Meaningful human control of partially automated driving systems
Insights from interviews with Tesla users

Suryana, Lucas Elbert; Nordhoff, Sina; Calvert, Simeon; Zgonnikov, Arkady; van Arem, Bart

DOI
10.1016/j.trf.2025.04.026

Publication date
2025

Document Version
Final published version

Published in
Transportation Research Part F: Traffic Psychology and Behaviour

Transportation Research Part F: Psychology and Behaviour 113 (2025) 213–236

Contents lists available at ScienceDirect

Transportation Research Part F: Psychology and Behaviour

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

Meaningful human control of partially automated driving systems: Insights from interviews with Tesla users

Lucas Elbert Suryana [a,d,*], Sina Nordhoff [a,b], Simeon Calvert [a,d], Arkady Zgonnikov [c,d], Bart van Arem [a]

[a] Department of Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, the Netherlands
[b] Electric Vehicle Research Center, Institute of Transportation Studies, University of California, Davis, United States
[c] Department of Cognitive Robotics, Faculty of Mechanical Engineering, Delft University of Technology, Delft, the Netherlands
[d] Centre for Meaningful Human Control, Delft University of Technology, Delft, the Netherlands

ARTICLE INFO

Keywords:
Partially automated driving systems
Meaningful human control
Semi-structured interviews
Empirical evaluation

ABSTRACT

Partially automated driving systems are designed to perform specific driving tasks—such as steering, accelerating, and braking—while still requiring human drivers to monitor the environment and intervene when necessary. This shift of driving responsibilities from human drivers to automated systems raises concerns about accountability, particularly in scenarios involving unexpected events. To address these concerns, the concept of meaningful human control (MHC) has been proposed. MHC emphasises the importance of humans retaining oversight and responsibility for decisions made by automated systems. Despite extensive theoretical discussion of MHC in driving automation, there is limited empirical research on how real-world partially automated systems align with MHC principles. This study offers two main contributions: (1) an empirical evaluation of MHC in partially automated driving, based on 103 semi-structured interviews with users of Tesla's Autopilot and Full Self-Driving (FSD) Beta systems; and (2) a methodological framework for assessing MHC through qualitative interview data. We operationalise the previously proposed tracking and tracing conditions of MHC using a set of evaluation criteria to determine whether these systems support meaningful human control in practice. Our findings indicate that several factors influence the degree to which MHC is achieved. Failures in tracking—where drivers' expectations regarding system safety are not adequately met—arise from technological limitations, susceptibility to environmental conditions (e.g., adverse weather or inadequate infrastructure), and discrepancies between technical performance and user satisfaction. Tracing performance—the ability to clearly assign responsibility—is affected by inconsistent adherence to safety protocols, varying levels of driver confidence, and the specific driving mode in use (e.g., Autopilot versus FSD Beta). These findings contribute to ongoing efforts to design partially automated driving systems that more effectively support meaningful human control and promote more appropriate use of automation.

* Corresponding author at: Department of Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, the Netherlands.

E-mail address: l.e.suryana@tudelft.nl (L.E. Suryana).

https://doi.org/10.1016/j.trf.2025.04.026
Received 14 November 2024; Received in revised form 28 April 2025; Accepted 28 April 2025

Available online 9 May 2025

1. Introduction

Partially automated driving systems—classified as SAE Level 2 automation—are designed to assist drivers with specific tasks such as steering, accelerating, and braking, while still requiring human drivers to maintain vigilance and be prepared to take control when necessary (SAE International, 2021). The deployment of these systems raises critical questions about the allocation of driver responsibility, especially in unexpected situations. Recent fatal collisions involving Tesla's Autopilot and Ford's BlueCruise have brought these concerns to prominence, particularly in instances where neither the driver nor the system adequately responded to visible obstacles (National Transportation Safety Board, 2017, 2018; Robins-Early, 2024). Currently, manufacturers such as Tesla explicitly assign oversight responsibility to the driver, as clearly stated in official safety documentation, which instructs users to remain attentive and ready to intervene when required (Tesla, 2024). Consequently, a driver's failure to take timely corrective action may render them liable in the event of a collision.

However, supervising partially automated driving systems presents significant challenges for human drivers (Martinho et al., 2021). Empirical studies involving Tesla Autopilot users have shown that prolonged exposure to reliably performing Level 2 automation often results in "passenger-like viewing behaviours," including extreme cases such as drivers sleeping at the wheel (Nordhoff et al., 2023). These behaviours illustrate the risks associated with overreliance on automation, leading to reduced attention and increased distraction. This phenomenon echoes Bainbridge's seminal analysis of the "ironies of automation," which demonstrated how automation can undermine operator engagement, foster overdependence, and degrade manual driving skills over time Bainbridge (1983). Supporting this perspective, Banks et al. (2018b) found that drivers responsible for monitoring partially automated systems frequently become complacent, raising concerns about their capacity to intervene effectively. Banks further argued that attributing fault to drivers for failures arising from the design and implementation of Level 2 and Level 3 systems is ethically questionable. Moreover, research indicates that drivers of partially automated vehicles are often held disproportionately accountable for collisions, even in situations where system limitations significantly constrain their ability to respond (Li et al., 2016; Awad et al., 2020; Beckers et al., 2022). These findings underscore ongoing concerns regarding the fair distribution of responsibility in the context of partially automated driving.

1.1. Meaningful human control

Delegating control to automated systems—those capable of executing tasks with varying degrees of autonomy, ranging from partial to full automation—may give rise to responsibility gaps, in which it becomes unclear which human agent should be held accountable for the outcomes of the system's actions (Matthias, 2004; Santoni de Sio & Mecacci, 2021). To address this challenge, the concept of meaningful human control (MHC) has gained increasing prominence in scholarly debates on responsibility attribution within automated contexts (Santoni de Sio & Van den Hoven, 2018). Originally proposed in relation to autonomous weapon systems (Docherty, 2015), MHC emphasises the principle that humans must retain some degree of control over automated decisions to remain morally and legally accountable for the system's behaviour (Santoni de Sio & Van den Hoven, 2018).

Although MHC was initially formulated in the context of fully automated systems—those functioning without human intervention, such as SAE Level 5 vehicles—it has since been expanded to encompass a broader range of automated technologies, including systems that still require human supervision, such as partially automated driving systems (Mecacci & Santoni de Sio, 2020). This wider applicability is particularly relevant in road transport, where system deployment must consider not only the vehicle's operational capabilities but also the complex nature of the transport ecosystem, including interactions with human drivers, pedestrians, cyclists, infrastructure, and the inherently unpredictable dynamics of traffic and weather.

To enhance understanding of how MHC applies across different levels of vehicle automation, the CCAM Taxonomy provides a useful conceptual framework (Connected Automated Driving, 2024). According to this classification, SAE Level 5 vehicles are fully autonomous and operate without any human intervention. In contrast, SAE Levels 1 to 4 represent varying degrees of automation, each requiring some level of human oversight. For instance, Level 1 systems incorporate minimal automation, such as basic cruise control, while Level 4 systems are highly automated but may still necessitate driver intervention under certain conditions.

Designing automated systems in accordance with the principles of MHC is essential for addressing responsibility gaps—particularly in contexts where ethical decision-making depends upon clearly defined parameters for human intervention and accountability (Cavalcante Siebert et al., 2023; Santoni de Sio & Mecacci, 2021). Even when human operators are not directly managing a system's real-time functions, they must retain meaningful control over its behaviour to ensure ongoing oversight and the appropriate assignment of responsibility.

While there is broad consensus in the literature regarding the importance of maintaining MHC in the context of automation (Mecacci & Santoni de Sio, 2020; Cavalcante Siebert et al., 2023; Calvert et al., 2024), there is less agreement on how MHC should be conceptualised and implemented (George et al., 2023; Santoni de Sio & Van den Hoven, 2018; Kwik, 2022; Steen et al., 2023). Despite these differing interpretations, Robbins (2023) identify the framework developed by Santoni de Sio and Van den Hoven (2018) as a valuable foundation for designing systems in line with MHC principles. In their work, Santoni de Sio and Van den Hoven (2018) proposed a philosophical framework through which systems can be evaluated for meaningful human control, outlining two key conditions that must be satisfied: *tracking* and *tracing*.

The *tracking* condition requires that automated systems respond appropriately to the relevant reasons of the human agents involved in their design and deployment. These "reasons" can be understood as expectations—that is, the considerations that justify how an automated system ought to behave to align with human values, objectives, and societal norms (Veluwenkamp, 2022). For clarity, the term 'expectations' will be used throughout this paper to refer to these reasons. In essence, the tracking condition stipulates that the behaviour of automated systems should reflect the expectations of the relevant human stakeholders.

The *tracing* condition, by contrast, requires that automated systems be designed in a manner that enables their actions to be attributed to at least one human agent involved in their development or operation. Tracing presupposes the existence of an individual who not only understands the system's functionality but also accepts moral responsibility for its behaviour.

Taken together, the tracking and tracing conditions proposed by Santoni de Sio and Van den Hoven (2018) provide a foundational conceptual framework for operationalising meaningful human control in cooperative and automated driving contexts (Calvert & Mecacci, 2020), as well as for the broader design and engineering of automated systems, including automated vehicles (Cavalcante Siebert et al., 2023).

1.2. Evaluation of MHC over partially automated driving systems

To ensure that MHC principles are upheld, comprehensive assessments of partially automated driving systems are essential. This involves examining how well these systems comply with MHC principles by evaluating both the tracking and tracing conditions (Mecacci & Santoni de Sio, 2020). In the context of automated driving systems, tracking emphasises that the system should respond to the expectations of its designers and the humans who interact with it. For example, if a driver of a partially automated system expects the system to comply with road regulations, the system should behave in accordance with those regulations to effectively track the driver's expectations.

Tracing, on the other hand, requires that at least one human agent involved in the design or operation of the system understands its capabilities and accepts moral responsibility for its actions. In the context of automated driving systems, this means that drivers must be fully aware of their supervisory role and receive adequate training to supervise and intervene when necessary (Cabrall et al., 2019).

Several studies evaluating MHC have employed the tracking and tracing framework as a basis for analysis. For instance, Calvert et al. (2020) used the framework to evaluate partially automated driving systems, while Calvert et al. (2021) applied these criteria to assess cooperative vehicles and truck platooning systems.

While these contributions offer valuable insights into the assessment of partially automated driving systems, they primarily rely on hypothetical scenarios or post-incident analyses. Notably absent from much of the existing literature are the subjective experiences of real-world users of such systems. Yet these experiential insights are critical for understanding how users interact with automated driving technologies in everyday contexts. This perspective is essential for ensuring appropriate system use, a core element of both MHC and broader traffic safety considerations (Cavalcante Siebert et al., 2023).

Recent work by Suryana et al. (2024) has begun to address this gap by examining drivers' perceptions of safety and trust in relation to the tracking dimension of MHC. However, comprehensive evaluations of MHC compliance—encompassing both the tracking and tracing conditions—based on users' subjective experiences remain limited in the current literature.

1.3. Research gaps and objectives

1. **Theoretical Gap**: There is a lack of clarity regarding the application of tracking and tracing methodologies to assess MHC in real-world driving contexts. This issue is particularly critical, as previous studies have demonstrated that drivers frequently exhibit unsafe behaviours—such as complacency, falling asleep behind the wheel, or engaging in non-driving activities—while using automated systems (Wörle & Metz, 2023; Nordhoff et al., 2023). Such behaviours challenge adherence to MHC principles and raise concerns about whether these systems are genuinely under meaningful human control in everyday driving scenarios.
2. **Practical Gap**: Existing assessments of MHC have largely neglected the subjective experiences of drivers operating partially automated systems in real-world settings. For example, the ways in which drivers perceive their supervisory role, interpret system behaviour, and how their perceptions of accountability evolve over time remain insufficiently explored. These experiential factors are essential for determining whether partially automated systems are truly under meaningful human control.
3. **Methodological Gap**: Current approaches to evaluating MHC often overlook critical elements of human-automation interaction. They fail to investigate whether the system's performance consistently aligns with human expectations, or whether drivers fully comprehend their responsibilities and are capable of reclaiming control when necessary. These limitations hinder the effective evaluation of meaningful human control in real-world driving contexts.

To address these gaps, this study applies the framework of MHC to real-world driving contexts, drawing on previously collected interview data from users of Tesla Autopilot and Full Self-Driving Beta systems (Nordhoff et al., 2023). By moving beyond hypothetical scenarios and post-accident analyses, this research offers a dynamic assessment of MHC in everyday driving situations. It further investigates how drivers perceive their responsibility in supervising automation, the evolution of their trust and safety perceptions, and how they interpret system behaviour—dimensions that have been largely neglected in prior evaluations. Finally, by employing a qualitative methodology that captures the nuanced and context-dependent nature of human-automation interaction, this study provides a more comprehensive approach to evaluating MHC compliance. Collectively, these contributions deepen the understanding of meaningful human control in partially automated driving systems, offering valuable insights for both theoretical development and practical design improvements aimed at enhancing the safety and accountability of driving automation.

2. Method

2.1. Dataset

This study draws on a dataset comprising 103 semi-structured interviews with active users of Tesla's Autopilot and Full Self-Driving (FSD) Beta systems. The interviews focused on participants' real-world experiences and interactions with these technologies, capturing a broad range of topics including perceived safety, trust, control, and responsibility.

Although participants were not explicitly introduced to the concept of Meaningful Human Control (MHC), the interviews contained numerous responses that align with its theoretical components—specifically, aspects related to tracking (e.g., alignment between system behaviour and human expectations) and tracing (e.g., attributions of responsibility and control). This made the dataset well-suited for retrospective analysis through the lens of the MHC framework.

Details regarding recruitment and study procedures are provided in the following subsections.

2.1.1. Recruitment

The dataset utilised in this study was collected through a recruitment process and interview procedure approved by the Human Research Ethics Committee of Delft University of Technology (ID: 2316). Participants were initially identified through special interest groups related to Tesla vehicles on various social media platforms, including Discord, Facebook, Twitter, Reddit, YouTube, Instagram, Tesla Motors Club, and the Tesla Motors Forum. Snowball sampling was subsequently employed, with participants referring others via email. As Full Self-Driving (FSD) Beta was available only to residents of North America and Canada during the study period, recruitment efforts were predominantly focused on these regions. Eligibility for participation was determined based on self-reported access to Autopilot and FSD Beta. FSD Beta users were individuals selected by Tesla according to safety scores and ownership status. Prior to granting access, Tesla provided the following usage guidelines to FSD Beta users:

"Full Self-Driving is in limited early access Beta and must be used with additional caution. It may do the wrong thing at the worst time, so you must always keep your hands on the wheel and pay extra attention to the road. Do not become complacent. When Full Self-Driving Beta is enabled, your vehicle will make lane changes off highway, select forks to follow your navigation route, navigate around other vehicles and objects, and make left and right turns. Use Full Self-Driving Beta only if you will pay constant attention to the road, and be prepared to act immediately, especially around blind corners, crossing intersections, and in narrow driving situations. Every driver is responsible for remaining alert and active when using Autopilot and must be prepared to take action at any time. As part of receiving FSD Beta, your vehicle will collect and share VIN-associated vehicle driving data with Tesla to confirm your continued eligibility for FSD Beta feature. If you wish to be removed from the limited early access FSD Beta please email xxx."

2.1.2. Procedure

Interviews were conducted remotely via Zoom, with both audio and video recordings. To ensure consistency and minimise interview bias, a predefined interview protocol was developed using Qualtrics (https://www.qualtrics.com). The link to the protocol was shared with participants via Zoom's chat function at the commencement of the interview, enabling them to follow the questions and progress through them independently. This approach was specifically designed to reduce the interviewer's potential influence on participants' responses.

At the outset of the interviews, participants provided their informed consent to take part in the study. The first section of the interview primarily comprised open-ended questions, focusing on participants' perceptions and experiences with Autopilot and Full Self-Driving (FSD) Beta, including aspects such as feelings of safety, trust, and typical usage (see Table A.5). For example, participants were asked to describe situations in which they felt unsafe using these systems, as well as how their trust and safety perceptions evolved over time. The second section of the interview comprised closed-ended questions concerning participants' socio-demographic profile, travel behaviour (e.g., age, gender, education level, frequency of Autopilot/FSD Beta use), and their general attitudes towards traffic safety.

The interviewer's role was primarily observational, intended to minimise bias by allowing participants to navigate the questionnaire independently. However, follow-up questions were posed to clarify responses or explore new themes that emerged during the interview. Participants were also encouraged to skip any questions that had already been addressed. The interviews lasted an average of 78 minutes, resulting in approximately 12,200 words of transcribed data.

To ensure the integrity of the data, four interviews conducted in German were excluded from the analysis to avoid potential issues associated with missing transcriptions or mistranslations that could arise from translating the responses into English. Consequently, 99 of the original 103 interviews were considered suitable for further analysis.

2.2. Data analysis

An evaluation framework for MHC was developed to assess whether Tesla's Full Self-Driving (FSD) Beta and Autopilot systems align with the expectations of relevant human agents (tracking), and to what extent individuals involved in the operation and design of these systems understand their capabilities and recognise their moral accountability for the systems' actions (tracing). This evaluation follows a structured five-step process, as detailed below.

2.2.1. MHC component identification

Tracking component: (1) human agents and (2) their expectations In this step, we identified the human agents and their safety expectations in order to evaluate tracking alignment. Using the MHC taxonomy as a framework (Fig. 1), we defined two categories

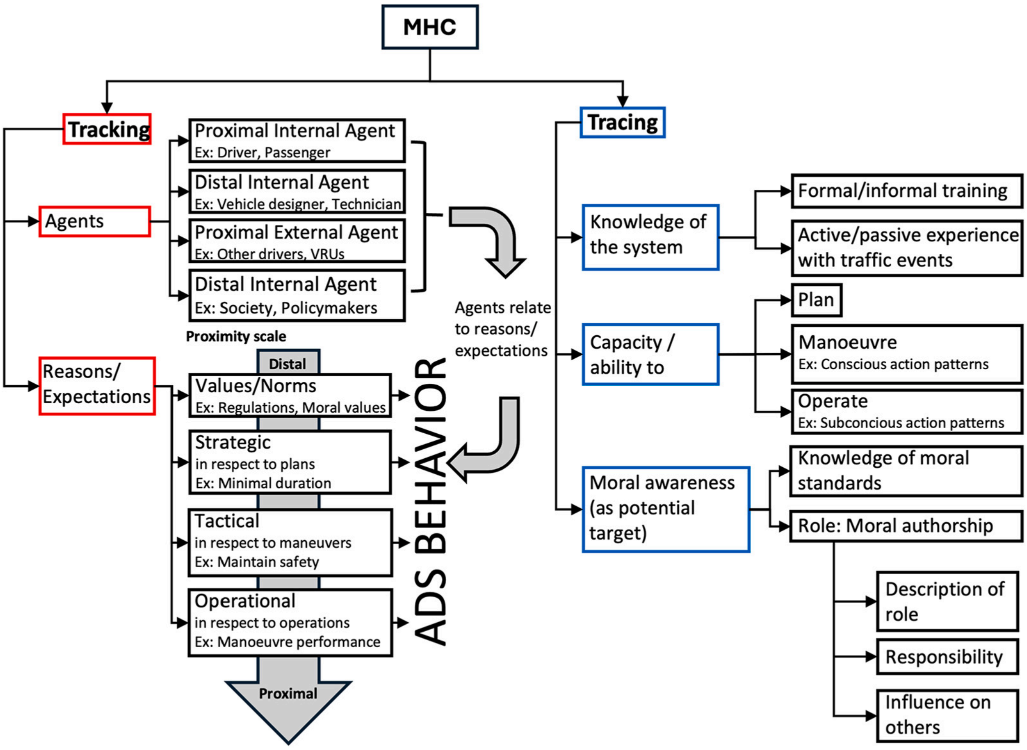

Fig. 1. Taxonomy of tracking and tracing, adapted from the work of Calvert and Mecacci (2020).

of human agents based on their relationship with the system: *drivers*, classified as proximal internal agents (those who interact directly with the system), and *manufacturers*, classified as distal internal agents (those responsible for designing and regulating system functionality). We also defined their *safety expectations* as tactical expectations, reflecting real-world interactions. Specifically, drivers expected the system to prevent accidents (e.g., by providing collision warnings or automatically applying the brakes in emergency situations), while manufacturers expected the system to comply with safety standards (e.g., meeting regulatory requirements for collision avoidance). This categorisation provided a structured framework for evaluating whether system behaviour aligns with the safety expectations of these human agents.

Tracking component: (3) features that influence vehicle behaviour In addition to defining human agents and their expectations, this step also identifies the active safety features in Tesla's Autopilot and FSD Beta systems that directly influence the vehicle's behaviour and contribute to meeting safety expectations. These features act as key indicators of how effectively the system tracks and responds to the expectations of human agents.

- *Automatic Emergency Braking (AEB):* Detects vehicles or obstacles in the vehicle's path and applies the brakes if necessary.
- *Forward/Side Collision Warning (F/SCW):* Alerts the driver to potential collisions with slower-moving or stationary vehicles or obstacles alongside the vehicle.
- *Blind Spot Monitoring (BSM):* Warns the driver when a vehicle or obstacle is detected in the blind spot during lane changes.
- *Lane Departure Avoidance (LDA):* Applies corrective steering to assist in keeping the vehicle within its intended lane.

These features were selected because they directly influence the vehicle's behaviour and are critical for ensuring safety in real-world driving scenarios. By focusing on these features, we were able to assess how effectively the system tracks and responds to the expectations of human agents, thus providing a robust foundation for evaluating alignment with the MHC principle.

Tracing component To evaluate tracing, it is necessary to identify a human agent who understands the system's capabilities and recognises their moral accountability in the design and operation of the system. In the case of Tesla, we selected the *driver* as the accountable human, as the company explicitly assigns this responsibility to drivers through its operational guidelines (Tesla, 2024). Prior to engaging Autopilot, drivers must agree to 'keep their hands on the wheel at all times' and to 'remain in control of and responsible for their vehicle at all times.' This requirement highlights the driver's role as the primary human responsible for overseeing system performance and intervening when necessary.

Table 1
Tracing evaluation criteria.

Criteria	Details
Knowledge	(1) To stay alert and (2) To keep both hands on the steering wheel
Capability	To be able to perform corrective action
Moral awareness	To maintain operational responsibility

2.2.2. Defining MHC evaluation criteria

Tracking evaluation criteria To assess whether Tesla's Autopilot and Full Self-Driving (FSD) Beta systems align with human agents' safety expectations, we adapted two evaluation criteria: *Safety of the Intended Functionality (SOTIF)* (International Organization for Standardization, 2019) and *Perceived Safety and Trust (PST)*. The SOTIF framework was selected because evaluating the safety of automated driving systems necessitates a standardised approach, while PST was included because even technically safe systems may fail to align with human expectations if their behaviour is perceived as unpredictable or unreliable.

These criteria were chosen to evaluate both the technical performance of the system and the subjective experiences of drivers. For SOTIF, we employed an adjusted version, termed ad-SOTIF, to compare drivers' descriptions of system behaviour with Tesla's official specifications. If the system's behaviour aligned with the manufacturer's descriptions, it was classified as ad-SOTIF (+); deviations were classified as ad-SOTIF (-).

For PST, we assessed drivers' perceptions of safety and trust based on their interview responses. As our study evaluates safety expectations through driver perceptions, PST serves as a proxy for assessing tactical expectations, as depicted in the tracking taxonomy in Fig. 1. Trust was incorporated as a criterion due to its strong positive relationship with perceived safety, given that trust is often modelled as a function of perceived safety (Nordhoff et al., 2021). This approach enabled us to capture additional facets of drivers' safety experiences that may not be explicitly expressed through the word "safe" in interviews, thereby providing a more comprehensive understanding of their perceptions. Positive perceptions, such as feelings of reliability or confidence, were classified as PST (+), while negative perceptions, such as distrust or feelings of risk, were classified as PST (-).

This dual approach enabled us to assess both the technical alignment of the system with its intended functionality and the subjective experiences of drivers, thereby ensuring a comprehensive evaluation of whether the system meets the safety expectations of both drivers and manufacturers.

Tracing evaluation criteria To evaluate driver compliance with tracing requirements, we derived three criteria from Tesla's usage guidelines (Nordhoff et al., 2023), as outlined in Section 2.1.1, and aligned them with the MHC tracing taxonomy (Calvert & Mecacci, 2020). These criteria include: *knowledge* (staying alert and keeping hands on the steering wheel), *capability* (performing corrective actions), and *moral awareness* (maintaining operational responsibility).

These instructions provide the foundation for operationalising the criteria. For instance, the requirement to "keep hands on the wheel" was categorised under knowledge, while "be prepared to act immediately" was mapped to capability. By grounding the criteria in both Tesla's instructions and the MHC framework, this step ensures a structured evaluation of driver compliance with tracing requirements. The criteria are summarised in Table 1.

2.2.3. Locating MHC-related content

The process of locating MHC-related content in the interview data depends on whether the questions are already aligned with the tracking and tracing evaluation criteria. In instances where the questions were not directly related, additional steps were required to identify and extract relevant content. For example, in our study, the interview questions were not explicitly designed to address tracking criteria, necessitating a more detailed preprocessing and keyword search approach. This method was essential due to the unstructured nature of the data, and keyword searches enabled us to systematically identify segments of the interviews that discussed specific safety features (e.g., Automatic Emergency Braking or Lane Departure Avoidance). In contrast, the tracing criteria were addressed through specific questions in the interview protocol, facilitating the direct extraction of relevant responses. Below, we outline the distinct methodologies used for locating content related to tracking and tracing.

Locating tracking-related content: (1) data preprocessing To systematically identify tracking-related content in the interview data, we began by preprocessing the transcribed text, transforming it into word tokens and applying several cleaning steps. Preprocessing ensures that only meaningful content is retained, eliminating noise that could affect the accuracy of subsequent keyword searches. In line with best practices in text analysis (Banks et al., 2018a; Hickman et al., 2022), we used the NLTK Python package (NLTK Project, 2024) to perform the following steps: (1) removal of newline characters and extra spaces, (2) tokenisation into individual words, (3) elimination of short words (< 2 characters) and long words (> 30 characters), numbers, and punctuation, (4) conversion to lowercase and filtering of English stopwords, and (5) lemmatisation to normalise words to their root forms. For example, the sentence *One advantage of the larger Autopilot was that it could automatically stop at traffic lights.* was transformed into the list *['one', 'advantage', 'larger', 'autopilot', 'automatically', 'stop', 'traffic', 'lights']*. This preprocessed dataset was then utilised in subsequent keyword searches.

Locating tracking-related content: (2) identifying seed words To identify tracking-related content, we selected seed words associated with the four active safety features (AEB, F/SCW, BSM, and LDA), which act as indicators of relevant discussions within the

Table 2
Seeds related to active safety features.

Category	Sub-category	Knowledge-based seeds	Seeds for keyword search
Active safety features	AEB	emergency AND braking	(emergency OR urgent OR disaster OR immediate OR assistance) AND (braking OR deceleration OR steering OR traction OR acceleration)
	F/SCW	collision AND warning	(collision OR accident OR collide OR crash OR head-on OR mishap) AND (warning OR warn OR alert OR indication OR danger OR caution)
	BSM	blind AND spot AND monitor	(blind OR mistaken OR sight OR impossible) AND (spot OR place OR there OR where) AND (monitor OR tracking OR surveillance OR alerting OR evaluation OR utilization)
	LDA	lane AND keeping	(lane OR road OR freeway OR crossing OR roadway OR highway OR ramp) AND (keeping OR kept OR keeps OR putting OR bringing OR maintain)

interview data. These safety features represent broader themes, while the seed words serve as specific indicators to help locate pertinent content. Adopting a knowledge-based approach (Watanabe & Zhou, 2022), we chose initial seed words—such as "emergency," "braking," "collision," "warning," "blind spot," and "lane departure"—based on their strong association with these safety features.

Once the initial seed words were selected, we applied the pre-trained Global Vector (GloVe) model in Python to enhance the seed word set. The GloVe model, a machine learning technique for generating word embeddings, was employed to identify synonyms and semantically related terms that may not have been initially considered. Details of the pre-trained model and setup instructions are available at the official GloVe project page (https://nlp.stanford.edu/projects/glove/). This enrichment process strengthened the robustness of the keyword search by ensuring that a broader range of relevant terms could be identified within the interview data. For example, the seed word "braking" was enriched with terms such as "deceleration," "traction," and "acceleration," while "collision" was expanded to include "accident," "collide," and "crash." The final enriched seed word set was then used in a systematic search to locate content pertinent to the active safety features. Table 2 presents a detailed overview of the initial and enriched seed words, illustrating the outcomes of this process. By combining both expert knowledge and machine learning techniques, this step ensured that the keyword search algorithm effectively identified relevant interview content.

Locating tracking-related content: (3) keyword search algorithm To systematically identify tracking-related content in the interview data, we applied a keyword search algorithm proposed by Suryana et al. (2024), which utilises enriched seed words to detect relevant segments. This algorithm was applied to the lemmatised tokens generated during the data preprocessing phase. It employed the enriched seed words, generated by the GloVe model, to scan the tokenised data and identify segments where the seed words appeared (see Algorithm 1).

The algorithm incorporated logical operators to refine the search process. The 'OR' operator allowed the inclusion of synonyms for the seed words, while the 'AND' operator ensured that paired seed words, as defined in Table 2, appeared together within a 20-token sliding window in the lemmatised, tokenised data. The choice of a 20-token window was informed by prior work (Suryana et al., 2024), which demonstrated that this window size effectively captures meaningful contextual relationships between related terms in similar textual analyses. For instance, when applying the knowledge-based seed words for AEB, "emergency" and "braking," the algorithm would detect the occurrence of the word "emergency" in the token sequence and then scan the subsequent 20 tokens to check for the presence of the paired seed word "braking." If both seed words were found within this 20-token window, the corresponding segment of the original transcribed interview would be extracted for further analysis using classifications such as ad-SOTIF(+), ad-SOTIF(-), PST(+), or PST(-).

Algorithm 1: Keyword Search Algorithm (Suryana et al., 2024).

Data: Seed words (seeds), tokenised data (tokenList), Buffer size (bufferSize)
Result: Retrieved list (list)

```
bufferSize ← 20;
list ← [];
threshold ← ∑(seed in seeds);
for token, index in tokenList do
    if token in seeds then
        tokenBuffer ← tokenList[index : index + bufferSize];
        seedCount ← ∑_{seed in seeds}(seed in tokenBuffer);
        if seedCount > threshold then
            list ← tokenBuffer;
```

Locating tracing-related content: direct extraction from interview responses To identify and extract interview segments related to the tracing evaluation criteria defined in Step 2, we focused on responses to specific questions in the interview protocol: Q25, Q26, Q34, and a question concerning the maintenance of control and responsibility (see Appendix A). For example, Q25 asked, "Do you typically keep your hands on the steering wheel at all times?" and Q26 asked, "Are you typically fully attentive and alert at all times?", both of which directly relate to drivers' knowledge. Similarly, Q34 ("Do you typically stay prepared to take corrective actions at all times?") provided insights into drivers' capability. The question regarding the maintenance of control and responsibility was not explicitly stated in Appendix A, but it could be inferred from drivers' responses. For instance, when drivers read the question aloud and responded with statements such as, "Do not maintain control and responsibility for my car? I strongly disagree," it addressed the moral awareness criterion, ensuring that drivers recognised their accountability for the system's behaviour. The extracted responses were then prepared for qualitative assessment, facilitating a focused and efficient evaluation of drivers' understanding of their responsibilities.

2.2.4. MHC evaluation

Tracking evaluation: content analysis Following the extraction of tracking-related conversation segments in Step 3, a content analysis was conducted to classify the content based on the evaluation criteria defined in Step 3: ad-SOTIF and PST. The objective of this step was to determine whether Tesla's Full Self-Driving (FSD) Beta and Autopilot systems comply with the tracking requirements of the MHC framework.

For ad-SOTIF, we compared drivers' descriptions of active safety features in the interview data with the intended functionalities outlined on Tesla's official website (Tesla, 2024). If the descriptions aligned with the manufacturer's specifications, the features were classified as ad-SOTIF (+). For instance, if a driver described Automatic Emergency Braking (AEB) as functioning consistently with Tesla's description (e.g., braking automatically when an obstacle is detected), this was categorised as ad-SOTIF (+). Conversely, if drivers reported discrepancies or failures in system behaviour (e.g., AEB not activating when required), the features were classified as ad-SOTIF (-).

For PST, we evaluated drivers' perceptions of safety and trust based on their interview responses. To assess trust, we identified terms such as "depend," "rely," and "trust," which indicated whether drivers had a positive level of trust in the system. Similarly, terms related to safety, such as "relax," "risk," and "safe," were used to gauge drivers' perceived safety. These terms were selected based on established questionnaires for evaluating trust (Choi & Ji, 2015) and perceived safety (Xu et al., 2018). If drivers expressed confidence in the system's reliability and felt safe using it, the content was classified as PST (+). For example, a driver stating, "I feel relaxed using Autopilot because it handles most situations well," would be categorised as PST (+). In contrast, if drivers expressed distrust or felt unsafe (e.g., "I don't trust the system to handle sudden stops"), the content was classified as PST (-).

This qualitative analysis enabled us to classify the extracted content into four categories: ad-SOTIF(+), ad-SOTIF(-), PST(+), and PST(-). By combining these classifications, we were able to assess not only the technical alignment of the system with its intended functionality but also the subjective experiences of drivers. This dual approach ensured a comprehensive evaluation of whether the system meets the safety expectations of both drivers and manufacturers, as outlined by the MHC framework. The results of this analysis provided a structured basis for understanding how well Tesla's systems track and respond to human agents' needs, highlighting both areas of alignment and potential gaps.

Tracing evaluation: thematic analysis To evaluate whether drivers' experiences with Tesla's Autopilot and FSD Beta systems comply with the tracing evaluation criteria, we conducted a thematic analysis of their responses. This involved categorising responses into subcategories that reflected drivers' understanding of their responsibilities, knowledge, and capabilities. Following inductive coding principles (Nordhoff, 2024), we performed open coding, reviewing the extracted responses line-by-line to identify recurring themes, such as "keeping hands on the wheel," "monitoring the road," or "feeling responsible for interventions." These themes were then grouped into broader subcategories based on their similarities and distinctions. For instance, responses mentioning "hands on the wheel" and "staying alert" were grouped under a subcategory such as Compliance with Hands-on Requirements. To ensure robustness, we retained only those subcategories mentioned by at least five drivers, as this frequency threshold helped validate the relevance and significance of each subcategory. In cases where a single quote applied to multiple subcategories, each relevant subcategory was assigned a frequency count of one. This systematic approach ensured that the subcategories were both data-driven and representative of drivers' experiences, providing a structured foundation for further analysis.

2.2.5. Illustrative quotes

Tracking quotes Representative quotations were selected from the classified content to illustrate the findings. These quotations exemplify each of the four classifications: ad-SOTIF(+), ad-SOTIF(-), PST(+), and PST(-). For each category, excerpts from the interview data were chosen to clearly represent either alignment with or deviation from manufacturer specifications (ad-SOTIF), or to reflect positive or negative driver perceptions regarding safety and trust (PST).

Tracing quotes To provide concrete examples of the subcategories identified in Step 4, we selected up to three representative quotations per subcategory. Priority was given to quotes that clearly exemplified the theme and reflected common driver experiences.

3. Results

This section presents the results of our evaluation of Tesla's Full Self-Driving (FSD) Beta and Autopilot systems in relation to the concept of meaningful human control (MHC), with a specific focus on the tracking and tracing requirements. To illustrate how drivers' feedback aligns with these requirements, we include quotations that reflect their experiences. Each quote is accompanied by the participant ID number for reference. To highlight key insights, we have selected several representative quotes. The results, supported by these quotations, are further discussed in Sections 3.1 and 3.2.

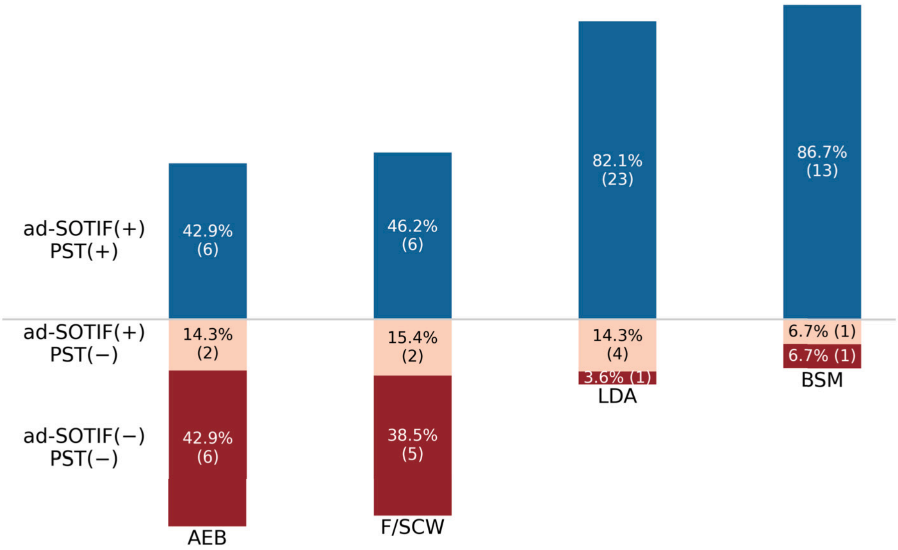

Fig. 2. Tracking evaluation results for the four active safety features of Tesla vehicles—Automatic Emergency Braking (AEB), Forward/Side Collision Warning (F/SCW), Lane Departure Avoidance (LDA), and Blind Spot Monitoring (BSM)—are presented. For each feature, the stacked bars represent the percentage of instances in which each performance category—alignment with intended functionality (ad-SOTIF) or perceived safety and trust (PST)—was mentioned in the interviews. The numbers in parentheses below the percentages indicate the total number of mentions for each category.

3.1. *Tracking evaluation results*

Our tracking evaluation revealed a varied distribution of safety features across the tracking evaluation criteria (Fig. 2). For example, in the ad-SOTIF(+) PST(+) category, Lane Departure Avoidance (LDA) and Blind Spot Monitoring (BSM) exhibit a higher percentage distribution compared to Automatic Emergency Braking (AEB) and Forward/Side Collision Warning (F/SCW). Percentage distributions above 80% indicate that LDA and BSM more frequently meet both driver and manufacturer safety expectations, compared to instances where they fail to align with one or both expectations. In contrast, there were no instances of ad-SOTIF(-) PST(+), suggesting that when the features did not perform as intended, drivers never held a positive perception of them.

To provide a more detailed insight into the tracking evaluation of the safety features, we classified them into three categories based on the co-occurrence of positive and negative instances of ad-SOTIF and PST. It is important to note that each safety feature could be assigned to multiple categories depending on the variation in user experiences:

- *Inconsistent tracking:* Safety features that were described both as (1) functioning as intended and generating positive perceptions of safety and trust, and (2) not functioning as intended and generating negative perceptions. A feature was assigned to this category when the number of instances with ad-SOTIF(+) PST(+) was comparable to those with ad-SOTIF(-) PST(-), indicating inconsistency in performance and perception.
- *Gap between performance and perceived safety/trust:* Safety features that technically functioned as intended but failed to generate positive perceptions of safety and trust. In such cases, although the features aligned with manufacturers' safety expectations, they failed to meet drivers' expectations. Features with a notable proportion of ad-SOTIF(+) PST(-) instances were assigned to this category.
- *Consistent tracking:* Safety features that not only functioned as intended but also consistently elicited positive perceptions of safety and trust among drivers. Features were included in this category when there was a high occurrence of ad-SOTIF(+) PST(+) and a low occurrence of ad-SOTIF(-) PST(-), indicating strong alignment with both driver and manufacturer safety expectations.

It is important to emphasise that although only a limited number of illustrative quotations are presented in the following sections, each theme was derived from multiple participant responses. The frequency with which each theme was mentioned varied; some were discussed by a larger number of participants, while others emerged less frequently. Moreover, individual responses often encompassed multiple themes, as participants' experiences with the system frequently addressed several aspects of its performance simultaneously. These tracking categories were developed to capture the full range of relevant patterns observed in the data, ensuring that both commonly and less frequently reported experiences were taken into account.

3.1.1. *Inconsistent tracking*

Automatic Emergency Braking (AEB) and Forward/Side Collision Warning (F/SCW) were found to align with both drivers' and vehicle manufacturers' safety expectations in certain scenarios, while failing to do so in others. To better understand how AEB and F/SCW can both meet and fall short of these expectations, we analysed participant feedback regarding the performance of these systems. The data revealed several recurring themes in drivers' perceptions of these features:

- **Effective functionality - ad-SOTIF(+) PST(+)**
  Participants mentioned that AEB performs well in detecting vehicles ahead that the driver may not see, thereby preventing potential collisions. One user expressed appreciation for this feature:

  "I would have actually hit someone, but they stopped suddenly for some reason, maybe someone was crossing.. I didn't see it, but the system detected it and prevented a collision by performing an emergency brake. It worked really well, and I'm very grateful (R047 - AEB)"

  Similarly, the F/SCW feature proved effective in alerting drivers to potential collisions from the front and sides. One participant expressed their appreciation for this feature:

  "An autopilot averted a potential accident.. I was very impressed. While driving, my car started.. telling me to take control immediately. I looked in my blind spot, and a car in the next lane veered into mine. I didn't see it, but Autopilot did and reacted right away (R091 - F/SCW)"

- **False positive and false negative errors - ad-SOTIF(-) PST(-)**
  Despite the overall positive performance, drivers also reported instances where AEB and F/SCW did not function as intended. For example, there were cases in which the automated systems failed to respond to debris on the highway and missed alerting the drivers. Both situations are considered false negatives, which led to feelings of unsafety among participants.

  "I think what's unsafe is just right now it only has a front collision warning.. It doesn't gonna detect anything..comes to you from the side.. If it does it.. only if you drive at the really slow speed. (R058 - F/SCW)"

  "If I didn't take over, it would drive right over the piece of wood and probably created a lot of damage that might have caused an accident because hitting at highway speeds, a piece of debris.. Tesla uses cameras as their technology, but you could probably detect better debris and just alert.. like they have some alerts when you're driving if it's uncertain. So they could do that to make it safer (R073 - AEB and F/SCW)"

  Additionally, the system sometimes engaged in "phantom braking," a false positive case in which the brakes were applied without the presence of an actual obstacle. This led to annoyance among drivers:

  Autopilot take care of 99% of driving.. The only issues.. it's not a perfect system.. there are a lot of false positives, particularly in one lane roads where in cars are coming at you fast. It sometimes thinks it's going into your lane and does a phantom brake. In the case.. it.. annoys you by saying, "hey, there's a forward collision warning" when it's not. (R078 - AEB and F/SCW)

- **Software issues - ad-SOTIF(-) PST(-)**
  Drivers also reported unsettling software issues, including automatic warnings upon vehicle reboot and inconsistencies in alarm triggering. These problems contributed to undesirable experiences among drivers:

  "It was.. scary enough that.. a non informed user would not know what to do. Autopilot would constantly disengage my visualization.. rebooting about every three seconds, and every time it rebooted, a forward collision warning would occur. It would not slow down my car, but it would make like the super loud multiple beeps like I'm gonna hit something. (R055 - F/SCW)"

Overall, our analysis indicates that the AEB and F/SCW systems generally align with both driver and vehicle manufacturer safety expectations under typical driving conditions. Specifically, AEB was frequently noted for its effectiveness in detecting vehicles ahead, while F/SCW was recognised for its ability to alert drivers to potential frontal and lateral collisions.

However, participants also reported instances where these systems failed to meet expectations. These failures included both false positives and false negatives. A commonly reported false positive was phantom braking—where the vehicle applied the brakes without a discernible obstacle. False negatives included failures to detect road debris, side collisions, or to provide timely warnings to the driver. In addition, several participants reported software inconsistencies, such as unexpected system reboots, which further undermined the reliability of the tracking function. These issues suggest that while AEB and F/SCW often perform as intended, limitations remain that affect their consistency and overall effectiveness.

3.1.2. Gap between performance and perceived safety/trust

Three safety features—AEB, F/SCW, and LDA—were classified under this category.[1] This classification was based on driver responses indicating that these features generally functioned as intended but were nonetheless associated with negative perceptions of safety and trust (PST). Based on participants' feedback, we identified several potential causes for this perception gap, which are outlined below.

[1] Blind Spot Monitoring (BSM) was not included in this category, as the only driver who expressed low trust in the BSM system attributed this to the placement of the warning symbol rather than concerns about its functional performance.

- **Premature collision warnings - ad-SOTIF(+) PST(-)**
  One participant described a situation in which their vehicle issued collision warnings for vehicles that were still a considerable distance away and then abruptly applied the brakes. Although this suggests that the warning system was effective at alerting the driver to potential future collisions, the participant felt that these warnings were unnecessary, as the vehicle in front still had sufficient time to complete the turn before the partially automated driving system reached it. This issue led to frustration with the system:

  "One of the annoying things.. this is a little tiny bit as safe.. You're from Holland.. So you are on the right side of the road.. So when you're driving.. someones turning left in front of you and they're like way ahead.. like the test slams on the brakes. Sometimes with the forward collision warning and.. it's like 200 meters ahead of you.. like they'll easily turn out past you.. But.. rear ending potential.. that's the worry. (R006 - F/SCW)"

- **Inadequate distance - ad-SOTIF(+) PST(-)**
  Finally, one driver reported that the vehicle failed to maintain a sufficiently safe distance from a parked vehicle, even though it did not result in a collision. This situation made them feel unsafe:

  "I'm not gonna say terribly unsafe, but uncomfortable. I do feel very unsafe if there's vehicles parked on the right hand side and the vehicles attempting to maintain the lane, but it comes far too close to the vehicles on the right hand side. That is very I feel that's very unsafe and that's very stressed. (R041 - LDA)

- **Inappropriate braking - ad-SOTIF(+) PST(-)**
  The driver described experiences in which the vehicle's emergency braking behaviour was problematic, particularly due to hesitation after braking in heavy traffic. While the system still performed effectively in preventing collisions, the driver implicitly expressed concerns about safety due to this behaviour. Specifically, if traffic clears and speeds up, such behaviour could disrupt the flow of traffic, as it acts in a way that is not anticipated by other drivers.

  "The emergency brake checking that goes on, where the car will break.. in heavy traffic. It's okay when the car's hitting the brakes and hesitating. But when it opens up, and we're moving faster, and there's more space.. People are anticipating you to stay at your speed.. You don't want.. hitting the brakes at those speeds.. those are the biggest situations. (R081 - AEB)"

Although AEB, F/SCW, and LDA successfully tracked vehicle manufacturers' safety expectations—such as braking to prevent collisions in heavy traffic, issuing warnings of potential collisions, and maintaining lane position—drivers reported several concerns that negatively affected their perceptions of safety and trust. These issues suggest that the active safety features did not consistently align with drivers' safety expectations. For example, AEB was reported to brake unnecessarily or hesitate in dense traffic conditions, disrupting traffic flow and raising safety concerns. F/SCW was criticised for issuing premature warnings and engaging in unnecessary braking when no imminent threat was present, often leading to frustration. LDA was noted for failing to maintain a safe lateral distance from parked vehicles, which resulted in driver discomfort and a diminished sense of security.

3.1.3. Consistent tracking

Two active safety features—BSM and LDA—were categorised as exhibiting consistent tracking, as they were reported to effectively meet both vehicle manufacturers' and drivers' safety expectations in most scenarios.

The specific situations in which drivers indicated that LDA successfully aligned with their safety expectations are outlined below.

- **Long trips - ad-SOTIF(+) PST(+)**
  Two drivers highlighted how their vehicle performed exceptionally well in maintaining its lane during long trips, particularly on highways. They described the feature as "flawless," suggesting that they perceived the system as both safe and trustworthy.

  I did a.. 7500 Mile road trip from Connecticut to California and back.. 99% of the trip on the highways.. was done using Autopilot. And it worked pretty much flawless. (R026 - LDA)

  Additionally, one driver noted that the system outperformed human drivers, particularly in maintaining focus and avoiding complacency during extended drives. This suggests that the driver trusted the system to remain vigilant and to avoid complacency.

  For autopilot.. used on a highway, I would say I'm the worst driver in the fact that it does a better job of long distance drives keeping lane centred. You know, watching.. not becoming complacent, I guess, which is so easy on a longer drive. (R087 - LDA)

- **Managing complex highway infrastructure - ad-SOTIF(+) PST(+)**
  Two drivers highlighted how their vehicle assisted them in navigating complex highway traffic. One driver emphasised that FSD Beta maintained their lane and did not drift into incoming on-ramps, noting that it performed merging manoeuvres better than Autopilot.

There's been a few highways where FSD beta can be engaged at highway speeds. And it does solve many of the problems they've had with navigate on autopilot, and then it merges better. It doesn't shift over into incoming on ramps like navigating like I'll it does it steers better. It maintains speed better overall. (R007 - LDA)

Another driver described a highway they considered a "scary" place to drive due to the numerous on- and off-ramps on both sides of the interstate. They noted that Autopilot kept them in the correct lane, something they felt they might not have been able to maintain on their own.

I don't think I could get through Atlanta if I didn't have Autopilot because their interstate is twice as wide as ours and they have on ramps and off ramps on both sides of the interstate. And it is a crazy, hectic, scary place to drive and. If I didn't have autopilot keeping me where I needed to be, I don't think I could do it. Nerves of steel there and I don't have it. (R099 - LDA)

- **Less mental workload - ad-SOTIF(+) PST(+)**
  Drivers consistently reported experiencing a reduced mental workload when using the vehicle's lane-keeping features. This reduction in cognitive effort was attributed to the vehicle's ability to handle routine tasks, such as maintaining lane position and adjusting for nearby traffic. One driver described how the system's reliability in keeping the car centred in the lane fostered a sense of security, allowing them to relax and trust the technology:

  I trust full self-driving to keep me in my lane. So no, I don't pay as close attention to where I am in the lane. I trust that it's keeping me in the lane. (R044 - LDA)

  One driver emphasised that the system's effectiveness in maintaining lane position significantly reduced fatigue, leading to a more positive driving experience.

  And the reason it makes it a lot less fatigue.. is that you don't have to mentally think about all the micro adjustments. So when you're driving down the road, you have to constantly make sure you're centred in the lane, make sure you're keeping distance from the car in front of you.. That's my experience.. I really positive with Autopilot now for Full Self Driving Beta. (R079 - LDA)

  Another driver described how the system allowed them to shift their focus towards broader situational awareness, which they considered a safer and more efficient way of driving:

  I'm no longer having to concentrate on keeping that car.. I just simply don't even think about it anymore. In fact, it's odd when I take it off of all the pilot effect.. this is like starting out driving again all over because it's just something you get used to that the car keeps it so well on its lane that you just don't think about that anymore. What you do is looking ahead. You're looking for other things happening to you.. and you're making sure that you react appropriately. Does really good. (R062 - LDA)

Despite the reported excellence of LDA, one driver highlighted a situation where LDA kept them in the wrong lane, which led to them feeling scared. While LDA was still functioning as intended by keeping the vehicle within the lane, it did so in the wrong lane.

- **Stay on the wrong lane - ad-SOTIF(-) PST(-)**
  "It's really scary. It just does all sorts of weird things today. I was like coming home from work and it stayed. It was two lane road. It stayed in the left lane, which turned into a turn lane and it just like blew right through the turn lane and just kept writing through. We call it here as suicide lanes where you have a you can make a left or a right turn either direction. And it just kept driving right through it. (R076 - LDA)"

The following aspects were highlighted regarding the effectiveness of the BSM system.

- **Safe lane changing - ad-SOTIF(+) PST(+)**
  Drivers consistently praised the vehicle's BSM system as a crucial safety feature that enhances the driving experience during lane changes. One participant highlighted that the system effectively monitors the vehicle's surroundings and facilitates safer lane changes by detecting vehicles in blind spots. This feature was reported to significantly increase their confidence and sense of safety while manoeuvring between lanes.

  A very complete functionality, features and.. ability to.. monitor everything around you and that lets you change lanes if there's a car in your blind spot or coming through and you're using it and stuff like that. Definitely makes me feel much safer when I'm doing it. (R026 - BSM)

  Another participant expressed appreciation for the BSM feature, noting that it helped prevent accidental lane changes resulting from limited peripheral vision. They found the BSM display particularly useful for enhancing situational awareness and reducing the likelihood of unintended lane merges.

When we go on vacation.. we're gonna be doing a lot of miles.. driving across the country. It takes a lot of.. drive.. I.. really enjoy it because I am blind on my entire right side. I have no peripheral vision, so it makes it with the screen being there and you know, blind spot awareness and all of those interesting features. It makes it harder for me to accidentally merge into someone if I don't look forward enough to the side to see if anybody's in there. (R099 - BSM)

- **Understanding what the system perceives - ad-SOTIF(+) PST(+)**
  Drivers reported that the Blind Spot Monitoring (BSM) system enhances their awareness of the vehicle's surroundings. One participant expressed appreciation for the system's graphical display, which allowed them to compare the vehicle's sensor feedback with their own visual observations. This feature was perceived as highly accurate and contributed to a greater sense of situational awareness.

  I would say.. 90% of the time my eyes are on the road. You typically monitor vehicle and its surroundings at all times.. I also enjoy the graphic that it gives you so you can understand our like to constantly compare with the vehicle sees to what I see and see what I can spot. That vehicle doesn't yet. And for the most part it's. Like 95% accurate. (R032 - BSM)

  Another participant emphasised that the visual feedback provided by the system on the display screen enhanced their confidence, as it allowed them to see exactly what the vehicle was detecting. This level of transparency contributed to a greater sense of safety and environmental awareness, reinforcing their trust in the system's ability to identify and avoid potential hazards.

  For both of them it's. You know, I feel safer because I see the perception. On the screen so I can see what it sees. And you know that gives me confidence of. Knowing exactly what it is seeing compared to.... And a lot of the perception part of this avoids that. Avoid those situations or helps avoid the situations. (R087 - BSM)

However, one participant reported a case in which the BSM system did not function as intended, resulting in an unsafe situation. According to the driver, the malfunction was caused by direct sunlight interfering with the sensor's ability to detect surrounding vehicles.

- **Weather-related sensor limitations - ad-SOTIF(-) PST(-)**
  "The place where I feel it's starting to get unsafe is the changing weather conditions. And sometimes lighting. That's one other one. When you get a bright hit of sunlight across into one of the panel doors, it'll just blind the camera. It can't compensate, and some levels. And I think they're gonna have to improve some of the cameras all around the car to be able to decrease their contrast to avoid it. These are the situation with you so unsafe. (R061 - BSM)"

Both LDA and BSM were noted for effectively tracking both vehicle manufacturers' and drivers' safety expectations. Specifically, LDA was praised for its ability to maintain lane position during extended highway travel and in complex driving environments, contributing to reduced driver fatigue and mental workload. BSM was commended for enhancing safety and driver confidence during lane changes by reliably monitoring blind spots and improving overall situational awareness. This feature was particularly valued by drivers with limited peripheral vision, who found the system especially beneficial.

Despite these strengths, instances of tracking failures were reported. BSM occasionally failed to function correctly due to sensor interference from direct sunlight. In the case of LDA, one participant reported a failure to maintain lane position, although the precise cause of this issue could not be determined.

3.1.4. Summary

To summarise the tracking evaluation results, we aggregated the commonly reported situations for each safety feature across the three categories discussed in Sections 3.1.1 to 3.1.3 (Table 3). The classification of 'tracked' or 'not tracked' is based on the presence of recurring themes in participant responses for each safety feature, as described in the corresponding sections. If a particular theme was mentioned by participants, the safety feature was classified accordingly in the table.

This analysis revealed that for each safety feature, there are situations in which the feature successfully tracked both the driver's and the vehicle manufacturer's safety expectations, as well as situations where it did not. Notably, failures to track the vehicle manufacturer's safety expectations were always accompanied by failures to track the driver's safety expectations. However, the reverse was not always true; in some cases, the system met the manufacturer's expectations but failed to align with the driver's expectations.

3.2. Tracing evaluation results

Using thematic analysis, we evaluated tracing by identifying ten subcategories within participants' responses, corresponding to the four tracing evaluation criteria. We also analysed the number of drivers who mentioned each subcategory (Table 4). These subcategories provide insight into how drivers operationalise the tracing criteria in practice, offering a deeper understanding of how responsibility, knowledge, and control are perceived and enacted.

Table 3
Assessment of the tracking condition of meaningful human control based on common situations mentioned by users of partially automated driving systems. A positive mark (+) indicates that the respective human agent's expectations are tracked, while a negative mark (–) indicates that the expectations are not tracked. The final column indicates whether both the driver's and the automaker's expectations are tracked.

Safety Feature	Described situation	Tracking of driver's expectations (PST)	Tracking of automaker's expectations (ad-SOTIF)	Human expectations are ..
BSM	Driver can detect objects in their blind spot while driving	+	+	Tracked
	BSM's sensors dysfunction due to weather such as sunlight	–	–	Not tracked
LDA	Driving on long highway trips with complex driving conditions	+	+	Tracked
	Drivers don't have to perform minor adjustments of the vehicle within its lane	+	+	Tracked
	LDA keeps the vehicle on the wrong lane	–	–	Not tracked
	LDA maintains lane, but the distance with surrounding objects is too close	–	+	Partially tracked
F/SCW	F/SCW warns the driver of unseen potential front and side collisions with sound and on-screen icons	+	+	Tracked
	F/SCW warns the driver of potential collisions that are still distant	–	+	Partially tracked
	F/SCW responds to false positive information and does not react to false negatives	–	–	Not tracked
	Annoying warnings after system reboots	–	–	Not tracked
	Warning dysfunctions when vehicle with high speed approaches	–	–	Not tracked
AEB	AEB brakes to prevent collision in unforeseen/unexpected situations	+	+	Tracked
	AEB responds to false positive information and does not react to false negatives	–	–	Not tracked
	AEB brakes to prevent collision, but the driver dislikes the way it brakes	–	+	Partially tracked

Table 4
Sub-categories related to tracing evaluation criteria. For each sub-category, count indicates the number of participants who mentioned each sub-category.

Tracing evaluation criteria	Sub-categories	Count
Knowledge: keeping both hands on the steering wheel	Driving with both hands on the steering wheel	39
	Driving with one hand on the steering wheel	13
	Driving mode	16
Knowledge: staying alert	Observation of the surrounding situations	17
	Highway	7
	Driving mode	26
Capacity: corrective action	Control over steering wheel	19
	Control over the pedals	10
	Driving mode	28
Maintaining operational responsibility	Agree to maintain control and responsibility	19

The frequency of mentions also indicates that certain subcategories were discussed more frequently than others. For instance, the 39 references to driving with both hands on the steering wheel suggest that a relatively large number of participants either understood or actively practised this behaviour. This number is notably higher than the 13 mentions of driving with only one hand on the wheel.

To provide deeper insight, the following sections offer detailed explanations and representative quotations for each tracing evaluation criterion, along with their corresponding subcategories.

3.2.1. Knowledge: keeping both hands on the steering wheel

This tracing requirement concerns whether drivers possess adequate knowledge regarding system use. Specifically, we assessed whether participants understood the importance of keeping both hands on the steering wheel and whether they reported complying with this guideline. Based on the interview data, we identified three categories of responses that addressed this topic.

- **Driving with both hands on the steering wheel**
  Several participants reported consistently keeping both hands on the steering wheel. This behaviour was often attributed to legal requirements, with some noting that they adhered to this practice to avoid reprimands or penalties.

  "Do you typically keep your hands on the steering wheel at all times? I do (R041)"

  "According to the law, the hands must be on the wheel. I actually keep my hands on the wheel, and I feel the resistance. (R047)"

“I do keep my hands on the steering wheel mostly so I don’t get dinged. (R067)”

- **Driving with one hand on the steering wheel**
  Other participants reported typically keeping only one hand on the steering wheel. For some, this was primarily to meet the system’s torque detection requirements, while others adopted this behaviour when using Autopilot, often describing a more relaxed driving posture during such instances.

  “I typically keep one hand on the steering wheel at all times. I keep it there just enough to satisfy the torquing requirement, where there needs to be weight on the system.” (R054)

  “Do you always keep both hands on the wheel? No, generally, I keep one hand. So, I have my left hand always on the wheel. It’s usually on my knee, on the door, or on my elbow.(R087)”

  “Yes, I keep my hands a little bit stream at all times. When.. I’m not have my hands on the steering wheel, I either have one hand like on the bottom.. like one hand in this picture. But I at least always have one hand on the steering wheel. (R098)”

- **Driving mode**
  Participants reported varied behaviours concerning hand placement on the steering wheel depending on the driving mode. While some consistently used one hand when operating Autopilot, others indicated that they were more likely to keep both hands on the wheel when using FSD Beta. The responses also revealed a range of strategies for maintaining system engagement while using Autopilot, including resting hands underneath the wheel, intermittently jiggling it to satisfy system prompts, or applying continuous pressure with one hand to meet torque detection requirements.

  “I’ll usually have just one hand.. just lean my hand on the bottom of the steering wheel and let the weight kind of be enough to do it, so that’s generally how I drive with just to put enough pressure on it. Keep it constant pressure on it so it never really warns me about not putting pressure on. I tried to do things around, just occasionally do it, but that becomes more effort than just letting your hand rest on the steering wheel when I’m driving with it.. We usually just keep my hand sitting there resting there and it works.(R021)”

  “Then do you typically keep your hands on the steering wheel at all time? Autopilot no, FSD beta yes (R033)”

  “With Autopilot.. depends on where we’re.. if we’re on the highway.. where there are no obvious issues up ahead that I can see, what I’ll typically do is rest my hands underneath the wheel. And then as the prompts come up, I’ll just jiggle the wheel a little bit to make the prompt go away. With.. beta, most of the time.. Especially during turns. Typically.. I’ll have my hands at.. seven and four or something, and just let the wheel sort of brush up against my hands. And sometimes I’ll keep my hands off the wheel.. if I’m comfortable in this situation. But I’ve kind of learned not to do that. (R051)”

Overall, the evaluation of drivers’ knowledge regarding the requirement to keep both hands on the steering wheel revealed a range of practices. While some participants consistently used both hands in adherence to legal requirements and to avoid penalties, others adopted a more relaxed approach—maintaining one hand on the wheel primarily to satisfy the system’s torque detection, particularly when using Autopilot. Behaviours also varied by driving mode; drivers were generally more likely to maintain a hands-on approach when using FSD Beta compared to Autopilot. From a tracing perspective, although drivers appeared to understand the requirement to keep their hands on the steering wheel, their actual behaviours demonstrated considerable variation.

3.2.2. Knowledge: staying alert

According to the vehicle manufacturer, drivers are required to remain alert at all times. Based on this requirement, we identified three subcategories of participant responses:

- **Awareness of surrounding situations**
  Participants expressed varied perspectives regarding situational awareness. Some reported that using Autopilot and FSD Beta enhanced their attentiveness, allowing them to focus further down the road and experience reduced fatigue. One participant noted experiencing heightened alertness while using the technology, citing improved contextual awareness and the ability to more effectively scan the driving environment. Another emphasised the importance of remaining fully attentive, particularly when using the beta version, to stay aware of the surrounding conditions.

  “Autopilot and FSD beta allow you to actually be more attentive in general then not having autopilot or FSD, and that’s because the car is taking care of the rudimentary things for you.. That allows you to focus further down on the road or it allows you to see things that maybe you wouldn’t have seen otherwise and it allows you to be less fatigued to where you’re able to. You’re able to be more alert than you would be otherwise. That doesn’t mean that you don’t also get distracted at times, but I think when you are paying attention, I think it allows you to pay better attention to the road than without autopilot or FSD. (R027)”

"For Autopilot and FSD always be alert and attentive. If it's a beta, it's required to be fully attentive and alert at all times. Autopilot. I know other owners, they're kind of relaxing, not paying attention. For me on Autopilot, it helps me become more attentive of my surroundings during driving.. When I'm driving myself, I usually look forward in once in a while, look left and right, but with autopilot, I'm able to watch.. all the mirrors all the time, making sure what I'm aware of what's going on around me. (R075)"

"Typically, fully attentive and alert at all times. Pretty excessively alert. As one of the things I love about the beta and regular autopilot as well, when they drive it, actually more aware because I can actually look around and take in.. where all the cars are around me. I understand.. what's going on, where it was.. I definitely enjoy it more when I'm not micromanaging those things and I'm able to take in and be more contextually aware. (R085)"

- **Highway**
  Participants' experiences with staying alert while driving varied depending on the driving context. In particular, some reported reduced attentiveness when using Autopilot on highways—especially on familiar routes, during low-traffic conditions, or in the absence of external distractions. The following quotations illustrate this variability in driver alertness:

  "And typically fully attentive and alert, you have to be.. Autopilot ..not so much.. I've noticed.. I'll be driving along and.. be able to read a sign along the road or something that. Before, you.. you wouldn't take the time to read and add. But.. if it's on a Interstate highway, it's no problem. (R010)"

  "I typically fully attentive and alert at all times? No, I've gotten comfortable with it over time, so I don't fully pay a attention anymore, specially on roads that I'm familiar with or on highways that I'm familiar with. (R016)"

  "I would say there are moments.. where I haven't been fully attentive. I obviously don't let that happen for like minutes.. I'm not gonna pull my phone out, look at it, but there's definitely times when driving on the highway, I look ahead and there's nobody for a kilometer ahead of me. And so I will look after the side and look at something in the scenery and then look back again or look at the passenger beside me and then look back again. Not for long periods of time. But longer than you could get away with if you were actually the one driving, I would say. (R045)"

- **Driving mode**
  Participants reported variations in their levels of attentiveness and alertness depending on the driving mode. One participant noted that their awareness was lower when using Autopilot compared to FSD Beta, even when feeling fatigued. Another acknowledged being less attentive in Autopilot mode, explaining that it enabled multitasking behaviours that would not be possible during conventional driving. A third participant stated that their level of attentiveness while using Autopilot was slightly lower than when using FSD Beta. The following quotations offer further insight into these reported differences:

  "Are you typically fully attentive and alert at all times? I'd say with Autopilot I have been in situations where I've driven really exhausted and I tend to have pretty good situational awareness even when I'm.. super exhausted. But I would say like.. when I use autopilot, it's not always 100 percent peak performance.. With FSD beta.. I'm always alert and fully attentive.(R051)"

  ".. If I were to grade these on how I feel on where I have to be fully attentive and alert at all times, FSD beta requires the most, Autopilot requires less (R063)"

  "..typically fully attentive, fully alert.. at all times.. Less.. in autopilot.. I would say that Autopilot does allow you to do other things that you might not normally do if you were driving the car normally. (R088)"

The evaluation of drivers' knowledge regarding the need to remain alert while driving revealed three key insights in relation to the tracing condition. First, some participants reported that Autopilot and FSD Beta enhanced their attentiveness by allowing them to focus further ahead and reduce fatigue. Second, attentiveness varied depending on the driving context; some drivers reported decreased focus on highways, particularly on familiar routes. Third, alertness differed across driving modes, with participants generally reporting lower levels of awareness while using Autopilot compared to FSD Beta.

Overall, although participants demonstrated knowledge of the requirement to stay alert—thus formally satisfying the tracing condition—their actual behaviours reflected variability in alertness depending on context and system use.

3.2.3. Capacity: corrective action

The tracing condition of MHC requires that drivers not only understand the functionality of partially automated systems but also retain the capacity to operate them effectively. Consistent with this requirement, the vehicle manufacturer in our study mandates that drivers must be able to perform corrective actions. Based on our analysis of participant responses, we identified three sub-categories reflecting this capacity.

- **Control over the steering wheel**
  Participants demonstrated readiness to take corrective action through their hand positioning while using automated driving features. One participant emphasised maintaining a firm grip on the steering wheel, deliberately placing their hands in the lower corner to enable a rapid response to unexpected lane drift. Another respondent noted that their approach to hand positioning was influenced by how Autopilot handled specific driving situations. Additionally, one participant reported that resting one hand on the wheel was ineffective for performing minor corrective actions when using Autopilot.

  "Often when I'm driving on the highway.. if it's just me, I'll just have one hand resting on top of the wheel, making minor corrective action. But that doesn't work very well with Autopilot. It thinks that I'm not touching my car so. So I yeah, do like the nine and three or five and seven to use gravity. (R028)"

  "Do I prepare to take corrective actions? Absolutely, whether it's holding that steering wheel really hard in case it wants to just drift off really quick or.. really hard.. that's.. why I hold my hands. The way they're steering wheels made, that's also why I hold my hands in that lower corner as opposed to up top when you hold it up top. If the car is gonna jerk itself off to the right, especially being left handed, it can only go so far before the centre beam and the steering wheel will block it. But if you hold it on the bottoms, it has much less travelled before you can get your hand on it. And if anything else, it's going to stop as soon as it hits you hit one of that centre peg.(R061)"

  ".. Stay prepared to take corrective actions, like more.. than if I was just tracking upon myself. Because technically there's someone else driving the.. car, you know that they're not very good at driving the car, so I have to pay more attention.. I guess I'm pretty good spatial awareness, so I take a lot.. for granted.. in terms how you place your hands on the steering wheel. (R081)"

- **Control over the pedals**
  Participants described their interactions with the brake pedal as a means of demonstrating their readiness to perform corrective actions. One participant noted the importance of being prepared to intervene, particularly in situations where FSD or Autopilot might fail to brake in time. Another explained that they positioned their foot in a comfortable location to enable rapid braking or acceleration when necessary. A third participant emphasised the ease with which they could engage the brake, as their foot was already positioned similarly to when operating vehicles with lower levels of automation.

  "Do you typically stay prepared to make corrective actions at all time? Absolutely, especially with FSD, better you have to be prepared. You have to kind of.. exit plan. If it comes down to it with Autopilot, not as much. But there are times where you may have to be ready to press the brakes because the car is not breaking in time and it's getting a little bit too close to the car ahead of you. (R064)"

  "I always stay prepared to take corrective action with FSD beta. But with autopilot on the highway.. feel a little bit more comfortable with my foot. Like to the side.. But it is really easy to lift my foot and hit the brake if I need to.. I would say like it's pretty much the same as when I use cruise control on older cars or other cars in the past. It's the same place. I would put my foot. (R074)"

  "Do you typically stay prepared to take corrective actions at all times? Mostly I keep my feet just like back. I'll wait from the pedals. Just getting a comfortable position unless.. a location where I don't have as much trust in Autopilot, FSD beta, or there's a lot of cars around me. Then I have my feet ready to like, break or accelerate or anything like that mostly depends on the situation. (R078)"

- **Driving mode**
  Participants' perspectives on their preparedness to take corrective action varied depending on the driving mode. One respondent strongly emphasised the importance of maintaining vigilance while using FSD Beta, describing a constant state of readiness to intervene. Another participant highlighted a contrast between the two modes, reporting a heightened level of awareness—described as being "hyper-aware"—when using FSD Beta, and a lower level of preparedness when using Autopilot. Additionally, some participants indicated that their trust in Autopilot increased over time, leading to a more relaxed driving posture and a perception of the system as being safer. The following quotations illustrate these perspectives:

  "Do you typically stay prepared to take corrective action at all times? Autopilots a little bit less than FSD beta. As long as I'm in a comfortable realm, there's no situations around me. I am prepared, but my guards down a little bit more. With FSD beta. I'm always ready to take over. (R017)"

  "Do you typically stay prepared to take corrective actions at all times? I certainly do that.. for all the reasons.. But autopilot, I feel like I'm typically less prepared because I'm more relaxed. I'm more letting my guard down. Because.. I trust it more. It's never done as much wrong as.. I'm looking at the scenery I'm looking.. I'm enjoying the ride versus driving pretty much. So definitely less prepared on autopilot.. Autopilot is safer in my opinion. (R048)"

  "Next question that typically stay prepared to take corrective action. Autopilot.. not so much. I mean, my hand is on the wheel.. With Beta. I'm very.. ready to take control. It's hyper aware. (R096)"

The evaluation of drivers' capacity to perform corrective actions—another key component of the tracing condition—revealed that participants demonstrated readiness by adjusting their grip on the steering wheel and maintaining their foot near the brake pedal. Several drivers also adopted specific strategies to enable rapid intervention when necessary. Preparedness varied by driving mode: participants reported feeling more vigilant while using FSD Beta and more relaxed when using Autopilot. Overall, the findings suggest that drivers exhibited different levels of readiness to take corrective action, influenced by both the automation mode and their individual driving strategies.

3.2.4. Maintaining operational responsibility

The final tracing requirement concerns the maintenance of operational responsibility. According to the tracing condition, at least one human must be aware that they hold moral responsibility for the outcomes of the system's actions. This aligns with the vehicle manufacturer's guidance, which stipulates that drivers are accountable for the operation of the partially automated driving systems.

Unlike the other tracing evaluation criteria, only one subcategory was identified in this area: agreement among participants that drivers must maintain control and assume responsibility.

Participants expressed this recognition in varying ways. One respondent mentioned feeling personally responsible for ensuring that the vehicle did not make errors. Another indicated a strong belief that, in the event of an incident, they would be held fully liable and could not shift blame to Autopilot in a legal context. A third respondent explicitly emphasised the importance of maintaining control and responsibility, acknowledging that they would accept fault in the case of a collision. The following quotations provide illustrative examples:

"The responsibility is definitely mine,. I wrecked my car.. not tesla fault.. indeed. (R032)"

Did not maintain control in this? No, I disagree with that. I mean..I get that I'm completely responsible for it. I'm gonna lose in court if I say Autopilot made me did it, or autopilot did it.(R067)

"I'm paying attention to what it's doing, backing it up to make sure it doesn't make a mistake.. But.. if it does, I'm responsible for it. So I have to be really paying attention to it. So I'm vigilant. But.. I feel like probably secure that it's doing a good job. (R072)"

Overall, the evaluation indicates that drivers are aware of their responsibility to oversee the vehicle's operation and recognise their accountability in the event of system errors or legal consequences.

4. Discussion

4.1. Theoretical implications

This section discusses how the proposed MHC evaluation framework can be applied to systems based on real-world driving experiences, offering new insights into the dynamic nature of meaningful human control (MHC), particularly in relation to the tracking and tracing components. The findings highlight the interplay between system performance and human factors, contributing to the existing body of literature by emphasising the roles of contextual variability, subjective risk perception, and the interaction between human engagement and system behaviour in the assessment of MHC.

The tracking evaluation revealed notable variations in how different safety features align with both human- and manufacturer-defined safety expectations across varying driving contexts. Features such as Blind Spot Monitoring (BSM) and Lane Departure Avoidance (LDA) demonstrated strong alignment with the tracking component of MHC during routine scenarios, such as highway lane-keeping. Drivers particularly valued BSM's warning system and visual interface for identifying vehicles in blind spots, consistent with findings from Kim et al. (2024), who reported that user interfaces offering surrounding information enhance driver trust and reduce perceived risk.

However, in emergency or unexpected driving situations—such as encounters with sudden obstacles—features like Automatic Emergency Braking (AEB) and Forward/Side Collision Warning (F/SCW) exhibited performance inconsistencies. Although drivers acknowledged their effectiveness in preventing collisions, these systems were less reliable in consistently meeting the tracking requirements of MHC. This observation aligns with Cicchino (2017), who found that such features significantly reduce front-to-rear crash rates but are not universally effective. These results underscore the importance of ensuring that partially automated systems are capable of dynamically adapting to diverse and unpredictable driving environments in order to uphold meaningful human control.

Misalignment between driver and manufacturer safety expectations often arises from technological limitations—such as sensor failures in adverse weather conditions, including bright light impairing sensor performance—or from mismatched expectations, such as drivers perceiving AEB braking as overly hesitant. Tesla's manual (Tesla, 2024) explicitly acknowledges limitations such as obscured lane markings or weather-related interference (e.g., rain); nevertheless, drivers expressed safety concerns when these limitations manifested in practice. These findings underscore the importance of addressing root causes, including the enhancement of sensor reliability and better alignment of system behaviour with human expectations.

In addition, drivers' risk perception played a critical role in the tracking component of MHC. False positives, such as phantom braking, diminished trust in the system, whereas successful interventions, such as timely collision avoidance, improved perceived safety. This dynamic highlights the need for the tracking component of MHC to account for both objective system performance and the subjective experiences of human drivers.

The tracing evaluation reveals how human factors shape the effectiveness of partially automated systems in meeting tracing criteria. Drivers frequently engaged selectively with system warnings or interventions based on their personal risk assessments. For instance, some participants reported disregarding Forward/Side Collision Warning (F/SCW) alerts when they judged the following vehicle to be at a safe distance, indicating a disconnect between system logic and human judgement.

The paradox of trust also emerged as a critical influence on tracing compliance. While drivers expressed appreciation for features like LDA for reducing cognitive load—consistent with findings by Miller and Boyle (2019), who demonstrated increased workload in the absence of LDA—over-reliance on such features often resulted in complacency. This supports the argument of Bainbridge (1983), who described the "ironies of automation," wherein human vigilance diminishes as system reliability increases. Young and Stanton (2002) further conceptualise this effect through "mental underload," where reduced task demands lower attentional capacity and compromise readiness to intervene.

Although manufacturers attempt to mitigate this risk by assigning drivers the responsibility to remain engaged, in practice, drivers often disengage during routine operation, becoming "out-of-the-loop" (Endsley, 2017). This challenge is exacerbated by system design approaches that overlook human cognitive limitations in sustained attention and monitoring tasks (Lee & See, 2004). The resulting paradox exposes a fundamental design flaw: assigning moral responsibility alone is insufficient to guarantee continuous vigilance. As argued by Hansson et al. (2021), systems that promote over-reliance while simultaneously expecting uninterrupted human supervision raise significant ethical concerns.

The level of driver engagement was found to vary depending on system behaviour and driving context. Participants tended to be more engaged in complex or high-risk driving scenarios, while disengagement was more common during routine tasks. This dynamic aligns with findings by Robins-Early (2024) and Oskina et al. (2023), who demonstrated that subjective risk perceptions—such as preferences for lateral distance during automated overtaking—significantly influence trust and perceived safety. For instance, some drivers reported experiencing stress when LDA maintained a minimal lateral buffer, even if the manoeuvre was technically safe.

Personal preferences, driving modes, and situational contexts also played a significant role in tracing performance. Drivers were more likely to comply with hand placement requirements in urban environments when using FSD Beta—perceived as riskier—while adopting minimal contact strategies (e.g., resting a hand or applying intermittent torque) on highways with Autopilot, which was perceived as more stable. Additionally, some participants admitted to deliberately manipulating the system by applying weight to the steering wheel to simulate compliance with hand detection requirements.

This pattern of selective adherence highlights the complex interplay between individual attitudes, perceived risk, and contextual factors. It suggests that tracing performance cannot be fully understood without accounting for how drivers interpret and respond to system cues within specific driving environments.

4.2. Practical implications

This section offers practical recommendations based on the insights obtained from the MHC evaluation, with a focus on enhancing system design and addressing subjective driver experiences.

To improve system design, it is essential to address environmental limitations such as glare from sunlight, adverse weather conditions, and faded lane markings, all of which can impair system reliability. For example, Blind Spot Monitoring (BSM) sensors that are susceptible to sunlight interference could be redesigned using alternative sensing technologies or with added redundancy to ensure consistent performance. Minimising false positives and false negatives—such as phantom braking or missed hazard detections—is also critical for sustaining driver trust. Potential solutions include refining object detection algorithms and integrating contextual awareness to reduce unnecessary alerts. These design improvements are urgent, particularly in light of findings by Paula et al. (2023), who reported that 78% of drivers were unable to override phantom braking, thereby heightening safety risks.

Moreover, system behaviour should be calibrated to align more closely with human expectations of safety and trust. For instance, Automatic Emergency Braking (AEB) could be adjusted to engage earlier in emergency scenarios, reflecting drivers' preferences for proactive intervention. This recommendation is supported by Koglbauer et al. (2018), who demonstrated that braking behaviour significantly influences perceived safety.

Addressing subjective driver experiences is equally critical for improving the effectiveness of partially automated systems. Clear and intuitive user interfaces can enhance driver understanding of system behaviour. For example, visual or auditory cues explaining why a warning was issued or why braking occurred could reduce confusion and strengthen trust. Encouraging driver engagement is also essential. Systems should actively prompt drivers to assume control in edge cases—such as when lane markings are unclear—to mitigate over-reliance on automation. Furthermore, educating drivers about system capabilities and limitations remains a key priority. Comprehensive training programmes can support more effective use of automation by emphasising the importance of staying attentive and prepared to intervene.

To address the broader challenges of driver engagement and over-reliance on automation, we propose several actionable recommendations. First, adaptive Human-Machine Interfaces (HMIs) could be developed to tailor hand placement reminders based on the driving context. For instance, stricter prompts may be warranted on highways, where automation is typically perceived as reliable, whereas fewer prompts may be appropriate in urban settings, where drivers are more naturally engaged due to increased perceived risk.

Second, enhanced training and regulatory measures are essential to reinforce driver readiness. Scenario-based training modules could prepare drivers to respond effectively in low-risk contexts, while policies mandating multi-layered engagement checks—extending beyond easily circumvented measures such as steering torque detection—would promote sustained vigilance.

Finally, reassessing preparedness expectations is crucial. Given the natural constraints of human attention and the observed tendency to over-rely on automation, vehicle manufacturers should reconsider assumptions about how quickly and effectively drivers can retake control. By incorporating these recommendations, automated driving systems can better align with human capabilities and limitations, thereby ensuring that drivers remain meaningfully engaged and ready to intervene when necessary.

4.3. Are Tesla's partially automated driving systems under meaningful human control?

Based on our evaluation of **tracking** and **tracing compliance**, we conclude that Tesla's FSD Beta and Autopilot systems do not fully satisfy the requirements of meaningful human control (MHC). In contrast to previous studies that primarily assess MHC through hypothetical scenarios or post-incident analyses (Calvert et al., 2021, 2020), our evaluation provides a more nuanced understanding of real-world system behaviour and its implications for MHC compliance. Below, we summarise the key findings that support this conclusion.

4.3.1. Failures in tracking compliance

Tesla's systems frequently failed to track safety expectations under challenging environmental conditions and in the presence of degraded infrastructure. These shortcomings underscore a lack of robustness in the perception systems, which are essential for maintaining alignment with both driver and manufacturer safety expectations. For example, adverse weather conditions—such as rain, snow, or glare—were reported to impair sensor functionality, resulting in failures to detect obstacles or maintain appropriate lane positioning. Similarly, infrastructure-related issues, including faded lane markings and poorly maintained roads, further exacerbated these limitations, as the systems rely heavily on visual inputs for accurate operation.

In high-risk or unpredictable scenarios, features such as Automatic Emergency Braking (AEB) and Forward/Side Collision Warning (F/SCW) often struggled to effectively track safety expectations. Issues such as phantom braking—where the system erroneously detects obstacles and applies the brakes unnecessarily—and false negatives—where genuine hazards go undetected—further compromised performance. These inconsistencies not only eroded driver trust but also diminished the system's capacity to meet safety expectations in critical situations.

Moreover, even when systems conformed to technical specifications, such as adhering to predefined braking thresholds, they frequently failed to meet driver expectations. Subjective perceptions of safety often diverged from objective system performance. For instance, participants described AEB interventions as overly cautious or hesitant, despite the system functioning within its intended parameters.

4.3.2. Failures in tracing compliance

Failures in tracing compliance stem from inconsistent driver adherence to safety protocols, over-reliance on automation in low-risk scenarios, and systemic design shortcomings that inadvertently promote disengagement. Drivers frequently demonstrated selective adherence to recommended behaviours, such as maintaining hands on the steering wheel and staying alert. Higher compliance was observed in high-risk contexts—such as urban environments with FSD Beta—where the perceived complexity of the driving environment prompted greater vigilance. Conversely, in low-risk contexts such as highway driving with Autopilot, compliance levels declined substantially as drivers placed greater trust in the system's reliability.

This variability indicates a troubling dependence on driver confidence rather than on system robustness to ensure safe operation. It also reveals a fundamental challenge in tracing compliance: current systems often fail to support sustained driver engagement and accountability, especially during routine or low-demand driving conditions.

An inverse relationship was observed between driver confidence and preparedness to perform corrective actions. When drivers perceived the system to be safe—such as during routine highway driving with Autopilot—their vigilance and readiness to intervene declined. This over-reliance on automation introduces significant risk, as drivers may be insufficiently prepared to take control in emergency situations, thereby undermining the system's ability to maintain meaningful human control.

Although participants generally acknowledged their moral responsibility for overseeing the system's operation, misuse of automation features was common. For example, several drivers reported circumventing safety protocols by applying weight to the steering wheel to simulate hand presence. This behaviour reveals a deeper structural issue: moral responsibility alone is insufficient to guarantee adherence to safety protocols. Current system designs, which rely predominantly on basic compliance checks such as steering torque verification, inadvertently facilitate complacency and improper use.

To address this challenge, it is not enough simply to remind drivers of their responsibilities. Instead, automated systems must be proactively designed to promote continuous driver engagement and situational awareness. This includes implementing more robust human–machine interaction strategies that help ensure drivers remain alert and ready to assume control when necessary.

4.4. Limitations

Despite the insights gained from our research, several limitations may impact the interpretation and generalisability of our findings. First, the data used primarily reflect drivers' subjective perceptions. While these perceptions are valuable for understanding user experiences, they may not always correspond to actual driving conditions. To address this limitation, we recommend that future research complement subjective reports with objective data (e.g., telemetry or kinematic data), allowing for statistical analysis that can more robustly assess system performance and user interaction.

Second, although all participants used Tesla's Autopilot or FSD Beta—both classified as SAE Level 2 systems—the original data collection did not ask participants to identify the automation level of their vehicle. While this could be seen as a limitation, we argue that it does not compromise the validity of our findings for three reasons: (1) users typically interpret automation through system behaviour and driver responsibility, rather than formal SAE terms; (2) knowledge-related questions in the interviews revealed that participants generally understood system limitations and the need for supervision (Nordhoff & Hagenzieker, 2024); and (3) Tesla communicates these limitations clearly through system prompts and manuals. Nonetheless, we recommend that future studies explicitly examine users' awareness of automation classifications or mental models, particularly where this may influence trust, expectations, or driver behaviour.

Third, our study is constrained by the fact that we only considered drivers and vehicle manufacturers as human agents in the evaluation of meaningful human control (MHC). However, it is important to acknowledge that other stakeholders—such as other road users, members of the public, lawmakers, and government authorities—also play a significant role in the operation, deployment, and governance of automated driving systems (Calvert & Mecacci, 2020). These stakeholders contribute to the broader sociotechnical context in which automated systems function. We therefore recommend that future research broaden the scope of analysis by including additional human agents to provide a more holistic evaluation of MHC.

Fourth, our evaluation of human expectations was primarily limited to safety. While safety remains a central concern in the context of driving automation, other expectations—such as comfort, regulatory compliance, and time efficiency—are also relevant. Future studies should incorporate a wider range of human expectations to enable a more comprehensive understanding of how MHC is established and maintained in partially automated systems.

Fifth, our findings are based exclusively on data collected from users of Tesla's FSD Beta programme in the United States and Canada. Variations in the design, implementation, and user interfaces of automated driving systems across manufacturers may lead to different user experiences and perceptions. Consequently, we recommend that future research include a more diverse sample of both automakers and participants to enhance the representativeness and generalisability of findings related to meaningful human control.

Sixth, while the dataset used in this study offers valuable insights into driver interactions with Tesla's Autopilot and FSD Beta systems, several potential biases should be acknowledged. Selection bias may have influenced the sample, as participants were recruited exclusively through online platforms. This recruitment method may have excluded individuals who are not active on such platforms, potentially leading to the underrepresentation of certain demographic groups. As a result, the sample may not fully reflect the diversity of all users of these systems. In addition, response bias may have affected the quality of interview data. Given the remote nature of the interviews (e.g., conducted via Zoom), participants may have tailored their responses to align with perceived social norms or desirability. These potential biases should be considered when interpreting the findings of this study.

Seventh, as a qualitative study, the analysis is subject to potential researcher bias. Although the study employed structured analytical frameworks—such as inductive category development and the application of ad-SOTIF and PST assessments—the interpretation of participant responses necessarily involved a degree of researcher judgement. To enhance objectivity, future studies could incorporate inter-coder validation or triangulation methods to strengthen the reliability and transparency of qualitative coding processes.

Eighth, the retrospective and indirect nature of our evaluation of MHC poses an inherent methodological limitation. The dataset was initially collected without explicitly introducing the MHC framework or assessing participants' understanding of its core components—namely, tracking and tracing. As a result, we were unable to directly evaluate participants' awareness, interpretation, or valuation of MHC as a concept. While our indirect approach yielded contextually relevant and theoretically grounded insights, it was not originally designed with the explicit goal of evaluating MHC. Future research should therefore be conducted with the explicit intent to assess MHC—meaning that while similar questions might be asked, they would be purposefully framed within the MHC framework. This would allow for more focused interpretation, targeted measurement, and potentially more valid conclusions about how users understand and experience meaningful human control in automated driving contexts.

5. Conclusion

Evaluating meaningful human control (MHC) over partially automated driving systems presents considerable challenges, stemming from the complex interactions between human drivers and automation, as well as the variability inherent in real-world driving contexts. This study offers a systematic assessment of how such systems adhere to MHC principles beyond post-incident analysis and hypothetical scenarios, by focusing on their operation in everyday, real-world situations.

The contributions of this study are twofold. First, we evaluated the extent to which partially automated driving systems in real-world contexts comply with the requirements of MHC. Second, we introduced a novel methodological approach for assessing the tracking and tracing dimensions of MHC, using qualitative data derived from in-depth interviews with users of Tesla's Autopilot and FSD Beta systems. This approach provides richer insights into drivers' lived experiences with automation and offers a practical framework for examining MHC in automated systems.

We evaluated tracking based on how consistently various safety features performed their intended functions, alongside drivers' perceptions of safety and trust. Tracing was assessed through drivers' knowledge of the requirement to keep both hands on the steering wheel and remain alert, their capacity to execute corrective actions, and their awareness of moral responsibility for the system's operation. By applying this evaluation framework, we found that while subsystems of Tesla's FSD Beta and Autopilot demonstrate partial adherence to the principles of meaningful human control, several significant challenges remain. These include inconsistencies in tracking both driver and manufacturer safety expectations, as evidenced by the comparatively weaker tracking performance of

F/SCW and AEB relative to BSM and LDA. Such issues are frequently linked to technological limitations—such as false positives, false negatives, and sensor vulnerabilities under adverse environmental conditions—as well as misaligned user expectations.

Inconsistencies in MHC also arise from variability in driver interaction, including selective adherence to safety protocols, over-reliance on automation, and misuse of system features. For example, adherence to guidelines—such as maintaining hand contact with the steering wheel and staying alert—was found to be inconsistent and often shaped by perceived risk. Drivers exhibited greater caution with FSD Beta in urban environments compared to the more relaxed use of Autopilot on highways.

These findings highlight the urgent need for further technological development, user-centred design improvements, and regulatory attention to ensure stronger and more consistent meaningful human control in partially automated driving systems.

CRediT authorship contribution statement

Lucas Elbert Suryana: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sina Nordhoff:** Writing – review & editing, Supervision, Resources, Methodology, Data curation, Conceptualization. **Simeon Calvert:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Arkady Zgonnikov:** Writing – review & editing, Supervision. **Bart van Arem:** Writing – review & editing, Supervision.

Acknowledgements

The data used in this research were originally collected by the second author to develop a conceptual framework for automation disengagements (Nordhoff, 2024). They were subsequently reused in follow-up publications examining the misuse of standard Autopilot and FSD Beta systems (Nordhoff et al., 2023), as well as the alignment between drivers' perceptions and the extent to which automated driving systems meet those expectations (Suryana et al., 2024). In the present study, we employed the dataset to investigate a distinct research question: to what extent do partially automated driving systems in real-world settings comply with the tracking and tracing conditions of meaningful human control (Santoni de Sio & Van den Hoven, 2018).

We acknowledge the use of OpenAI ChatGPT and DeepSeek to enhance the clarity and conciseness of the manuscript and to proofread it; the original text was entirely authored by the researchers. This work was supported by the Indonesia Endowment Fund for Education (LPDP) under Grant No. 0006552/TRA/D/19/lpdp2021.

Appendix A. Questionnaire

See Table A.5.

Table A.5
Interview questions.

Question number	Question
Q1	Do you have the Full Self-Driving Beta (FSD Beta) feature? (1 = Yes, 2 = No)
Q2	Before the first time of using Autopilot and FSD Beta, did you watch / read / listen to information on how to use it? (1 = Yes, 2 = No)
Q3	Please mention the type of information you consulted on how to use Autopilot and FSD Beta (website of Tesla (www.tesla.com), car dealer / sales point, online communities and forums, YouTube videos, newspapers and magazines, friends, family, colleagues, driver manual)
Q4	Please describe your experience with using Autopilot and FSD Beta and the benefits and risks associated with using it. Please explain your answer
Q5	Have your expectations of using Autopilot and FSD Beta been fulfilled? Why / why not?
Q6	Why do you use Autopilot and FSD Beta?
Q7	Did you ever stop using Autopilot and FSD Beta (for prolonged periods of time)?
Next, we would like to explore your perceptions regarding four general statements about the operation of Autopilot and FSD Beta	
Q8	The current Autopilot does make driving autonomous. Is that correct? (1 = Yes, 2 = No, 3 = I don't know)
Q9	There are no safety issues with Autopilot. Is that correct? (1 = Yes, 2 = No, 3 = I don't know)
Q10	Autopilot is a hands-on feature. Is that correct? (1 = Yes, 2 = No, 3 = I don't know)
Q11	Tesla FSD Beta is safer than a human. Is that correct? (1 = Yes, 2 = No, 3 = I don't know)
With the next section, we would like to explore your perceptions of safety while using Autopilot and FSD Beta	
Q12	Do you feel safe when Autopilot and FSD Beta is active? Why / why not?
Q13	What / how do you feel when you feel safe / unsafe? Please explain
Q14	What is it about Autopilot and FSD Beta that is safe / unsafe? Please explain.
Q15	Now please remember the situation / s in which you typically feel unsafe when Autopilot and FSD Beta is active and describe these situations.
Q16	What can Autopilot and FSD Beta do to support your safety in Autopilot and FSD Beta? Please explain
Q17	Does feeling safe / feeling unsafe impact how you use Autopilot and FSD Beta on your next drives / in the future? Please explain.
Q18	Has your perceived safety changed over time? If so, how?

Table A.5 (*continued*)

Question number	Question
With the next section, we would like to explore your trust in Autopilot and FSD Beta.	
Q19	How would you position your level of trust in Autopilot and FSD Beta. (1 = I don't trust it at all, 2 = I don't trust it, 3 = I neither don't trust it at all nor trust it a lot, 4 = I trust it, 5 = I trust it a lot)
Q20	What can Autopilot and FSD Beta do to support your trust in Autopilot and FSD Beta?
Q21	Does your trust / distrust in Autopilot and FSD Beta impact how you use Autopilot and FSD Beta on your next drives / in the future? Please explain.
Q22	Has your trust changed over time? If so, how?
Q23	When you do compare yourself with other drivers, Autopilot, and FSD Beta, do you think you are ... (1 = A much worse driver, 2 = A worse driver, 3 = Not a better nor a worse driver, 4 = A better driver, 5 = A much better driver) (De Craen, 2010)
With the next section, we would like to explore how you typically use Autopilot and FSD Beta.	
Q24	How do you typically place your hands on the steering wheel when Autopilot and FSD Beta is active? Please select the image that serves as the best representation of your placement of your hands on the steering wheel when Autopilot / FSD Beta is active and explain your answer.
Q25	Do you typically keep your hands on the steering wheel at all times?
Q26	Are you typically fully attentive and alert at all times?
Q27	How often do you typically engage in other secondary activities while Autopilot and FSD Beta is active? (Never, rarely, occasionally, frequently, always; monitoring the road ahead, talking to fellow travellers, observing the landscape, using the phone for music selection, using the phone for navigation, using the phone for calls, eating and drinking, using the phone for texting, watching videos / TV shows, sleeping)
Q28	Do you disengage Autopilot and FSD Beta? Why / why not?
Q29	Does Autopilot and FSD Beta disengage? When / in which situations?
Q30	How do you typically place your eyes when Autopilot and FSD Beta is active?
Q31	Do you typically keep your eyes on the road at all times?
Q32	Do you typically monitor the vehicle and its surroundings at all times?
Q33	How do you typically place your feet when Autopilot and FSD Beta is active?
Q34	Do you typically stay prepared to take corrective actions at all times?
Q35	Has your use of Autopilot (in terms of how you placed your hands on the steering wheel, eyes on the road, and feet) changed over time? If so, how?

Data availability

Data will be made available on request.

References

Awad, E., Levine, S., Kleiman-Weiner, M., Dsouza, S., Tenenbaum, J. B., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2020). Drivers are blamed more than their automated cars when both make mistakes. *Nature Human Behaviour*, *4*, 134–143. https://doi.org/10.1038/s41562-019-0762-8.

Bainbridge, L. (1983). Ironies of automation. In *Analysis, design and evaluation of man–machine systems* (pp. 129–135). Elsevier.

Banks, G. C., Woznyj, H. M., Wesslen, R. S., & Ross, R. L. (2018a). A review of best practice recommendations for text analysis in R (and a user-friendly app). *Journal of Business and Psychology*, *33*, 445–459. https://doi.org/10.1007/s10869-017-9528-3.

Banks, V. A., Eriksson, A., O'Donoghue, J., & Stanton, N. A. (2018b). Is partially automated driving a bad idea? Observations from an on-road study. *Applied Ergonomics*, *68*, 138–145. https://doi.org/10.1007/s10869-017-9528-3.

Beckers, N., Siebert, L. C., Bruijnes, M., Jonker, C., & Abbink, D. (2022). Drivers of partially automated vehicles are blamed for crashes that they cannot reasonably avoid. *Scientific Reports*, *12*, Article 16193. https://doi.org/10.1038/s41598-022-19876-0.

Cabrall, C. D., Eriksson, A., Dreger, F., Happee, R., & de Winter, J. (2019). How to keep drivers engaged while supervising driving automation? A literature survey and categorisation of six solution areas. *Theoretical Issues in Ergonomics*, *20*, 332–365. https://doi.org/10.1080/1463922X.2018.1528484.

Calvert, S., Johnsen, S., & George, A. (2024). Designing automated vehicle and traffic systems towards meaningful human control. In *Research handbook on meaningful human control of artificial intelligence systems* (pp. 162–187). Edward Elgar Publishing.

Calvert, S. C., van Arem, B., Heikoop, D. D., Hagenzieker, M., Mecacci, G., & de Sio, F. S. (2020). Gaps in the control of automated vehicles on roads. *IEEE Intelligent Transportation Systems Magazine*, *13*, 146–153. https://doi.org/10.1109/MITS.2019.2926278.

Calvert, S. C., & Mecacci, G. (2020). A conceptual control system description of cooperative and automated driving in mixed urban traffic with meaningful human control for design and evaluation. *IEEE Open Journal of Intelligent Transportation Systems*, *1*, 147–158. https://doi.org/10.1109/OJITS.2020.3021461.

Calvert, S. C., Mecacci, G., Heikoop, D. D., & Janssen, R. (2021). How to ensure control of cooperative vehicle and truck platoons using meaningful human control. *European Journal of Transport and Infrastructure Research*, *21*, 95–119. https://doi.org/10.18757/ejtir.2021.21.2.5354.

Cavalcante Siebert, L., Lupetti, M. L., Aizenberg, E., Beckers, N., Zgonnikov, A., Veluwenkamp, H., Abbink, D., Giaccardi, E., Houben, G.-J., Jonker, C. M., et al. (2023). Meaningful human control: Actionable properties for ai system development. *AI and Ethics*, *3*, 241–255. https://doi.org/10.1007/s43681-022-00167-3.

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

Cicchino, J. B. (2017). Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates. *Accident Analysis and Prevention*, *99*, 142–152. https://doi.org/10.1016/j.aap.2016.11.009.

Connected Automated Driving (2024). Taxonomy of automated driving systems. https://www.connectedautomateddriving.eu/methodology/taxonomy/. (Accessed 2 February 2025).

Docherty, B. L. (2015). Mind the gap: The lack of accountability for killer robots. https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots. (Accessed 6 November 2024).

Endsley, M. R. (2017). Autonomous driving systems: A preliminary naturalistic study of the tesla model s. *Journal of Cognitive Engineering and Decision Making*, *11*, 225–238. https://doi.org/10.1177/1555343417695197.

George, A., Cavalcante Siebert, L., Abbink, D., & Zgonnikov, A. (2023). Feasible action-space reduction as a metric of causal responsibility in multi-agent spatial interactions. In *ECAI 2023* (pp. 819–826). IOS Press.

Hansson, S. O., Belin, M-Å., & Lundgren, B. (2021). Self-driving vehicles—an ethical overview. *Philosophy & Technology*, *34*, 1383–1408. https://doi.org/10.1007/s13347-021-00464-5.

Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text preprocessing for text mining in organizational research: Review and recommendations. *Organizational Research Methods*, *25*, 114–146. https://doi.org/10.1177/1094428120971683.

International Organization for Standardization (2019). *ISO/PAS 21448: Road vehicles - safety of the intended functionality*. Pre-standards document, Geneva, Switzerland.

Kim, S., He, X., van Egmond, R., & Happee, R. (2024). Designing user interfaces for partially automated vehicles: Effects of information and modality on trust and acceptance. *Transportation Research. Part F, Traffic Psychology and Behaviour*, *103*, 404–419. https://doi.org/10.1016/j.trf.2024.02.009.

Koglbauer, I., Holzinger, J., Eichberger, A., & Lex, C. (2018). Autonomous emergency braking systems adapted to snowy road conditions improve drivers' perceived safety and trust. *Traffic Injury Prevention*, *19*, 332–337. https://doi.org/10.1080/15389588.2017.1407411.

Kwik, J. (2022). A practicable operationalisation of meaningful human control. *Laws*, *11*, 43. https://doi.org/10.3390/laws11030043.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*, 50–80. https://doi.org/10.1518/hfes.46.1.50-30392.

Li, J., Zhao, X., Cho, M.-J., Ju, W., & Malle, B. F. (2016). From trolley to autonomous vehicle: Perceptions of responsibility and moral norms in traffic accidents with self-driving cars. Technical Report, SAE Technical Paper.

Martinho, A., Herber, N., Kroesen, M., & Chorus, C. (2021). Ethical issues in focus by the autonomous vehicles industry. *Transport Reviews*, *41*, 556–577. https://doi.org/10.1080/01441647.2020.1862355.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, *6*, 175–183. https://doi.org/10.1007/s10676-004-3422-1.

Mecacci, G., & Santoni de Sio, F. (2020). Meaningful human control as reason-responsiveness: The case of dual-mode vehicles. *Ethics and Information Technology*, *22*, 103–115. https://doi.org/10.1007/s10676-019-09519-w.

Miller, E. E., & Boyle, L. N. (2019). Behavioral adaptations to lane keeping systems: Effects of exposure and withdrawal. *Human Factors*, *61*, 152–164. https://doi.org/10.1177/0018720818800538.

National Transportation Safety Board (2017). *Collision between a car operating with automated vehicle control systems and a tractor-semitrailer truck near Williston*. Technical Report, Accident Report NTSB/HAR-17/02 PB2017-102600.

National Transportation Safety Board (2018). *Collision between a sport utility vehicle operating with partial driving automation and a crash attenuator*. Technical Report, Accident Report NTSB/HAR-20/01 PB2020-100112.

NLTK Project (2024). Natural language toolkit. https://www.nltk.org. (Accessed 19 October 2023).

Nordhoff, S. (2024). A conceptual framework for automation disengagements. *Scientific Reports*, *14*, 8654. https://doi.org/10.1038/s41598-024-57882-6.

Nordhoff, S., & Hagenzieker, M. (2024). "I will raise my hand and say 'I over-trust autopilot'. I use it too liberally"–drivers' reflections on their use of partial driving automation, trust, and perceived safety. *Transportation Research. Part F, Traffic Psychology and Behaviour*, *107*, 1105–1124. https://doi.org/10.1016/j.trf.2024.09.021.

Nordhoff, S., Lee, J. D., Calvert, S. C., Berge, S., Hagenzieker, M., & Happee, R. (2023). (Mis-) use of standard autopilot and full self-driving (FSD) Beta: Results from interviews with users of Tesla's FSD beta. *Frontiers in Psychology*, *14*, Article 1101520. https://doi.org/10.3389/fpsyg.2023.1101520.

Nordhoff, S., Stapel, J., He, X., Gentner, A., & Happee, R. (2021). Perceived safety and trust in sae level 2 partially automated cars: Results from an online questionnaire. *PLoS ONE*, *16*, Article e0260953. https://doi.org/10.1371/journal.pone.0260953.

Oskina, M., Farah, H., Morsink, P., Happee, R., & van Arem, B. (2023). Safety assessment of the interaction between an automated vehicle and a cyclist: A controlled field test. *Transportation Research Record*, *2677*, 1138–1149. https://doi.org/10.1177/03611981221112423.

Paula, D., Bauder, M., Pfeilschifter, C., Petermeier, F., Kubjatko, T., Böhm, K., Riener, A., & Schweiger, H.-G. (2023). Impact of partially automated driving functions on forensic accident reconstruction: A simulator study on driver reaction behavior in the event of a malfunctioning system behavior, https://doi.org/10.3390/s23249785.

Robbins, S. (2023). The many meanings of meaningful human control. *AI and Ethics*, 1–12. https://doi.org/10.1007/s43681-023-00320-6.

Robins-Early, N. (2024). Driver in fatal Texas crash was using Ford's auto driving system, officials say. https://www.theguardian.com/business/2024/apr/11/ford-mustang-mach-e-automated-blue-cruise-system-fatal-texas-crash. (Accessed 24 April 2024).

SAE International (2021). Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. Standard J3016_202104. https://doi.org/10.4271/J3016-202104.

Santoni de Sio, F., & Van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. *Frontiers in Robotics and AI*, *5*, 15. https://doi.org/10.3389/frobt.2018.00015.

Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology*, *34*, 1057–1084. https://doi.org/10.1007/s13347-021-00450-x.

Steen, M., van Diggelen, J., Timan, T., & van der Stap, N. (2023). Meaningful human control of drones: Exploring human–machine teaming, informed by four different ethical perspectives. *AI and Ethics*, *3*, 281–293. https://doi.org/10.1007/s43681-022-00168-2.

Suryana, L. E., Nordhoff, S., Calvert, S. C., Zgonnikov, A., & Van Arem, B. (2024). A meaningful human control perspective on user perception of partially automated driving systems: A case study of tesla users. In *2024 IEEE intelligent vehicles symposium (IV)* (pp. 409–416).

Tesla (2024). Autopilot and full self-driving capability. https://www.tesla.com/support/autopilot. (Accessed 19 October 2024).

Tesla (2024). Model 3 Owner's manual. https://www.tesla.com/ownersmanual/model3. (Accessed 19 August 2024).

Veluwenkamp, H. (2022). Reasons for meaningful human control. *Ethics and Information Technology*, *24*, 51. https://doi.org/10.1007/s10676-022-09673-8.

Watanabe, K., & Zhou, Y. (2022). Theory-driven analysis of large corpora: Semisupervised topic classification of the un speeches. *Social Science Computer Review*, *40*, 346–366. https://doi.org/10.1177/0894439320907027.

Wörle, J., & Metz, B. (2023). Misuse or abuse of automation? Exploring drivers' intentions to nap during automated driving. *Transportation Research. Part F, Traffic Psychology and Behaviour*, *99*, 460–472. https://doi.org/10.1016/j.trf.2023.10.023.

Xu, Z., Zhang, K., Min, H., Wang, Z., Zhao, X., & Liu, P. (2018). What drives people to accept automated vehicles? Findings from a field experiment. *Transportation Research. Part C, Emerging Technologies*, *95*, 320–334. https://doi.org/10.1016/j.trc.2018.07.024.

Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors*, *44*, 365–375. https://doi.org/10.1518/0018720024497709.