leaving the insights of the manufacturers themselves unexamined. Future research should not leave out the manufacturer side and take into account the views of the industry, for example through expert interviews.

Based on this, the focus of future research should be more on-field testing, using cutting-edge technology to validate the results of simulative methodologies. Because previous

trust research has focused on technological aspects, future research in this domain must not overlook other critical influencing mechanisms besides the technology of AVs itself. These include, in particular, the significance of involving independent certifiers as well as a basic legislative structure that establishes an international foundation for AVs. This also means that demographic and country specifics are taken into account. This also implies that demographic and country-specific factors should be considered. It may be argued that Europeans react differently to technical improvements than Asians or North Americans. Asian communities, for example, appear to be more tech-savvy than other populations, which may lead to a high level of initial trust. This raises the question of how trust in other technologies may be transferred to AVs.

# 6 Conclusion

Although fully automated AVs may still be in the distant future, the fascination of this technology endures. However, the dream of AVs raises some questions that need to be addressed in order to make the benefits of AVs socially viable. The question of whether society and individuals can trust AVs is one of these concerns. One of the most critical factors in AVs' becoming socially accepted and used is trust. As a result, scientists and industry have spent a long time figuring out how to appropriately increase trust in an AV.

We believe that increasing trust in AVs is not a binary issue (whether people trust AVs or not), and that trust must be built gradually and highly individual. Trust in AVs has so far been investigated from two different perspectives: psychological, where participants could freely voice their opinions on AVs to research the behaviors; and technological, where specific implementations are provided and the behavior of the users of these implementations is analyzed. By performing a literature search that identified 55 relevant scientific papers, the findings of this study synthesize the two viewpoints into a complete perspective. A TA was carried out based on these 55 scientific papers, and the outcomes offer characteristics that foster trust from the relevant perspectives.

# A Appendix

## A.1 List of Reviewed Literature

| Author and Year | Title | View |
|---|---|---|
| Buckley et al. (2018b) | A qualitative examination of drivers' responses to partially automated vehicles | Psychological |
| J. Lee et al. (2016) | A Question of Trust: An Ethnographic Study of Automated Cars on Real Roads | Technological |
| Sheng et al. (2019) | A Case Study of Trust on Autonomous Driving | Technological |
| Niu et al. (2018) | Anthropomorphizing information to enhance trust in autonomous vehicles | Technological |
| Hengstler et al. (2016) | Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices | Psychological |
| Koester & Salge (2020) | Building Trust in Intelligent Automation: Insights into Structural Assurance Mechanisms for Autonomous Vehicles | Psychological |
| Cioroaica et al. (2020) | Building trust in the untrustable | Psychological |
| Morra et al. (2019) | Building Trust in Autonomous Vehicles: Role of Virtual Reality Driving Simulators in HMI Design | Technological |

| | | |
|---|---|---|
| Khastgir et al. (2018) | Calibrating trust through knowledge: Introducing the concept of informed safety for automation in vehicles | Technological |
| Ekman et al. (2018) | Creating Appropriate Trust in Automated Vehicle Systems: A Framework for HMI Design | Technological |
| Filip et al. (2016) | Designing and calibrating trust through situational awareness of the vehicle (SAV) feedback | Psychological |
| Khan et al. (2021) | Digital Labels: Influencing Consumers Trust and Raising Cybersecurity Awareness for Adopting Autonomous Vehicles | Technological |
| Sonoda & Wada (2017) | Displaying System Situation Awareness Increases Driver Trust in Automated Driving | Technological |
| Abe et al. (2018) | Driver Trust in Automated Driving Systems: The Case of Overtaking and Passing | Technological |
| Ha et al. (2020) | Effects of explanation types and perceived risk on trust in autonomous vehicles | Psychological |
| Ruijten et al. (2018) | Enhancing Trust in Autonomous Vehicles through Intelligent User Interfaces That Mimic Human Behavior | Technological |

| | | |
|---|---|---|
| Q. Zhang et al. (2020) | Expectations and Trust in Automated Vehicles | Psychological |
| Wintersberger et al. (2020) | Explainable Automation: Personalized and Adaptive UIs to Foster Trust and Understanding of Driving Automation Systems | Psychological |
| Ekman et al. (2019)<br><br>Technological | Exploring automated vehicle driving styles as a source of trust information | Technological |
| Sun et al. (2020) | Exploring Personalised Autonomous Vehicles to Influence User Trust | Technological |
| J. D. Lee & Kolodge (2020) | Exploring Trust in Self-Driving Vehicles Through Text Analysis | Psychological |
| Y. Ma et al. (2020) | Factors affecting trust in the autonomous vehicle: A survey of primary school students and parent perceptions | Psychological |
| Wintersberger et al. (2019) | Fostering User Acceptance and Trust in Fully Automated Vehicles: Evaluating the Potential of Augmented Reality | Technological |
| Mühl et al. (2020) | Get Ready for Being Chauffeured: Passenger's Preferences and Trust While Being Driven by Human and Automation | Technological |

| | | |
|---|---|---|
| Azevedo-Sa et al. (2021) | How internal and external risks affect the relationships between trust and driver behavior in automated driving systems | Technological |
| Shahrdar et al. (2019) | Human Trust Measurement Using an Immersive Virtual Reality Autonomous Vehicle Simulator | Technological |
| Filip et al. (2017) | Human factors considerations for cooperative positioning using positioning, navigational and sensor feedback to calibrate trust in CAVs | Psychological |
| Sun & Zhang (2021) | Improvement of Autonomous Vehicles Trust Through Synesthetic-Based Multimodal Interaction | Technological |
| von Sawitzky et al. (2019) | Increasing Trust in Fully Automated Driving: Route Indication on an Augmented Reality Head-up Display | Technological |
| Forster et al. (2017) | Increasing anthropomorphism and trust in automated driving functions by adding speech output | Technological |
| Körber et al. (2018) | Introduction matters: Manipulating trust in automation and reliance in automated driving | Technological |

| | | |
|---|---|---|
| Choi & Ji (2015) | Investigating the Importance of Trust on Adopting an Autonomous Vehicle | Psychological |
| R. H. Ma et al. (2021) | Investigating what level of visual information inspires trust in a user of a highly automated vehicle | Technological |
| Rovira et al. (2019) | Looking for Age Differences in Self-Driving Vehicles: Examining the Effects of Automation Reliability, Driving Risk, and Physical Impairment on Trust | Psychological |
| Li et al. (2019) | No Risk No Trust: Investigating Perceived Risk in Highly Automated Driving | Technological |
| Jayaraman et al. (2019) | Pedestrian Trust in Automated Vehicles: Role of Traffic Signal and AV Driving Behavior | Technological |
| Helldin et al. (2013) | Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving | Technological |
| Ajenaghughrure et al. (2020) | Risk and Trust in artificial intelligence technologies:A case study of Autonomous Vehicles | Technological |
| Petersen et al. (2019) | Situational Awareness, Driver's Trust in Automated Driving Systems and Secondary Task Performance | Psychological |

| | | |
|---|---|---|
| Seet et al. (2020) | Subtype Divergences of Trust in Autonomous Vehicles: Towards Optimisation of Driver–Vehicle Trust Management | Technological |
| Häuslschmid et al. (2017) | Supporting Trust in Autonomous Driving | Technological |
| Oliveira et al. (2020) | The influence of system transparency on trust: Evaluating interfaces in a highly automated vehicle | Technological |
| Waytz et al. (2014) | The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle | Technological |
| Large et al. (2019) | To Please in a Pod: Employing an Anthropomorphic Agent-Interlocutor to Enhance Trust and User Experience in an Autonomous, Self-Driving Vehicle | Technological |
| Ekman et al. (2016) | To See or Not to See: The Effect of Object Recognition on Users' Trust in "Automated Vehicles" | Technological |
| Luo et al. (2020) | Trust Dynamics in Human-AV(Automated Vehicle) Interaction | Technological |
| Gold et al. (2015) | Trust in Automation – Before and After the Experience of Take-over Scenarios in a Highly Automated Vehicle | Psychological |

| | | |
|---|---|---|
| Yokoi & Nakayachi (2021) | Trust in Autonomous Cars: Exploring the Role of Shared Moral Values, Reasoning, and Emotion in Safety-Critical Decisions | Psychological |
| Celmer et al. (2018) | Trust in Branded Autonomous Vehicles & Performance Expectations: A Theoretical Framework | Psychological |
| Kaur & Rampersad (2018) | Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars | Psychological |
| Ekman et al. (2020) | Trust in what? Exploring the interdependency between an automated vehicle's driving style and traffic situations | Technological |
| Dikmen & Burns (2017) | Trust in Autonomous Vehicles The Case of Tesla Autopilot and Summon | Psychological |
| Verberne et al. (2015) | Trusting a Virtual Driver That Looks, Acts, and Thinks Like You | Technological |
| Kraus et al. (2021) | What's Driving Me? Exploration and Validation of a Hierarchical Personality Model for Trust in Automated Driving | Technological |
| Omeiza et al. (2021) | Why Not Explain? Effects of Explanations on Human Perceptions of Autonomous Driving | Technological |

***Table 4:*** *List of Reviewed Literature*

## A.2 Overview of Themes

***Table 5:*** *Overview of Characteristics and Keywords of the Theme Information Exchange*

| **Theme** | **Characteristic** | **Keywords** |
|---|---|---|
| Information Exchange | Method of Information Exchange | Visual methods<br>Haptic feedback<br>Virtual reality<br>Audio feedback |
| | Error Information | Immediate justification<br>Feedback system<br>Disparity of situations<br>Malfunctions<br>Software errors<br>Algorithmic transparency<br>Certifications |
| | Explainability&Transparency of Actions | Sensing surroundings<br>Reason for actions<br>Plan for actions<br>Post-hoc timing<br>How communication<br>Conversational<br>Visual<br>Perceived Risk |
| | Situational Awareness | Sensing surroundings<br>Spatial information<br>Vehicle communication<br>Super-human driving<br>Infrastructure connection<br>Poor visibility |

***Table 6:*** *Overview of Characteristics and Keywords of the Theme User Perception and Comprehension*

| **Theme** | **Characteristic** | **Keywords** |
|---|---|---|
| User Perception and Comprehension | Mental model | Mental representation<br>Haptic feedback<br>Capability thoughts<br>Functionality thoughts<br>Discrepancy<br>Trust calibration<br>Limitations<br>Expectation |
| | Training | Capability thoughts<br>Functionality thoughts<br>Trust Calibration<br>Knowledge<br>Hands-on Experience<br>introduction |

***Table 7:*** *Overview of Characteristics and Keywords of the Theme Perceived Intelligence*

| **Theme** | **Characteristic** | **Keywords** |
|---|---|---|
| Perceived Intelligence | Anthropomorphism | Human-like |
| | | Has voice |
| | | Has name |
| | | Has gender |
| | | Has appearance |
| | | Interface |
| | | Rational thoughts |
| | Agent Competence | Performance |
| | | Purpose |
| | | Expectation |
| | | Behaves as expected |
| | | Drives to speed |
| | | Polite speech |
| | Adaptive Automation | Situational adaptivity |
| | | Driving style |
| | | Early indications |
| | | Gentle braking |
| | | Driving goals |
| | | Safety |
| | Customization and Personalization | Individuality |
| | | Data driven |
| | | Learning curve |
| | | Cognitive overload |
| | | Non critical systems |
| | | Personality |

| **Theme** | **Characteristic** | **Keywords** |
|---|---|---|
| Trust in the Technology | Explainability | Decision making |
| | | Uncertainty |
| | | Why action taken |
| | | understanding |
| | | simple explanations |
| | | deep explanations |
| | | precise |
| | | context dependent |
| | Transparency | Inner workings |
| | | Algorithmic transparency |
| | | AI black box |
| | | Certifications |
| | | Development process |
| | Usability | Good design |
| | | Intuitive |
| | | Experience |
| | | Body impairments |
| | | Accessibility |
| | Perceived Risk & Safety | Fears |
| | | Physical Harm |
| | | Uncertainty |
| | | Predictability |
| | | Safety measures |
| | | Negative vents |
| | Cybersecurity&Data Privacy | Vulnerability |
| | | Hacking |
| | | Privacy |
| | | Data streams |
| | | Manipulation |
| | | Robustness |
| | | Standards |
| | | Anonymizer |
| | | encryption |
| | | Human factors |

| | | |
|---|---|---|
| Trust in the Technology | Reliability & Accuracy | Consistency<br>Low failure rate<br>Predictability<br>Redundancy<br>Representative datasets<br>Country differences |
| | Technical Capability | Features<br>Problem solving<br>Generalization<br>Communication<br>Object detection<br>Safety |
| | Trialability | Demonstration<br>Experience<br>Mental model<br>understanding |
| | Predictability | moment to moment<br>Generalization<br>Consistency<br>Behavior<br>Determinism |

***Table 8:*** *Overview of Characteristics and Keywords of the Theme Trust in the Technology*

***Table 9:*** *Overview of Characteristics and Keywords of the Theme Trust in the Legislative*

| **Theme** | **Characteristic** | **Keywords** |
|---|---|---|
| Trust in the Legislative | Legal Framework | Standards<br>Accountability<br>Liability<br>Human-on -the-loop<br>International standards<br>responsibility |
| | Ethical Question | Moral issues<br>Algorithmic transparency<br>Discrimination<br>Prioritization |
| | Independent Certifications | Control mechanism<br>Comparability<br>Implementation<br>Algorithmic transparency<br>Certifications<br>Audits |

***Table 10:*** *Overview of Characteristics and Keywords of the Theme Trust in the Manufacturer*

| **Theme** | **Characteristic** | **Keywords** |
|---|---|---|
| Trust in the Manufacturer | Experience % Expertise | Benchmark |
| | | Newcomers |
| | | Technology on market |
| | | Predecessor technology |
| | Public Relations&Communication | Manufacturer awareness |
| | | Branding |
| | | Performance |
| | | Reliability |
| | | Expectation |
| | | Wording |
| | | Competence |

# References

Abe, G., Sato, K., & Itoh, M. (2018). Driver trust in automated driving systems: The case of overtaking and passing. *IEEE Transactions on Human-Machine Systems*, *48*(1), 85-94. doi: 10.1109/THMS.2017.2781619

Adnan, N., Md Nordin, S., bin Bahruddin, M. A., & Ali, M. (2018). How trust can drive forward the user acceptance to the technology? in-vehicle technology for autonomous vehicle. *Transportation Research Part A: Policy and Practice*, *118*, 819-836. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0965856418309030` doi: https://doi.org/10.1016/j.tra.2018.10.019

Ajenaghughrure, I. B., da Costa Sousa, S. C., & Lamas, D. (2020). Risk and trust in artificial intelligence technologies: A case study of autonomous vehicles. In *2020 13th international conference on human system interaction (hsi)* (p. 118-123). doi: 10.1109/HSI49210.2020.9142686

Audi AG. (2022). *Sleeping while driving into the weekend? not yet...* Retrieved 18/06/2022, from `https://www.audi.com/en/innovation/autonomous-driving/current-legal-situation-autonomous-driving.html`

Azevedo-Sa, H., Zhao, H., Esterwood, C., Yang, X. J., Tilbury, D. M., & Robert, L. P. (2021). How internal and external risks affect the relationships between trust and driver behavior in automated driving systems. *Transportation Research Part C: Emerging Technologies*, *123*, 102973. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0968090X21000103` doi: https://doi.org/10.1016/j.trc.2021.102973

Bansal, P., Kockelman, K. M., & Singh, A. (2016). Assessing public opinions of and interest in new vehicle technologies: An austin perspective. *Transportation Research Part C: Emerging Technologies*, *67*, 1-14. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0968090X16000383` doi: https://doi.org/10.1016/j.trc.2016.01.019

Beggiato, M., Hartwich, F., Schleinitz, K., Krems, J., Othersen, I., & Petermann-Stock, I. (2015, 11). What would drivers like to know during automated driving? information needs at different levels of automation.. doi: 10.13140/RG.2.1.2462.6007

Beggiato, M., Pereira, M., Petzoldt, T., & Krems, J. (2015). Learning and development of trust, acceptance and the mental model of acc. a longitudinal on-road study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *35*, 75-84. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1369847815001564` doi: https://doi.org/10.1016/j.trf.2015.10.005

Belanche, D., Casaló, L. V., & Flavián, C. (2012). Integrating trust and personal values into the technology acceptance model: The case of e-government services adoption. *Cuadernos de Economía y Dirección de la Empresa*, *15*(4), 192–204.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. Retrieved from `https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa` doi: 10.1191/1478088706qp063oa

Bruckes, M., Grotenhermen, J.-G., Cramer, F., & Schewe, G. (2019). Paving the way for the adoption of autonomous driving: Institution-based trust as a critical success factor. Retrieved from `https://aisel.aisnet.org/ecis2019_rp/87/`

Buckley, L., Kaye, S.-A., & Pradhan, A. K. (2018a). Psychosocial factors associated with intended use of automated vehicles: A simulated driving study. *Accident Analysis & Prevention*, *115*, 202-208. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0001457518301258` doi: https://doi.org/10.1016/j.aap.2018.03.021

Buckley, L., Kaye, S.-A., & Pradhan, A. K. (2018b). A qualitative examination of drivers' responses to partially automated vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, *56*, 167-175. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1369847817305168` doi: https://doi.org/10.1016/j.trf.2018.04.012

California Legislative Counsel. (2012). *Ssb-1298 vehicles: autonomous vehicles: safety and performance requirements.* Retrieved 29-12-2021, from `https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201120120SB1298`

Carlson, M. S., Drury, J. L., Desai, M., Kwak, H., & Yanco, H. A. (2014). Identifying factors that influence trust in automated cars and medical diagnosis systems. In *2014 aaai spring symposium series.*

Celmer, N., Branaghan, R., & Chiou, E. (2018). Trust in branded autonomous vehicles & performance expectations: A theoretical framework. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 1761-1765. Retrieved from `https://doi.org/10.1177/1541931218621398` doi: 10.1177/1541931218621398

Choi, J. K., & Ji, Y. G. (2015). Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human–Computer Interaction*, *31*(10), 692-702. Retrieved from `https://doi.org/10.1080/10447318.2015.1070549` doi: 10.1080/10447318.2015.1070549

Cioroaica, E., Buhnova, B., Kuhn, T., & Schneider, D. (2020). Building trust in the untrustable. In *Proceedings of the acm/ieee 42nd international conference on software engineering: Software engineering in society* (p. 21–24). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3377815.3381378` doi: 10.1145/3377815.3381378

Davis, F. D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International journal of man-machine studies*, *38*(3), 475–487.

De Angelis, M., Puchades, V. M., Fraboni, F., Pietrantoni, L., & Prati, G. (2017). Negative attitudes towards cyclists influence the acceptance of an in-vehicle cyclist detection system. *Transportation Research Part F: Traffic Psychology and Behaviour*, *49*, 244-256. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1369847816302431` doi: https://doi.org/10.1016/j.trf.2017.06.021

Dikmen, M., & Burns, C. (2017). Trust in autonomous vehicles: The case of tesla autopilot and summon. In *2017 ieee international conference on systems, man, and cybernetics (smc)* (p. 1093-1098). doi: 10.1109/SMC.2017.8122757

Du, N., Haspiel, J., Zhang, Q., Tilbury, D., Pradhan, A. K., Yang, X. J., & Robert, L. P. (2019). Look who's talking now: Implications of av's explanations on driver's trust, av preference, anxiety and mental workload. *Transportation Research Part C: Emerging Technologies*, *104*, 428-442. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0968090X18313640` doi: https://doi.org/10.1016/j.trc.2019.05.025

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, *58*(6), 697-718. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1071581903000387` (Trust and Technology) doi: https://doi.org/10.1016/S1071-5819(03)00038-7

Ekman, F., Johansson, M., Bligård, L.-O., Karlsson, M., & Strömberg, H. (2019). Exploring automated vehicle driving styles as a source of trust information. *Transportation Research Part F: Traffic Psychology and Behaviour*, *65*, 268-279. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1369847818306594` doi: https://doi.org/10.1016/j.trf.2019.07.026

Ekman, F., Johansson, M., Karlsson, M., Bligård, L.-O., & Strömberg, H. (2020, 12). Trust in what? exploring the interdependency between an automated vehicle's driving style and traffic situations. *Transportation Research Part F Traffic Psychology and Behaviour*, *76*, 59-71. doi: 10.1016/j.trf.2020.10.012

Ekman, F., Johansson, M., & Sochor, J. (2016). To see or not to see: The effect of object recognition on users' trust in "automated vehicles". In *Proceedings of the 9th nordic conference on human-computer interaction.* New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/2971485.2971551` doi: 10.1145/2971485.2971551

Ekman, F., Johansson, M., & Sochor, J. (2018). Creating appropriate trust in automated vehicle systems: A framework for hmi design. *IEEE Transactions on Human-Machine Systems*, *48*(1), 95-101. doi: 10.1109/THMS.2017.2776209

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological review*, *114*(4), 864. doi: https://doi.org/10.1037/0033-295X.114.4.864

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

Filip, G., Meng, X., Burnett, G., & Harvey, C. (2016). Designing and calibrating trust through situational awareness of the vehicle (sav) feedback. In *Iet international conference on intelligent and connected vehicles (icv 2016)* (p. 1-6). doi: 10.1049/cp.2016.1171

Filip, G., Meng, X., Burnett, G., & Harvey, C. (2017). Human factors considerations for cooperative positioning using positioning, navigational and sensor feedback to calibrate trust in cavs. In *2017 forum on cooperative positioning and service (cpgps* (p. 134-139). doi: 10.1109/CPGPS.2017.8075111

Forster, Y., Naujoks, F., & Neukum, A. (2017). Increasing anthropomorphism and trust in automated driving functions by adding speech output. In *2017 ieee intelligent vehicles symposium (iv)* (p. 365-372). doi: 10.1109/IVS.2017.7995746

Ghazizadeh, M., Lee, J. D., & Boyle, L. N. (2012). Extending the technology acceptance model to assess automation. *Cognition, Technology & Work*, *14*(1), 39–49.

Goertzel, B. (2014, December). Artificial General Intelligence: Concept, State of the Art, and Future Prospects. *Journal of Artificial General Intelligence*, *5*(1), 1-48. doi: 10.2478/jagi-2014-0001

Gold, C., Körber, M., Hohenberger, C., Lechner, D., & Bengler, K. (2015). Trust in automation – before and after the experience of take-over scenarios in a highly automated vehicle. *Procedia Manufacturing*, *3*, 3025-3032. Retrieved from `https://`

www.sciencedirect.com/science/article/pii/S2351978915008483 (6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015) doi: https://doi.org/10.1016/j.promfg.2015.07.847

Gray, H., Gray, K., & Wegner, D. (2007, 03). Dimensions of mind perception. *Science (New York, N.Y.)*, *315*, 619. doi: 10.1126/science.1134475

Ha, T., Kim, S., Seo, D., & Lee, S. (2020). Effects of explanation types and perceived risk on trust in autonomous vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, *73*, 271-280. Retrieved from https://www.sciencedirect.com/science/article/pii/S1369847820304587 doi: https://doi.org/10.1016/j.trf.2020.06.021

Häuslschmid, R., von Bülow, M., Pfleging, B., & Butz, A. (2017). Supporting trust in autonomous driving. In *Proceedings of the 22nd international conference on intelligent user interfaces* (p. 319–329). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3025171.3025198 doi: 10.1145/3025171.3025198

Helldin, T., Falkman, G., Riveiro, M., & Davidsson, S. (2013, 10). Presenting system uncertainty in automotive uis for supporting trust calibration in autonomous driving.. doi: 10.1145/2516540.2516554

Hengstler, M., Enkel, E., & Duelli, S. (2016). Applied artificial intelligence and trust—the case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change*, *105*, 105-120. Retrieved from https://www.sciencedirect.com/science/article/pii/S0040162515004187 doi: https://doi.org/10.1016/j.techfore.2015.12.014

Hobert, L., Festag, A., Llatser, I., Altomare, L., Visintainer, F., & Kovacs, A. (2015). Enhancements of v2x communication in support of cooperative autonomous driving. *IEEE Communications Magazine*, *53*(12), 64-70. doi: 10.1109/MCOM.2015.7355568

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors*, *57*(3), 407–434.

Holland-Letz, D., Kässer, M., Kloss, B., & Müller, T. (2019). *Start me up: Where mobility investments are going.* Retrieved 22-11-2021, from https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/start-me-up-where-mobility-investments-are-going

Jayaraman, S. K., Creech, C., Tilbury, D. M., Yang, X. J., Pradhan, A. K., Tsui, K. M., & Robert, L. P. (2019). Pedestrian trust in automated vehicles: Role of traffic signal and

av driving behavior. *Frontiers in Robotics and AI*, *6*. Retrieved from https://www.frontiersin.org/article/10.3389/frobt.2019.00117 doi: 10.3389/frobt.2019.00117

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53-71. Retrieved from https://doi.org/10.1207/S15327566IJCE0401_04 doi: 10.1207/S15327566IJCE0401_04

Kaur, K., & Rampersad, G. (2018). Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars. *Journal of Engineering and Technology Management*, *48*, 87-96. Retrieved from https://www.sciencedirect.com/science/article/pii/S0923474817304253 doi: https://doi.org/10.1016/j.jengtecman.2018.04.006

Khan, W. Z., Khurram Khan, M., Arshad, Q.-u.-A., Malik, H., & Almuhtadi, J. (2021). Digital labels: Influencing consumers trust and raising cybersecurity awareness for adopting autonomous vehicles. In *2021 ieee international conference on consumer electronics (icce)* (p. 1-4). doi: 10.1109/ICCE50685.2021.9427684

Khastgir, S., Birrell, S., Dhadyalla, G., & Jennings, P. (2018). Calibrating trust through knowledge: Introducing the concept of informed safety for automation in vehicles. *Transportation Research Part C: Emerging Technologies*, *96*, 290-303. Retrieved from https://www.sciencedirect.com/science/article/pii/S0968090X18309252 doi: https://doi.org/10.1016/j.trc.2018.07.001

Kitchenham, B. A., & Charters, S. (2007, 07 09). *Guidelines for performing systematic literature reviews in software engineering* (Tech. Rep. No. EBSE 2007-001). Keele University and Durham University Joint Report. Retrieved from https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf

Koester, N., & Salge, O. (2020). Building trust in intelligent automation: Insights into structural assurance mechanisms for autonomous vehicles. In J. F. George, S. Paul, R. De', E. Karahanna, S. Sarker, & G. Oestreicher-Singer (Eds.), *Proceedings of the 41st international conference on information systems, ICIS 2020, making digital inclusive: Blending the locak and the global, hyderabad, india, december 13-16, 2020.* Association for Information Systems. Retrieved from https://aisel.aisnet.org/icis2020/iot_smart/iot_smart/7

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2015). Why did my car just do that? explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *9*(4), 269–275.

Körber, M., Prasch, L., & Bengler, K. (2018). Why do i have to drive now? post hoc explanations of takeover requests. *Human factors*, *60*(3), 305–323.

Köster, N., & Salge, T.-O. (2021). Paving the way for trust in autonomous vehicles–how oems, authorities and certifiers can foster customers' trust in avs. In T. Bertram (Ed.), *Automatisiertes fahren 2021* (pp. 29–39). Wiesbaden: Springer Fachmedien Wiesbaden.

Koul, S., & Eydgahi, A. (2018, Dec.). Utilizing technology acceptance model (tam) for driverless car technology. *Journal of Technology Management amp;amp; Innovation*, *13*(4), 37–46. Retrieved from `https://www.jotmi.org/index.php/GT/article/view/2789` doi: 10.4067/S0718-27242018000400037

Kraus, J., Scholz, D., & Baumann, M. (2021). What's driving me? exploration and validation of a hierarchical personality model for trust in automated driving. *Human Factors*, *63*(6), 1076-1105. Retrieved from `https://doi.org/10.1177/0018720820922653` (PMID: 32633564) doi: 10.1177/0018720820922653

Kyriakidis, M., Happee, R., & de Winter, J. (2015, 07). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour*, *32*. doi: 10.1016/j.trf.2015.04.014

Körber, M., Baseler, E., & Bengler, K. (2018, 01). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied Ergonomics*, *66*, 18-31. doi: 10.1016/j.apergo.2017.07.006

Large, D. R., Harrington, K., Burnett, G., Luton, J., Thomas, P., & Bennett, P. (2019). To please in a pod: Employing an anthropomorphic agent-interlocutor to enhance trust and user experience in an autonomous, self-driving vehicle. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications* (p. 49–59). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3342197.3344545` doi: 10.1145/3342197.3344545

Lee, Hsieh, & Hsu. (2011). Adding innovation diffusion theory to the technology acceptance model: Supporting employees' intentions to use e-learning systems. *Journal of Educational Technology & Society*, *14*(4), 124–137.

Lee, J., Kim, N., Imm, C., Kim, B., Yi, K., & Kim, J. (2016). A question of trust: An ethnographic study of automated cars on real roads. In *Proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications* (p. 201–208). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3003715.3005405` doi: 10.1145/3003715.3005405

Lee, J., & Lee, K. M. (2022). Polite speech strategies and their impact on drivers' trust in autonomous vehicles. *Computers in Human Behavior*, *127*, 107015. Retrieved from https://www.sciencedirect.com/science/article/pii/S0747563221003381 doi: https://doi.org/10.1016/j.chb.2021.107015

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243-1270. Retrieved from https://doi.org/10.1080/00140139208967392 (PMID: 1516577) doi: 10.1080/00140139208967392

Lee, J. D., & Kolodge, K. (2020). Exploring trust in self-driving vehicles through text analysis. *Human Factors*, *62*(2), 260-277. Retrieved from https://doi.org/10.1177/0018720819872672 (PMID: 31502885) doi: 10.1177/0018720819872672

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50-80. Retrieved from https://doi.org/10.1518/hfes.46.1.50_30392 (PMID: 15151155) doi: 10.1518/hfes.46.1.50_30392

Lee, J.-G., Lee, K. M., & Ryu, S.-H. (2019). Vehicle politeness in driving situations. *Future Internet*, *11*(2). Retrieved from https://www.mdpi.com/1999-5903/11/2/48 doi: 10.3390/fi11020048

Li, M., Holthausen, B., Stuck, R., & Walker, B. (2019, 09). No risk no trust: Investigating perceived risk in highly automated driving. In (p. 177-185). doi: 10.1145/3342197.3344525

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable ai: A review of machine learning interpretability methods. *Entropy*, *23*(1), 18.

Linkov, V., Zámečník, P., Havlíčková, D., & Pai, C.-W. (2019). Human factors in the cybersecurity of autonomous vehicles: Trends in current research. *Frontiers in psychology*, *10*, 995.

Luo, R., Chu, J., & Yang, X. J. (2020). Trust dynamics in human-av (automated vehicle) interaction. In *Extended abstracts of the 2020 chi conference on human factors in computing systems* (p. 1–7). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3334480.3382970 doi: 10.1145/3334480.3382970

Ma, R. H., Morris, A., Herriotts, P., & Birrell, S. (2021). Investigating what level of visual information inspires trust in a user of a highly automated vehicle. *Applied Ergonomics*, *90*, 103272. Retrieved from https://www.sciencedirect.com/science/article/pii/S0003687020302210 doi: https://doi.org/10.1016/j.apergo.2020.103272

Ma, Y., Li, S., Qin, S., & Qi, Y. (2020). Factors affecting trust in the autonomous vehicle: A survey of primary school students and parent perceptions. In *2020 ieee 19th international conference on trust, security and privacy in computing and communications (trustcom)* (p. 2020-2027). doi: 10.1109/TrustCom50675.2020.00277

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709–734. Retrieved from `http://www.jstor.org/stable/258792`

McCarthy, J. (2004). What is artificial intelligence. *URL: http://www-formal. stanford. edu/jmc/whatisai. html*.

Mcknight, D., & Chervany, N. (2000, 08). What is trust? a conceptual analysis and an interdisciplinary model..

Mercedes-Benz Group. (2017). *Automated and autonomous driving. legal framework.* Retrieved 18/06/2022, from `https://group.mercedes-benz.com/innovation/case/autonomous/legal-framework.html`

Merritt, S. M., Heimbaugh, H., LaChapell, J., & Lee, D. (2013). I trust it, but i don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human factors*, *55*(3), 520–534.

Mishra, A. (1996, 01). Organizational responses to crisis: The centrality of trust.

Morgan, P. L., Voinescu, A., Alford, C., & Caleb-Solly, P. (2018). Exploring the usability of a connected autonomous vehicle human machine interface designed for older adults. In *International conference on applied human factors and ergonomics* (pp. 591–603).

Morgan, P. L., Voinescu, A., Williams, C., Caleb-Solly, P., Alford, C., Shergold, I., ... Pipe, A. (2017). An emerging framework to inform effective design of human-machine interfaces for older adults using connected autonomous vehicles. In *International conference on applied human factors and ergonomics* (pp. 325–334).

Morra, L., Lamberti, F., Pratticó, F. G., Rosa, S. L., & Montuschi, P. (2019). Building trust in autonomous vehicles: Role of virtual reality driving simulators in hmi design. *IEEE Transactions on Vehicular Technology*, *68*(10), 9438-9450. doi: 10.1109/TVT.2019.2933601

Muir, B., & Moray, N. (1996, 04). Trust in automation. part ii. experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, *39*, 429-60. doi: 10.1080/00140139608964474

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International journal of man-machine studies*, *27*(5-6), 527–539.

Muir, B. M. (1994). Trust in automation: Part i. theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, *37*(11), 1905-1922. Retrieved from `https://doi.org/10.1080/00140139408964957` doi: 10.1080/00140139408964957

Mühl, K., Strauch, C., Grabmaier, C., Reithinger, S., Huckauf, A., & Baumann, M. (2020). Get ready for being chauffeured: Passenger's preferences and trust while being driven by human and automation. *Human Factors*, *62*(8), 1322-1338. Retrieved from `https://doi.org/10.1177/0018720819872893` (PMID: 31498656) doi: 10.1177/0018720819872893

Niu, D., Terken, J., & Eggen, B. (2018). Anthropomorphizing information to enhance trust in autonomous vehicles. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *28*(6), 352-359. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/hfm.20745` doi: https://doi.org/10.1002/hfm.20745

Oliveira, L., Burns, C., Luton, J., Iyer, S., & Birrell, S. (2020). The influence of system transparency on trust: Evaluating interfaces in a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour*, *72*, 280-296. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1369847820304381` doi: https://doi.org/10.1016/j.trf.2020.06.001

Omeiza, D., Kollnig, K., Web, H., Jirotka, M., & Kunze, L. (2021). Why not explain? effects of explanations on human perceptions of autonomous driving. In *2021 ieee international conference on advanced robotics and its social impacts (arso)* (p. 194-199). doi: 10.1109/ARSO51874.2021.9542835

Paden, B., Čáp, M., Yong, S. Z., Yershov, D., & Frazzoli, E. (2016). A survey of motion planning and control techniques for self-driving urban vehicles. *IEEE Transactions on Intelligent Vehicles*, *1*(1), 33-55. doi: 10.1109/TIV.2016.2578706

Pandl, K. D., Thiebes, S., Schmidt-Kraepelin, M., & Sunyaev, A. (2020). On the convergence of artificial intelligence and distributed ledger technology: A scoping review and future research agenda. *IEEE Access*, *8*, 57075-57095. doi: 10.1109/ACCESS.2020.2981447

Parkinson, S., Ward, P., Wilson, K., & Miller, J. (2017). Cyber threats facing autonomous and connected vehicles: Future challenges. *IEEE transactions on intelligent transportation systems*, *18*(11), 2898–2915.

Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International journal of electronic commerce*, *7*(3), 101–134.

Penmetsa, P., Adanu, E. K., Wood, D., Wang, T., & Jones, S. L. (2019). Perceptions and expectations of autonomous vehicles–a snapshot of vulnerable road user opinion. *Technological Forecasting and Social Change*, *143*, 9–13.

Petersen, L., Robert, L., Yang, X. J., & Tilbury, D. (2019, 03). Situational awareness, driver's trust in automated driving systems and secondary task performance. doi: 10.2139/ssrn.3345543

Pietsch, B. (2021). *2 killed in driverless tesla car crash, officials say.* Retrieved 27/03/2022, from `https://www.nytimes.com/2021/04/18/business/tesla-fatal-crash-texas.html`

Raats, K., Fors, V., & Pink, S. (2020). Trusting autonomous vehicles: An interdisciplinary approach. *Transportation Research Interdisciplinary Perspectives*, *7*, 100201. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2590198220301123` doi: https://doi.org/10.1016/j.trip.2020.100201

Raj, A., Kumar, J. A., & Bansal, P. (2020). A multicriteria decision making approach to study barriers to the adoption of autonomous vehicles. *Transportation Research Part A: Policy and Practice*, *133*, 122-137. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0965856419305798` doi: https://doi.org/10.1016/j.tra.2020.01.013

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, *49*(1), 95–112.

Rödel, C., Stadler, S., Meschtscherjakov, A., & Tscheligi, M. (2014). Towards autonomous cars: The effect of autonomy levels on acceptance and user experience. In *Proceedings of the 6th international conference on automotive user interfaces and interactive vehicular applications* (p. 1–8). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/2667317.2667330` doi: 10.1145/2667317.2667330

Rousseau, R., Tremblay, S., & Breton, R. (2004). Defining and modeling situation awareness: A critical review. *A cognitive approach to situation awareness: Theory and application*, 3–21.

Rovira, E., McLaughlin, A. C., Pak, R., & High, L. (2019). Looking for age differences in self-driving vehicles: Examining the effects of automation reliability, driving risk, and physical impairment on trust. *Frontiers in Psychology*, *10*. Retrieved from `https://www.frontiersin.org/article/10.3389/fpsyg.2019.00800` doi: 10.3389/fpsyg.2019.00800

Ruijten, P. A. M., Terken, J. M. B., & Chandramouli, S. N. (2018). Enhancing trust in autonomous vehicles through intelligent user interfaces that mimic human behavior. *Multimodal Technologies and Interaction*, *2*(4). Retrieved from `https://www.mdpi.com/2414-4088/2/4/62` doi: 10.3390/mti2040062

SAE International. (2021). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles.* Retrieved 18-11-2021, from `https://www.sae.org/standards/content/j3016_202104/`

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, pp. 2296–2300).

Saleh, K., Hossny, M., & Nahavandi, S. (2017). Towards trusted autonomous vehicles from vulnerable road users perspective. In *2017 annual ieee international systems conference (syscon)* (p. 1-7). doi: 10.1109/SYSCON.2017.7934782

Schoettle, B., & Sivak, M. (2014, 11). A survey of public opinion about connected vehicles in the u.s., the u.k., and australia. In (p. 687-692). doi: 10.1109/ICCVE.2014.7297637

Schoettle, B., & Sivak, M. (2015). *Motorists' preferences for different levels of vehicle automation* (Tech. Rep.). University of Michigan, Ann Arbor, Transportation Research Institute.

Seet, M. S., Dragomir, A., Mathialagan, I., Ann, L. Y., Zaid, Z. b., Ramapatna, S. L., ... Bezerianos, A. (2020). Subtype divergences of trust in autonomous vehicles: Towards optimisation of driver–vehicle trust management. In *2020 ieee 23rd international conference on intelligent transportation systems (itsc)* (p. 1-6). doi: 10.1109/ITSC45102.2020.9294495

Shahrdar, S., Park, C., & Nojoumian, M. (2019). Human trust measurement using an immersive virtual reality autonomous vehicle simulator. In *Proceedings of the 2019 aaai/acm conference on ai, ethics, and society* (p. 515–520). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3306618.3314264` doi: 10.1145/3306618.3314264

Sheng, S., Pakdamanian, E., Han, K., Kim, B., Tiwari, P., Kim, I., & Feng, L. (2019). A case study of trust on autonomous driving<sup>*</sup>. In *2019 ieee intelligent transportation systems conference (itsc)* (p. 4368–4373). IEEE Press. Retrieved from `https://doi.org/10.1109/ITSC.2019.8917251` doi: 10.1109/ITSC.2019.8917251

Shin, D. (2020, 12). User perceptions of algorithmic decisions in the personalized ai system:perceptual evaluation of fairness, accountability, transparency, and explainabil-

ity. *Journal of Broadcasting Electronic Media*, *64*, 1-25. doi: 10.1080/08838151.2020.1843357

Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *Cutter business technology journal*, *31*(2), 47–53.

Singh, S. (2018). Critical reasons for crashes investigated in the national motor vehicle crash causation survey. *[Report No. DOT HS 812 506]. National Highway Traffic Safety Administration.*.

Slater, S. F., & Mohr, J. J. (2006). Successful development and commercialization of technological innovation: Insights based on strategy type. *Journal of product innovation management*, *23*(1), 26–33.

Sonoda, K., & Wada, T. (2017). Displaying system situation awareness increases driver trust in automated driving. *IEEE Transactions on Intelligent Vehicles*, *2*(3), 185-193. doi: 10.1109/TIV.2017.2749178

Stadler, S., Cornet, H., Novaes Theoto, T., & Frenkler, F. (2019). A tool, not a toy: using virtual reality to evaluate the communication between autonomous vehicles and pedestrians. In *Augmented reality and virtual reality* (pp. 203–216). Springer.

Stanton, N. A., & Young, M. S. (2000). A proposed psychological model of driving automation. *Theoretical Issues in Ergonomics Science*, *1*(4), 315-331. Retrieved from `https://doi.org/10.1080/14639220052399131` doi: 10.1080/14639220052399131

Stockert, S., Richardson, N. T., & Lienkamp, M. (2015). Driving in an increasingly automated world – approaches to improve the driver-automation interaction. *Procedia Manufacturing*, *3*, 2889-2896. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2351978915007982` (6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015) doi: https://doi.org/10.1016/j.promfg.2015.07.797

Sun, X., Li, J., Tang, P., Zhou, S., Peng, X., LI, H. N., & Wang, Q. (2020, 11). Exploring personalised autonomous vehicles to influence user trust. *Cognitive Computation*, *12*. doi: 10.1007/s12559-020-09757-x

Sun, X., & Zhang, Y. (2021). Improvement of autonomous vehicles trust through synesthetic-based multimodal interaction. *IEEE Access*, *9*, 28213-28223. doi: 10.1109/ACCESS.2021.3059071

Taeihagh, A., & Lim, H. S. M. (2019). Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transport reviews*, *39*(1), 103–128.

Taiebat, M., Brown, A. L., Safford, H. R., Qu, S., & Xu, M. (2018). A review on energy, environmental, and sustainability implications of connected and automated vehicles. *Environmental Science & Technology*, *52*(20), 11449-11465. Retrieved from https://doi.org/10.1021/acs.est.8b00127 (PMID: 30192527) doi: 10.1021/acs.est.8b00127

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, *94*(6), 1395-1415. Retrieved from https://doi.org/10.2307/796133 doi: 10.2307/796133

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, *46*(2), 186–204.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425–478.

Verberne, F., Ham, J., & Midden, C. (2012). Trust in smart systems: Sharing driving goals and giving information to increase trustworthiness and acceptability of smart systems in cars. *Human factors*, *54*, 799-810. doi: 10.1177/0018720812443825

Verberne, F., Ham, J., & Midden, C. (2015, 05). Trusting a virtual driver that looks, acts, and thinks like you. *Human Factors The Journal of the Human Factors and Ergonomics Society*, *57*, 1-15. doi: 10.1177/0018720815580749

von Sawitzky, T., Wintersberger, P., Riener, A., & Gabbard, J. L. (2019). Increasing trust in fully automated driving: Route indication on an augmented reality head-up display. In *Proceedings of the 8th acm international symposium on pervasive displays.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3321335.3324947 doi: 10.1145/3321335.3324947

Wakabayashi, D. (2018). *Self-driving uber car kills pedestrian in arizona, where robots roam.* Retrieved 27-03-22, from https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113-117. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022103114000067 doi: https://doi.org/10.1016/j.jesp.2014.01.005

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2), xiii–xxiii. Retrieved from http://www.jstor.org/stable/4132319

Wintersberger, P., Frison, A.-K., Riener, A., & Sawitzky, T. v. (2019). Fostering user acceptance and trust in fully automated vehicles: Evaluating the potential of augmented reality. *Presence*, *27*(1), 46-62. doi: 10.1162/pres_a_00320

Wintersberger, P., Nicklas, H., Martlbauer, T., Hammer, S., & Riener, A. (2020). Explainable automation: Personalized and adaptive uis to foster trust and understanding of driving automation systems. In *12th international conference on automotive user interfaces and interactive vehicular applications* (p. 252–261). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3409120.3410659` doi: 10.1145/3409120.3410659

Wintersberger, P., & Riener, A. (2016). Trust in technology as a safety aspect in highly automated driving. *i-com*, *15*(3), 297–310. Retrieved from `https://doi.org/10.1515/icom-2016-0034` doi: doi:10.1515/icom-2016-0034

Yokoi, R., & Nakayachi, K. (2021). Trust in autonomous cars: Exploring the role of shared moral values, reasoning, and emotion in safety-critical decisions. *Human Factors*, *63*(8), 1465-1484. Retrieved from `https://doi.org/10.1177/0018720820933041` (PMID: 32663047) doi: 10.1177/0018720820933041

Yuen, K. F., Wong, Y. D., Ma, F., & Wang, X. (2020). The determinants of public acceptance of autonomous vehicles: An innovation diffusion perspective. *Journal of Cleaner Production*, *270*, 121904.

Zhang, Q., Yang, X. J., & Robert, L. P. (2020). Expectations and trust in automated vehicles. In *Extended abstracts of the 2020 chi conference on human factors in computing systems* (p. 1–9). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3334480.3382986` doi: 10.1145/3334480.3382986

Zhang, T., Tao, D., Qu, X., Zhang, X., Lin, R., & Zhang, W. (2019). The roles of initial trust and perceived risk in public's acceptance of automated vehicles. *Transportation Research Part C: Emerging Technologies*, *98*, 207-220. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0968090X18308398` doi: https://doi.org/10.1016/j.trc.2018.11.018

Zhao, X., Robu, V., Flynn, D., Salako, K., & Strigini, L. (2019). Assessing the safety and reliability of autonomous vehicles from road testing. In *2019 ieee 30th international symposium on software reliability engineering (issre)* (pp. 13–23).

Zimmermann, M., & Wotawa, F. (2020, Sep 01). An adaptive system for autonomous driving. *Software Quality Journal*, *28*(3), 1189-1212. Retrieved from `https://doi.org/10.1007/s11219-020-09519-w` doi: 10.1007/s11219-020-09519-w

Zmud, J. P., & Sener, I. N. (2017). Towards an understanding of the travel behavior impact of autonomous vehicles. *Transportation Research Procedia*, *25*, 2500-2519. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2352146517305884` (World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016) doi: https://doi.org/10.1016/j.trpro.2017.05.281

# Assertion

*Ich versichere wahrheitsgemäß, die Arbeit selbstständig verfasst, alle benutzten Quellen und Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde sowie die Satzung des KIT zur Sicherung guter wissenschaftlicher Praxis in der jeweils gültigen Fassung beachtet zu haben.“ Bei Abgabe einer unwahren Versicherung wird die Masterarbeit mit „nicht ausreichend“ (5,0) bewertet.*



Brühl, July 16, 2022 Lars Benedict Martin