Investigating Contextual Variations in Explaining Plausible Narratives of Social Intention in Driving

A Literature Survey

Jang Hun Oh

Supervisor(s): Hayley Hung, Vitaliy Popov, Arthur Mercier

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
January 25, 2026

Name of the student: Jang Hun Oh
Final project course: CSE3000 Research Project
Thesis committee: Hayley Hung, Vitaliy Popov, Arthur Mercier, Ricardo Marroquim

Abstract

Intelligent systems in autonomous driving increasingly require the ability to infer social intentions to ensure safe and fluid interactions with human road users. However, current approaches typically frame this problem as objective trajectory prediction or fixed classification, ignoring the open-ended nature of human interpretation where a single physical behaviour can generate multiple plausible narratives. To address the gap between trajectory forecasting and narrative understanding, this research investigates how to systematically map the dimensions of variation in driving situations to the range of intention narratives they generate. A literature survey was conducted to distinguish between foundational human social norms and current algorithmic approaches. By integrating script theory with the 3Cs framework (Cues, Characteristics, Classes), this study developed a dimension extraction framework to analyse where objective observations diverge into subjective interpretations. Through comparative analysis of prototypical scenarios (lane merging and pedestrian negotiation), results revealed that current intelligent systems operate predominantly in geometric space, optimizing for physical feasibility, whereas human drivers operate in social space governed by normative scripts. The research concludes that narrative open-endedness is inversely proportional to the strength of physical and social constraints. That is, when constraints are weak, human internal scripts diverge from machine logic, leading to critical prediction errors. Consequently, future systems must shift from raw trajectory output to semantic narrative understanding to explicitly model this uncertainty and align machine reasoning with human expectations.

1 Introduction

1.1 Motivation

Nowadays, more intelligent systems appear in complex human social environments from hospital settings, to social gatherings and even in driving scenarios[12; 18; 42]. These intelligent systems require the ability to infer social intentions for safe and fluid interactions with the humans using those systems. This research focused on the domain of autonomous driving, a high-stakes environment where understanding the intent of other road users, such as a pedestrian near a cross walk or a driver changing their lane, can be the difference between a smooth manoeuvrer or a catastrophic collision.

In traditional papers, this problem had been framed as one of objective trajectory prediction. Specifically in pedestrian crossings, current approaches often relied on trajectory prediction models that treated intention as a fixed classification problem (e.g., will cross or will not cross)[18; 21; 32], ignoring the nuances of why an action was performed. However, this approach often ignored the intention of the human as intention is not a singular physical property, but a subjective narrative constructed by the observer based on their current situation and their own experiences.

For example in another situation when a car driver hesitates at an intersection, this could generate multiple equally plausible narratives of intent depending on the context and the surrounding driver's perspective. To a defensive driver, the hesitation might signal caution while to an aggressive driver, it might signal incompetence. These different interpretations are all equally valid narratives framed by the specific role and internal script of the observer. These divergent interpretations arise from what script theory, proposed by Fischer et al., [7] described as a mismatch between external scripts (the objective rules of the road) and internal scripts (the subjective expectations of the driver).

Currently, intelligent system communities lacked a systematic framework to map these variations[2; 12]. Without formalising how these dimensions (e.g., environmental cues) influenced perceived intention, designers could not effectively stress-test their models against ambiguity[2; 29]. By formalising these scenarios using script theory[7] and the 3Cs framework (Cues, Characteristics, Classes) proposed by Rauthmann et al.[29], this research aims to provide intelligent system designers with a structured roadmap. This benefits the community by moving beyond simple outcome prediction to a deeper narrative understanding, which is essential for creating robust systems that can negotiate open-ended social situations safely.

1.2 Research Questions

To address the gap between trajectory prediction and narrative understanding, the primary research question driving this study is:

> *How can we systematically map the dimensions of variation in prototypical driving situations to the range of plausible social intention narratives they generate?*

To answer this, the following secondary questions has been investigated:

- **SQ1 (Taxonomy):** What are the primary dimensions of variation that can be found in current literatures on social perception and intelligent systems that influence the perception of intention in the context of driving?
- **SQ2 (Framework):** How can external scripts and internal scripts be used to model the generation of divergent intention narratives in distinct driving case studies?
- **SQ3 (Analysis):** What is the relationship between the open-endedness or constraint level of a driving scenario and the degree of variation in plausible intention narratives?
- **SQ4 (Strategy):** Based on this analysis, what recommendations can be provided to intelligent systems designers for a research strategy that scales from constrained to unconstrained scenarios?

2 Theoretical Background

This section defines the socio-psychological concepts of the open-endedness in driving environments and situations, and

how script theory and the 3Cs framework can be used to explain the reasons for the open-endedness.

2.1 The Challenge of Open-Endedness

Traditional engineering approaches often attempt to model driving by listing a finite set of rules as motivated in [41]. However, social driving is defined by open-endedness. This concept refers to the breadth of possible ways a single physical behaviour can be explained by one or more intentions. Open-endedness arises when there is a mismatch between established norms and subjective interpretations, or when the environment is weakly regulated (e.g., an uncontrolled parking lot vs. a strictly controlled toll-gate). In these scenarios, a single set of physical cues can generate multiple equally plausible narratives.

2.2 Script Theory in Driving

To model the expectations that drive these narratives, the script theory proposed by Fischer et al[7] had been utilised. This theory proposed that humans navigate complex social situations by following mental sequences of expected behaviours for a given context. In the domain of driving, these scripts operate on two distinct levels:

- **External Scripts:** These represent the objective rules and physical constraints of a scenario that everyone in that scenario is familiar with and expects each other to follow. For example, traffic lights, lane markings and right-of-way laws.
- **Internal Scripts:** These represent the subjective mental model of the observer. One example of this could be a defensive driver versus an aggressive driver. Internal scripts filter the observation of the external event based on their own experience or driving style, leading to multiple interpretations of the same physical action.

2.3 The 3Cs Framework

To analyse how road users move from observation to interpretation, the 3Cs Framework (Cues, Characteristics, Classes) by Rauthmann et al.[29] has been adopted:

- **Cues:** Cues represent the physico-biological level of the situation. They are the objectively quantifiable elements physically present in the setting.Examples include surrounding vehicles, lanes, speed limit signs and engine noise. Cues serve as the raw input for the driver's perception to process to become meaningful.
- **Characteristics:** Characteristics capture the subjective-functional reality of the situation. They represent the psychological meaning ascribed to the cues by the observer. While cues describe the setting, characteristics describe the experience (e.g., identifying a driver as anxious, impatient, or aggressive). This distinction is critical because raw cues do not have inherent meaning until they are processed into psychological characteristics.
- **Classes:** Classes represent the abstract grouping or category of the situation. They provide a high-level label that overviews various cues and characteristics into a recognisable type (e.g., a general traffic situation or a lane change situation).

2.4 The Role of the Annotator

Recognizing this subjectivity requires a perspectivist turn in data labelling [2]. Ground truth is not an objective physical reality but a subjective measurement shaped by the annotator's internal script. For instance, labelling a driver as "aggressive" is an attribution of a psychological characteristic, not a measurement of a physical cue. Consequently, disagreement between annotators is not statistical noise to be eliminated, but a diamond standard signal reflecting the intrinsic complexity of social interaction [2]. A robust system must therefore model this ambiguity as valid social information rather than error.

3 Methodology

To address the research questions regarding the relationship between contextual dimensions and narrative open-endedness, this study adopted a multi-layered qualitative approach. First, a systematic literature survey was conducted to curate large numbers of literatures about driving and vulnerable road user (VRU) scenarios to differentiate between trajectory-based models and social narrative human interpretations. These sources were then analysed using a hybrid theoretical framework integrating the 3Cs (Cues, Characteristics, Classes) with script theory. This framework served as a structured lens to extract specific dimensions of variation and map them to the activation of internal and external scripts. Finally, these extracted dimensions were applied to construct comparative case studies, where specific parameters were manipulated to rigorously observe the open-endedness or convergence of intention narratives between human and machine agents.

3.1 Literature Survey Methodology

To ensure the gathered literature was strictly relevant to the research objectives, a structured search strategy was developed in parallel with specific scope definitions. This process involved defining a search query to capture the broad intersection of driving and social intention, while simultaneously applying inclusion criteria to ensure the results focused on narrative understanding rather than pure trajectory prediction.

Search Strategy and Term Elicitation

The systematic search was performed using the Scopus database. The specific terms and keywords can be seen in Table 1. These were elicited through an iterative process. Initial keywords were identified from the titles and abstracts of foundational starter papers provided in the beginning of this research. These were then expanded with synonyms and related concepts (e.g., expanding "intention" to include "plausible narrative" and "script theory") found during exploratory searches to ensure coverage of both technical and psychological domains.

The query combined four key conceptual blocks:

1. **Intention/Behavior Understanding:** Terms related to the estimation, prediction, and interpretation of human actions.
2. **Subjectivity/Context:** Terms capturing the variability of interpretation, including "narratives," "script theory," "social norms," and "contextual variation".

3. **The Driving Domain:** Specific terms to constrain the results to everyday automotive vehicle contexts (e.g., cars, traffic, ADAS).
4. **Exclusion Criteria:** A negative filter to remove irrelevant domains and technical specifications.

Category	Search Terms
Intention & Behavior	social intention*, intent* recognition, intent* estimat*, intent detection, intent* predict*, behavio*r interpret*, action understanding, social percept*, action prediction, goal inference, narrative expl*, plausible narrative*, intent* inferenc*, goal recogn*, goal estimat*, goal inferenc*, behavior recogn*, behavior estimat*, behavior inferenc*, behavior understand*
Subjectivity & Context	subjective annotation, plausible narrative*, alternative narrative*, script theory, social norm*, contextual variation, human factors, subjectivity, perspectiv*, multi-perspective, ambigu*, uncertainty, context model*, situation awareness, external script*, internal script*, multiple annotator perspective*, multi-perspective, situational script*, commonsense script*, situation*, scenario*, case?stud*
Domain	driving W/25 (car* OR vehicle* OR road* OR automobile* OR video* OR simulat* OR situat*) driver behaviour, driver performance, steering, pedestrian*, intelligent transport*, autonomous vehicle, self-driving car, automated vehicle, autonomous car, autonomous driving, traffic, ADAS, intelligent transport*
Exclusion Criteria	ship*, vessel*, maritime, subway, rail transit, two-wheeler*, Motorcyclist, bicycle, cyclist, drone*, UAV, indoor robot, crane, lifting, chemical plant, wheel loader, network security, bandwidth, respiratory, clinical, EEG, pedometer, step count*,Exercise, health, alcohol, Customer Insights, customer experience research, travel planning, travel assistant, willingness-to-pay, carsharing, Children, carbon mitigation, climate change, sexual behavior, crowd management, crowd control, public traffic, Foot Traffic

Table 1: Search query keywords used to obtain initial literatures for the literature survey

Selection Procedure

The initial query yielded over 600 results. To ensure a diverse and manageable selection of high-quality literature, the following stratification and filtering process was applied:

1. **Language Filter:** Only papers written entirely in English were included to ensure accurate interpretation.
2. **Subject Area Filter:** Results were grouped by Scopus subject area. To guarantee diversity, selection began with the least populated subject areas and progressed to the most populated. This prevented the over-representation of a single dominant field (e.g., Engineering) and minimised duplicate checks.
3. **Top-N Selection Strategy:** Within each subject area, two sorting lists were generated as of January 11, 2026:
   - **Most Recent:** The top 5 most recently published papers.
   - **Most Cited:** The top 5 most cited papers.

These papers then underwent a further full-text review to determine their suitability. This review utilised the thematic criteria visualised in Figure 1 as a pass/fail filter.

Figure 1: Visualisation of the inclusion criteria used during the full-text review. The concentric diagram illustrates the hierarchical scope: papers were retained only if they satisfied all inner boundary conditions, specifically focusing on short-duration social interactions modifiable by script theory from the driver's perspective.

Papers were discarded and replaced by the next best-ranked paper in their respective list if it was a duplicate from a previous subject area or if they failed to meet these criteria:

- **Social Component:** The scenario must involve interaction between agents (e.g., between a driver and pedestrian or between a driver and a model) rather than single-agent performance. This is because having multiple agents allows the generation of multiple equally plausible between these agents.
- **Situational Duration:** The study must focus on short situations (approx. 5 to 30 minutes) to prioritize immediate social interactions over long-term habits and to make the actions as reliable as possible without risking external noise[1].
- **Theoretical Relevance:** The paper must address elements modifiable by script theory and have modifiable cues to allow the creation of case studies that contains multiple narratives.
- **Perspective:** The study must present the scenario from the perspective of the driver or ego-vehicle. This criterion is necessary to maintain a consistent observer role when analysing internal scripts, ensuring the focus remains on how a driver interprets the intentions of other road users.

3.2 Dimension Extraction Framework

This research integrated script theory with the 3Cs framework (Cues, Characteristics, Classes) as defined in section 2, to extract the dimensions of variation in the driving domain. While the theoretical background defines these concepts individually, this methodology merges them into a single analytical model to trace the critical moment where objective observation diverges into subjective interpretation.

As illustrated in figure 2, the framework integrated the interaction, where physical cues, situational classes and characteristics served as triggers for the observer's external and internal script, which then generates multiple subjective narratives.

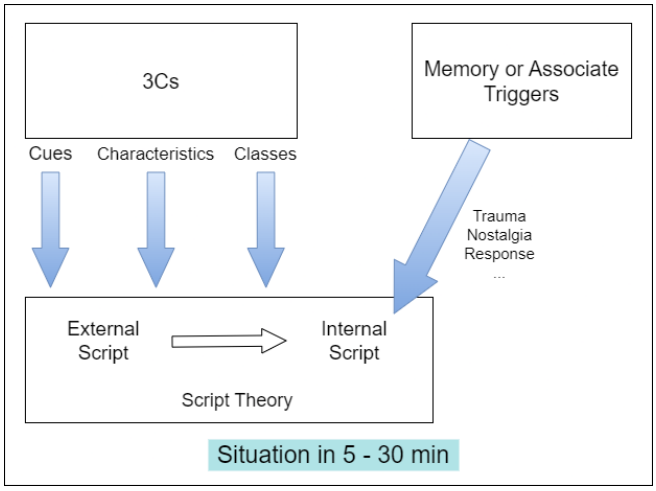

Figure 2: The dimension extraction framework linking the 3Cs to script theory. Cues, characteristics and classes trigger external and internal scripts, which generate multiple plausible narratives.

For each paper in the final selection, the driving scenario was analysed by extracting the following data points corresponding to the 3Cs framework to allow for the extraction of multiple plausible narratives that can be generated:

- **Cue extraction:** The specific physical variables the authors treated as significant inputs were identified. For example, did the model rely solely on kinematic cues (e.g., velocity or position) or did it incorporate social signaling cues (e.g., indicator timing or gaze direction)?
- **Characteristic Attribution:** The study was then analysed if it explicitly modelled the subjective meaning of those cues. These include any evidence of characteristic attribution where the model assigns a psychological state (e.g., "aggressive," "hesitant," "distracted") to an agent before predicting their action.
- **Class and Script Context:** The study was checked if it acknowledged the situational class (e.g., an unsignalised intersection) and the active scripts (e.g., Right-of-Way norms) that constrained the interaction.

By mapping these extracted elements onto the integrated framework, 2 case studies had been generated. For each case study, a constant class is used while manipulating specific physical cues (e.g., gap size, gaze direction) to create strong constraint and weak constraint variations. By comparing the human interpretation derived from the framework against the machine interpretation derived from standard geometric feasibility models the specific conditions under which algorithmic predictions diverge from human social reality has been isolated.

4 Literature Survey Results

To identify the current landscape of social intention in driving, the selected literature has been categorised into two primary domains: *Non-Intelligent Systems* and *Intelligent Systems*. This categorisation has been made to differentiate the foundational understanding of human social norms against the algorithmic approaches currently used to predict them. Within each domain, literature were further sub-categorised into the *Driver* and the *Vulnerable Road User (VRU)* groups as these were the 2 primary roles being observed.

4.1 Non-Intelligent Systems

This category examined how humans naturally perceived intention and negotiated space without the intervention of intelligence systems.

The Driver Role

Within the complex social-technological construct of the transportation system, the driver role is defined by a continuous process of negotiation and intention [36; 45]. Using the 3Cs framework as described in section 2, it appeared that human drivers operated in various classes where they had to process physical cues in each class to infer the intentions of other road users.

These cues include objectively quantifiable elements such as vehicle kinematics(eg., a vehicle's average speed or their acceleration and stopping behaviours), lateral positioning(a vehicle's movement across the width of a lane), and indicator timing(the time when a vehicle makes an explicit turn signals to signal a manoeuvre)[36; 14]. These cues served as the primary input for a driver's internal script. This internal script represented the driver's subjective working model of the situation, which they use to predict the overall sequence of events and execute specific individual actions like merging into another lane or yielding for another driver.

Drivers relied heavily on implicit communication through kinematics to assess the characteristics of others, often prioritising these signals over explicit Human-Machine Interfaces[36; 22]. A research by Tian et al.[36] indicated that specific kinematic cues, such as vehicle deceleration or lateral offset, directly shaped a driver's assessment of whether the other driver had an egoistic, altruistic, or a competitive perceived characteristic. However, it is notable that this study seemingly treats these assessments as a single objective reality. From a perspectivist lens, this approach potentially overlooks the idea of multiple narratives, as it does not account for how the observer's own internal script might lead to different interpretations of the same kinematic cues. Another research by Kauffmann et al. [14] identified that earlier indicator signalled significant increase in a driver's perceived

cooperativeness, thereby triggering a less defensive internal script for the observer.

The open-endedness of driving often stemmed from the gap between the external script and the driver's internal script. Research done by Gaymard et al. [9; 10] highlighted that drivers often adopted a system of informal social norms that justified legitimate transgressions, such as exceeding a speed limit when they felt it was safe or socially acceptable. These subjective interpretations meant that the same physical cue could generate multiple equally plausible narratives of intent depending on the observer's personality.

The VRU Role

The role of the Vulnerable Road User (VRU), such as pedestrians and cyclists, was defined by a continuous process of implicit communication and the negotiation of space with surrounding drivers, rather than simple binary choices[35; 28]. Using the 3Cs framework as described in section 2, the cues involved in these interactions such as physical proximity between a car and a pedestrian, the vehicle speed, and the gap acceptance threshold, served as the raw data from which road users derived meaning[35; 30].

However, the characteristics of a situation, including the psychological meaning of these cues (e.g., perceived stress or threat), often led to open-endedness. A single observed behaviour may support multiple equally plausible narratives of intention. For instance, a pedestrian lingering at a kerb might be interpreted as waiting for a gap, being distracted, or cautiously yielding despite having the right of way.

Negotiation in these scenarios is heavily influenced by the interplay between external and internal scripts. The external script provided the standardised version of the situation, such as the established rules of the road that grant pedestrians priority at zebra crossings. In contrast, a VRU's internal script is their personal working model, shaped by factors like age or risk tolerance.

Ambiguity often arised from a mismatch between these scripts, such as when a timid pedestrian's internal script caused them to hesitate, even when the external script granted them the right of way. This negotiation is rarely instantaneous as it is a temporal play where intent is progressively clarified through social signalling (e.g., gaze or posture changes)[31; 28]. This dynamic renders pure trajectory forecasting insufficient since forecasting assumes a fixed path based on current physics, but in a negotiation, the pedestrian's future path is conditional on the vehicle's response. A system that focused on stopping whenever a pedestrian is near, risks creating an indefinite deadlock where neither agent moves because the vehicle fails to signal the intent to yield.

A research by Rasouli et al.[28] into specific modalities highlights that gaze and head orientation are primary indicators used by VRUs to communicate awareness and initiate space negotiation. A VRU's change in head orientation acts as an implicit signal that they are actively checking the state of traffic or requesting the right of way, which helped to disambiguate their intention for other actors.

4.2 Intelligent Systems

This category reviews the evolution of algorithmic prediction from early probabilistic models to modern Large Language Models (LLMs). While achieving high quantitative accuracy, a critical analysis reveals a methodological gap in how these systems predominately treat intention as a geometric output, often failing to reconstruct the underlying reasoning process.

Predicting the Driver

The prediction of driver intention has transitioned through three technical paradigms. Early methodologies utilized Bayesian networks and Hidden Markov Models (HMMs) to map unobservable states to vehicular dynamics [13; 24]. While these provided transparency, they struggled with high-dimensional nonlinearity [43; 33]. The field subsequently shifted toward sequence learning (LSTMs, RNNs) and hybrid CNN architectures to capture temporal dependencies [19; 39; 47; 11]. However, these "black box" models prioritize outcome prediction over causal reasoning, often averaging diverse behaviour modes into unrealistic trajectories [23].

Recent advancements between 2023 and 2026 have established transformer architectures and Large Language Models (LLMs) as the dominant paradigm due to their self-attention mechanisms, which enable global context modelling and the parallel processing of multimodal inputs (LiDAR, radar, cameras, and HD maps)[5]. These systems excel at capturing long-range spatiotemporal dependencies and social correlations among multiple agents in dense traffic[5; 8].

Separate research by Wang et al. [40] and Peng et al. [25] replaced traditional recurrence with multi-head attention. This mechanism paralleled the cue selection phase of human cognition where rather than processing all input data sequentially, the attention heads assigned dynamic weights to specific interactions. This effectively isolated physical cues to construct a more accurate representation of the scenario's global class.

Furthermore, the integration of Large Language Models (LLMs) has introduced a behavioural cognition layer aimed at high-level reasoning and risk assessment[46; 5]. Specifically, an emerging research by Peng et al.[26] using a LC-LLM (Lane Change-Large Language Model), reformulated intention prediction as a language modelling problem, utilising Chain-of-Thought (CoT) reasoning to generate natural language explanations for predicted manoeuvrers. Foundation models are now being tested to interpret corner cases and provide semantic explanations for manoeuvrers, such as identifying a cut-in risk, and generating actionable advice for the autonomous system[46; 5].

Simultaneously, hybrid models are integrating Mamba (Selective State Space Models) with transformers and CNNs to maintain linear computational complexity while capturing the long-range dependencies required for real-time edge deployment [34].

Despite the increasing accuracy of these models, they suffer from a cold cognition approach [38]. This approach relies on brute-force learning that maps cues from sensor readings directly to action (predicted trajectory or manoeuvrer label). This explicitly skips the characteristic which represents

the driver's internal psychological script, emotional state, and subjective meaning of the situation.

By focusing solely on realised outcomes, these models fail to capture the open-endedness of behavioural interpretation where the same physical behavior can stem from multiple, equally plausible internal narratives. Furthermore, even state-of-the-art LLM-based approaches frequently performed post-hoc rationalisation rather than true intent modelling. This is because they generate a semantic narrative to justify a trajectory after it has been predicted, rather than simulating the driver's internal goal-setting process [46; 26]. Consequently, the system lacked a deep understanding of the subjective norms and internal scripts (e.g., an aggressive vs. a timid driver's perspective) that can lead to divergent intention narratives from identical environmental cues.

Predicting the VRU

Research into pedestrian intention estimation has pivoted towards leveraging high-dimensional sensory data to anticipate interactions before they manifest as physical actions. Current methodologies for VRU analysis can be categorised into two primary streams: trajectory forecasting (predicting future coordinates) and intention estimation (predicting high-level behavioural states).

The first primary input modality is vision-based approaches. These approaches primarily extracted physiological features to predict immediate actions[32]. Many systems, such as the hybrid framework developed by Li et al.[16], utilise pose estimation algorithms, such as Part Affinity Fields (PAFs), to generate human skeleton representations from monocular video frames. These models operate on the premise that skeletal keypoints, specifically those corresponding to the head, torso, and legs, signalled the shift from a stationary state to a crossing state [3].

In addition to body pose, head orientation and gaze are also treated as powerful predictors of a pedestrian's situational awareness and if they want to yield to the incoming vehicle or cross the street[4; 6]. For instance, dynamic fuzzy automata models incorporate head steering angles and torso tilt to determine if a pedestrian is actively checking for oncoming traffic at nighttime, acting as an indication for the internal script of caution [15]. While effective for binary crossing prediction, these vision-centric models often struggle to generalise across crowded or occluded scenarios where high-quality skeletal data is difficult to extract[20].

The second primary input modality is context-aware models. Context-aware models extend the predictive horizon by integrating environmental data and social interactions through Graph Convolutional Networks (GCNs) and spatiotemporal reasoning[17]. These frameworks move beyond individual kinematics to model the spatiotemporal relationships between a pedestrian and other entities in the scene, such as vehicles, traffic lights, and zebra crossings [44; 17]. For example, the pedestrian graph approach by Liu et al.[17], where the pedestrian and surrounding objects (vehicles, traffic lights) serve as nodes, allowed the system to reason about relationships over time.

Furthermore, some models accounts for group information, acknowledging that a pedestrian's crossing decision is often influenced by the behaviour of surrounding crowds[48; 20]. By using semantic segmentation to parse the global context, these systems can adapt to complex urban environments where localised pedestrian information must be weighted against broader environmental cues[20]

Despite these technical advancements, a critical methodological gap remains in how these systems conceptualise human intent. Most existing systems treat intention as a binary classification problem (Cross vs. No-Cross)[21], effectively conflating the trajectory prediction with the inference of social intent.

As argued by Hung et al.[12], with the exception of specific intention-focused datasets (see subsection 4.3), the majority of trajectory-based models models suffered from the intention by outcome bias wherein an intent is only validated if it results in a completed action. Crossing scenarios are not merely physical displacements but social negotiations. Before initiating a crossing, a VRU often seeks to establish trust or mutual awareness with the driver. However, current models often ignore the subtle cues used for this negotiation, such as subtle gestures and eye contact used to establish trust[37]. While recent approaches cast forecasting as a probability distribution over potential futures, they remain limited to a description of kinematics behaviour rather than the embodiment of intention narratives. A probabilistic model may simulate multiple trajectories, but without understanding the causal internal script driving those trajectories, it cannot distinguish between a high-risk gamble and a calculated negotiation. By committing to a single ground truth rather than accounting for multiple plausible narratives, intelligent systems risk making catastrophic errors in high-stakes driving environments where the correct response depends on the meaning of the action, not just its likelihood.

Section 4 will explicitly address this gap by applying the dimension extraction framework mentioned in the methodology section to real-world collision scenarios to demonstrate how the inclusion of characteristics resolves the ambiguity of unrealised intentions.

4.3 Notable Exclusions and Limitations of the Search

It is necessary to acknowledge that the strict inclusion criteria of the systematic search strategy in table 1 resulted in the exclusion of certain literatures that are relevant to the domain of intention estimation. A prominent example is the PIE dataset[27] by Rasouli et al. While this work is fundamental to the distinction between trajectory forecasting and intention estimation, it did not appear in the final search query, likely due to the specific combination of keywords. Although excluded from the systematic analysis, its contribution has been acknowledged in addressing the outcome bias. Unlike standard datasets that label intention based on physical displacement, PIE explicitly labels crossing intention independently of action. This aligns with the perspectivist approach proposed in this thesis, though future iterations of this review would require broader search terms to formally capture such datasets.

5 Case Studies

This section applies the theoretical framework established in Section 3.2 to two prototypical driving scenarios: a mandatory lane merge (Case A) and a pedestrian crossing interaction (Case B). The primary objective was to demonstrate the inverse relationship between contextual constraints and narrative open-endedness.

5.1 Case Study A: Lane Merging

Table 2 compares two variations of a mandatory lane change event. For case A1's scenario, vehicle A is forcing a mandatory merge in front of Vehicle B. For case A2, vehicle A executes the same mandatory merge infront of a vehicle B but with parameters adjusted for safety and social compliance. By keeping the class constant (mandatory merge) but varying the cues (gap size, signaling), these variations were made to show how open-endedness emerges from script conflict.

5.2 Case Study B: Pedestrian near Zebra Crossing

This case study tests interpretation in a shared space (Class: Urban uncontrolled zebra crossing). Unlike traditional outcome-based models [12], this analysis focuses on unrealized intentions where a pedestrian's intent may not manifest as immediate physical movement. In case B1, a pedestrian stands at the curb of an urban uncontrolled zebra crossing but is looking down at their smartphone. They are stationary but standing close to the road edge while a driver is heading towards the zebra crossing. In case B2, the pedestrian stands at the same curb on the edge of the uncontrolled zebra crossing but now they are looking directly at the oncoming traffic. They are stationary but standing close to the road edge while a driver is heading towards the zebra crossing.

6 Discussion and Recommendations

6.1 Discussion

Synthesizing the findings from Section 5 answers the third research question regarding the relationship between constraints and narrative variation. The analysis reveals a distinct inverse relationship where as the strength of physical or social constraints increases, the dominance of the external script rises, resulting in low open-endedness where human and machine interpretations converge.

In scenarios defined by strong constraints, such as the Cooperative Merge (Case A2) or Attentive Pedestrian (Case B2), ambiguity is minimised. Strong cues (e.g., a 30m gap or direct eye contact) act as clear triggers for shared external scripts, aligning the human driver's yielding intention with the machine's geometric safety assessment. Conversely, open-endedness peaks when constraints are weak. In the Aggressive Merge (Case A1), the machine's geometric logic (finding the gap feasible) clashes with the human's territorial defence script triggered by the hostile late signal [14]. Similarly, in the Distracted Pedestrian scenario (Case B1), the machine infers safety from a lack of motion, failing to capture the narrative risk of distraction.

To address these disconnects between geometric processing and social interpretation, this study proposes three strategies for future intelligent systems:

1. Shift from Trajectory to Narrative Output: Current cold cognition models prioritise outcome prediction over causal reasoning [38]. To bridge the gap exposed in Case A1, future architectures must move beyond mapping raw inputs to a single ground truth. Instead, systems should output semantic intention labels (e.g., *Yielding, Harassing, Distracted*) as an intermediate or primary objective. Explicitly modelling these characteristics allows the system to reason about social compliance and explain why an action is taken, rather than just calculating physical feasibility.

2. Uncertainty as a Safety Signal: In high open-endedness scenarios, systems must avoid forced binary decisions (e.g., Cross/No Cross). As seen in Case B1, a stationary but distracted pedestrian represents high narrative uncertainty despite low kinematic risk. A narrative-aware system should detect this script mismatch where physical cues suggest safety but psychological characteristics suggest risk, and output a high uncertainty state to widen safety margins.

3. Explainable Social Interfaces: Finally, adopting narrative understanding requires a parallel evolution in Human-Machine Interfaces (HMI). If an autonomous vehicle yields to an aggressive merger (Case A1) without explanation, a human passenger may perceive it as a sensor error. However, if the HMI explicitly communicates the inferred narrative (e.g., *"Yielding to Aggressive Driver"*), it aligns the passenger's internal script with the machine's logic, thereby maintaining trust in the system's social competence.

7 Responsible Research

This section outlines the measures taken to ensure the scientific integrity of this study. It includes the reproducibility of the literature survey, the ethical implications of social intention modelling in intelligent driving systems, and transparency regarding the use of artificial intelligence tools.

7.1 Reproducibility

To ensure the reproducibility of the literature survey presented in Section 4, the search strategy and selection process have been documented in the methodology in Section 3.

The systematic search was conducted using the Scopus database. To minimise selection bias and allow for replication, the exact search query keywords used to search for the literature papers are detailed in Table 1.

Given that the initial query yielded a substantial volume of results (exceeding 600 entries), a rigorous manual screening process was employed as detailed in subsection 3.1. Each paper was individually evaluated for relevance to ensure the final selection strictly aligned with the research objectives.

To extract the dimensions, a theoretical dimension extension framework had been created to aid in the creation of case studies. This framework is explained in detail in section 3.2

7.2 Ethical Concerns

The core motivation of this research causes ethical implications for the design of Intelligent Transportation Systems (ITS). As highlighted in subsection 1.1, autonomous driving is a high-stakes environment where misinterpreting social intent can lead to catastrophic collisions. Traditional sys-

Table 2: Comparison of Human vs. Machine Interpretation in Lane Merging Scenarios (Class: Mandatory Lane Change)

Feature	Case A1: Aggressive Merge	Case A2: Cooperative Merge
Cues	Gap: 9.4m (Critical); Late Signal (12m prior)	Gap: 30m (Safe); Early Signal (60m prior)
Psychological Characteristic	Hostile, Egoistic, Rule-Violator. Late signal implies priority of efficiency over social fairness [36].	Safe, Polite, Skilled. Early signal is perceived as cooperative [14].
Internal Script (Human)	**Territorial Defence.** Vehicle B's driver feels justified in punishing the behaviour to maintain social order.	**Yielding.** Vehicle B's driver allows vehicle A to merge safely.
Machine Interpretation	**Safe to Yield.** TSWHMM model [19] sees gap as feasible and assumes standard traffic rules apply.	**Safe to Yield.** Model sees geometric feasibility and alignment with traffic rules.
Outcome	**High Open-endedness (Divergence).** Machine predicts yield to vehicle A but vehicle B's driver may block instead.	**Low Open-endedness (Convergence).** Both agents predict yielding to vehicle A's driver.

Table 3: Comparison of Human vs. Machine Interpretation in Pedestrian Scenarios

Feature	Case B1: Distracted Pedestrian	Case B2: Attentive Pedestrian
Cues	Velocity: 0; Head: Down/Away; Eye contact: on their smartphone	Velocity: 0; Head: Facing Road; Eye contact: on oncoming traffic
Psychological Characteristic	Distracted, Unaware. Future actions unpredictable despite current stance.	Attentive, Aware, Waiting. Direct gaze confirms mutual awareness.
Internal Script (Human)	**Caution/Hesitation.** Driver slows down due to potential for sudden, unchecked crossing [31; 35].	**Mutual Awareness.** Driver trusts pedestrian sees them and prepares to yield/pass safely.
Machine Interpretation	**No Crossing.** Trajectory model projects stationary state forward (Velocity=0 implies no intent).	**Crossing Intention.** Model predicts crossing based on head orientation facing traffic.
Outcome	**High Open-endedness (Divergence).** Human prepares for sudden risk while machine assumes safety.	**Low Open-endedness (Convergence).** Machine assumes safety. Pedestrian's eye contact acts as a closing social signal to human.

tems that classify intention as a binary ground truth ignore the inherent ambiguity of human interaction. An ethical concern arises if future systems act with high confidence on ambiguous signals. By acknowledging that a single behaviour (e.g., hesitation) can generate multiple plausible narratives, this framework advocates for systems that maintain uncertainty rather than forcing potentially dangerous deterministic decisions.

Another ethical implication is that the proposed use of internal scripts to model multiple driver's perspectives introduces the risk of encoding human biases into algorithmic decision-making. If an autonomous vehicle's internal script is trained on data that disproportionately associates specific driving styles or certain demographics or regions, it may unfairly penalise or misinterpret the actions of individuals that are not part of these groups. Responsible research in this domain must ensure that the definition of situational norms is inclusive and does not reinforce harmful stereotypes that might cause unpredictable behaviours within the system.

7.3 Usage of AI

The usage of AI in this research, particularly Large Language Models(LLMs), are listed in this section.

- **Processing and Summarisation:** LLMs particularly Gemini and NotebookLM, were utilised to assist in the processing of the initial literature results of approximately 600 papers. AI tools were used to generate summaries of each paper individually to aid in the exclusion of irrelevant literature papers. However, the final selection of papers and categorisation of the literatures were performed manually to ensure accuracy.
- **Language Refinement:** LLMs were used to check for grammatical and spelling errors.

8 Conclusion

8.1 Summary of Research Findings

This research investigated the disconnect between algorithmic trajectory prediction and human social interpretation in the domain of autonomous driving. The primary research question sought to determine how contextual constraints influence multiple equally plausible narratives of social intentions.

By integrating the 3Cs framework (Cues, Characteristics, Classes) with script theory, this study concludes that current intelligent systems operate predominantly in geometric space, optimizing for physical feasibility (e.g., velocity vectors). In contrast, human interaction occurs in social space, governed by normative scripts.

An analysis of the case studies yielded a definitive answer to sub-research question 3, where open-endedness is struc-

turally inversely proportional to the strength of physical and social constraints. When constraints are strong (e.g., a cooperative merge scenario with a large gap between vehicles), the external script (traffic rules) dominates, and human-machine interpretations converge. Conversely, when constraints are weak or ambiguous (e.g., an aggressive merge with a small gap between vehicles), the external script fails. Humans then start to rely on their subjective internal scripts (e.g., territorial defence), whereas geometric intelligent models fails to capture these variables, leading to potentially critical prediction errors.

8.2 Contributions

The primary contribution of this work is the formalisation of script mismatches as the root cause of intelligent system failure in social settings. Unlike previous studies that attribute prediction errors to sensor noise or occlusion, this research demonstrates that errors often arise because the machine perceives a trajectory, while the human perceives a violation of norms. By mapping specific physical cues (like a late turn signal) to psychological characteristics (aggression), this thesis provides a theoretical blueprint for why even technically safe manoeuvrers can be perceived as socially dangerous.

8.3 Open Issues and Future Directions

While this study establishes a theoretical framework, several open issues remain. First, this research relied on qualitative case study analysis derived from literature and it lacks large-scale empirical validation with human subjects to quantify exactly when a script mismatch is triggered. Secondly, a major barrier to implementing the proposed framework is the scarcity of training data. Current datasets are heavily biased towards geometric labels (like bounding boxes and velocity). To validate the 3Cs framework, the community requires new datasets annotated with socio-psychological labels (e.g., "Hesitation," "Aggression," "Negotiation Phase") similar to the approach taken by the PIE dataset, but expanded to vehicle-to-vehicle interactions. Lastly, integrating an LLM-based reasoning layer may introduce significant inference latency. A key open issue is how to distil these complex script understandings into lightweight models capable of real-time inference in a short amount of time is required for safety-critical braking decisions.

8.4 Recommendations for Further Research

Firstly, future architectures should treat intention as a semantic classification task (e.g., outputting labels like "Yielding" or "Distracted") as an intermediate or primary objective. This forces the model to learn the why before predicting the trajectory of where.

Secondly, research should focus on systems that can detect script divergence. If a pedestrian's kinematics conflict with their context, the system should explicitly quantify this narrative uncertainty to adjust safety margins.

Finally, as autonomous systems begin to make decisions based on invisible social cues, they must be able to explain these decisions to human passengers. Future research should focus on Human-Machine Interfaces (HMI) that can translate the system's internal script assessment into natural language (e.g., Slowing down because the pedestrian looks distracted), thereby calibrating human trust in the automated system.

References

[1] Anna Belardinelli. Gaze-based intention estimation: Principles, methodologies, and applications in hri. *ACM Transactions on Human-Robot Interaction*, 13(3), 2024.

[2] Federico Cabitza, Andrea Campagner, and Valerio Basile. Toward a perspectivist turn in ground truthing for predictive computing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(6):6860–6868, June 2023.

[3] Dong Cao and Yunbin Fu. Using graph convolutional networks skeleton-based pedestrian intention estimation models for trajectory prediction. volume 1621, 2020.

[4] Xiaobo Chen, Shilin Zhang, Wei Xu, Dapeng Cheng, and Lei Yang. Robust pedestrian crossing intention prediction via uncertainty-guided transformer ensemble network for autonomous driving. *IEEE Transactions on Instrumentation and Measurement*, 74, 2025.

[5] Fulin Chu, Haoyu Li, Lili Xie, and Jingyuan Zhao. A survey of transformer architectures for autonomous driving. *Expert Systems with Applications*, 299, 2026.

[6] Anup Doshi and Mohan Manubhai Trivedi. On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes. *IEEE Transactions on Intelligent Transportation Systems*, 10(3):453 – 462, 2009.

[7] Frank Fischer, Ingo Kollar, Karsten Stegmann, and Christof Wecker. Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, 48(1):56–66, 2013.

[8] Kai Gao, Xunhao Li, Bin Chen, Lin Hu, Jian Liu, Ronghua Du, and Yongfu Li. Dual transformer based prediction for lane change intentions and trajectories in mixed traffic environment. *IEEE Transactions on Intelligent Transportation Systems*, 24(6):6203 – 6216, 2023.

[9] Sandrine Gaymard. Norms in social representations: Two studies with french young drivers. *European Journal of Psychology Applied to Legal Context*, 1(2):165 – 181, 2009.

[10] Sandrine Gaymard, Philippe Allain, François Osiurak, and Didier Le-Gall. The conditions of respect of rules in young and elderly drivers: An exploratory study. *European Journal of Psychology Applied to Legal Context*, 3(1):11 – 28, 2011.

[11] He Huang, Zheni Zeng, Danya Yao, Xin Pei, and Yi Zhang. Spatial-temporal convlstm for vehicle driving intention prediction. *Tsinghua Science and Technology*, 27(3):599 – 609, 2022.

[12] Hayley Hung, Litian Li, Jord Molhoek, and Jing Zhou. The discontent with intent estimation in-the-wild: The case for unrealized intentions. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '24, New York, NY, USA, 2024. Association for Computing Machinery.

[13] Jiaxuan Jiang, Jiapeng Liu, Miłosz Kadziński, and Xiuwu Liao. A bayesian network approach for dynamic behavior analysis: Real-time intention recognition. *Information Fusion*, 118, 2025.

[14] Nina Kauffmann, Franz Winkler, Frederik Naujoks, and Mark Vollrath. "what makes a cooperative driver?" identifying parameters of implicit and explicit forms of communication in a lane change scenario. *Transportation Research Part F: Traffic Psychology and Behaviour*, 58:1031 – 1042, 2018.

[15] Joon-Young Kwak, Byoung Chul Ko, and Jae-Yeal Nam. Pedestrian intention prediction based on dynamic fuzzy automata for vehicle driving at nighttime. *Infrared Physics and Technology*, 81:41 – 51, 2017.

[16] Yanfen Li, Hanxiang Wang, L. Minh Dang, Tan N. Nguyen, Dongil Han, Ahyun Lee, Insung Jang, and Hyeonjoon Moon. A deep learning-based hybrid framework for object detection and recognition in autonomous driving. *IEEE Access*, 8:194228 – 194239, 2020.

[17] Bingbin Liu, Ehsan Adeli, Zhangjie Cao, Kuan-Hui Lee, Abhijeet Shenoi, Adrien Gaidon, and Juan Carlos Niebles. Spatiotemporal relationship reasoning for pedestrian intent prediction. *IEEE Robotics and Automation Letters*, 5(2):3485 – 3492, 2020.

[18] Jinxin Liu, Yugong Luo, Zhihua Zhong, Keqiang Li, Heye Huang, and Hui Xiong. A probabilistic architecture of long-term vehicle trajectory prediction for autonomous driving. *Engineering*, 19:228 – 239, 2022.

[19] Pujun Liu, Ting Qu, Huihua Gao, and Xun Gong. Driving intention recognition of surrounding vehicles based on a time-sequenced weights hidden markov model for autonomous driving. *Sensors (Basel, Switzerland)*, 23(21), 2023.

[20] Zhuofan Liu, Yigang Xue, and Lan Zhu. Pedestrian crossing intention prediction in obstructed scenarios using graph-based feature reconstruction and multi-scale self-attentive encoding. *Engineering Applications of Artificial Intelligence*, 159, 2025.

[21] Javier Lorenzo, Ignacio Parra Alonso, Rubén Izquierdo, Augusto Luis Ballardini, Álvaro Hernández Saz, David Fernández Llorca, and Miguel Ángel Sotelo. Capformer: Pedestrian crossing action prediction using transformer. *Sensors*, 21(17), 2021.

[22] Linda Miller, Jasmin Leitner, Johannes Kraus, and Martin Baumann. Implicit intention communication as a design opportunity for automated vehicles: Understanding drivers' interpretation of vehicle trajectory at narrow passages. *Accident Analysis and Prevention*, 173, 2022.

[23] Mozhgan Nasr Azadani and Azzedine Boukerche. A novel multimodal behavior prediction method for automated vehicles. page 47 – 54, 2023.

[24] Nuria Oliver and Alex P. Pentland. Graphical models for driver behavior recognition in a smartcar. page 7 – 12, 2000.

[25] Jiahao Peng, Chao Sun, Zitong Chen, and Jiaru Zhong. Unipredictor: Unified network for vehicle intention and trajectory joint prediction. page 887 – 893, 2025.

[26] Mingxing Peng, Xusen Guo, Xianda Chen, Kehua Chen, Meixin Zhu, Long Chen, and Fei-Yue Wang. Lc-llm: Explainable lane-change intention and trajectory predictions with large language models. *Communications in Transportation Research*, 5, 2025.

[27] Amir Rasouli, Iuliia Kotseruba, Toni Kunic, and John Tsotsos. Pie: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6261–6270, 2019.

[28] Amir Rasouli, Iuliia Kotseruba, and John K. Tsotsos. Understanding pedestrian behavior in complex traffic scenes. *IEEE Transactions on Intelligent Vehicles*, 3(1):61 – 70, 2018.

[29] John F. Rauthmann and Ryne A. Sherman. *Conceptualizing and measuring the psychological situation*. 2021.

[30] S. Schmidt and B. Färber. Pedestrians at the kerb - recognising the action intentions of humans. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(4):300 – 310, 2009.

[31] Friederike Schneemann and Frederik Diederichs. Action prediction with the jordan model of human intention: a contribution to cooperative control. *Cognition, Technology and Work*, 21(4):711 – 721, 2019.

[32] Neha Sharma, Chhavi Dhiman, and S. Indu. Pedestrian intention prediction for autonomous vehicles: A comprehensive survey. *Neurocomputing*, 508:120 – 152, 2022.

[33] Qingyuan Song, Weiping Fu, Wen Wang, Yuan Sun, Denggui Wang, and Jincao Zhou. Quantum decision making in automatic driving. *Scientific Reports*, 12(1), 2022.

[34] Haibin Sun and Ning Feng. Mvim: A high-accuracy, lightweight neural network for real-time driver behavior recognition. *Expert Systems with Applications*, 296, 2026.

[35] Kai Tian, Athanasios Tzigieras, Chongfeng Wei, Yee Mun Lee, Christopher Holmes, Matteo Leonetti, Natasha Merat, Richard Romano, and Gustav Markkula. Deceleration parameters as implicit communication signals for pedestrians' crossing decisions and estimations of automated vehicle behaviour. *Accident Analysis and Prevention*, 190, 2023.

[36] Kai Tian, Jiaxun Wu, Tony Z. Qiu, Chaozhong Wu, Hui Zhang, Yi He, Naikan Ding, Wei Lyu, and Chongfeng Wei. A framework for analyzing driver safety-efficiency trade-offs at uncontrolled crosswalks: Towards social vehicle automation. *Safety Science*, 187, 2025.

[37] Balint Varga, Dongxu Yang, and Soren Hohmann. Intention-aware decision-making for mixed intersection scenarios. page 369 – 374, 2023.

[38] Koen Vellenga, H. Joe Steinhauer, Göran Falkman, Jonas Andersson, and Anders Sjögren. Ai vs. humans: Comparing road user intention recognition performance. *Transportation Research Part F: Traffic Psychology and Behaviour*, 118, 2026.

[39] Lei Wang, Jianyou Zhao, Mei Xiao, and Jian Liu. Predicting lane change and vehicle trajectory with driving micro-data and deep learning. *IEEE Access*, 12:106432 – 106446, 2024.

[40] Ling Wang, Qingjie Bian, Aibing Shu, Muamer Abuzwidah, Yingying Xing, and Wanjing Ma. Incorporating roadside traffic data to predict lane change behaviors within weaving segment for automated vehicles. *Journal of Transportation Engineering Part A: Systems*, 151(12), 2025.

[41] Yang Xing, Chen Lv, Huaji Wang, Dongpu Cao, and Efstathios Velenis. An ensemble deep learning approach for driver lane change intention inference. *Transportation Research Part C: Emerging Technologies*, 115, 2020.

[42] Xintong Yan, Jie He, Guanhe Wu, Changjian Zhang, and Chenwei Wang. A proactive recognition system for detecting commercial vehicle driver's distracted behavior. *Sensors*, 22(6), 2022.

[43] Wei Yang, Bo Wan, and Xiaolei Qu. A forward collision warning system using driving intention recognition of the front vehicle and v2v communication. *IEEE Access*, 8:11268 – 11278, 2020.

[44] Shilin Zhang, Xiaobo Chen, Wei Xu, Lei Yang, and Jian Yang. Evidential multimodal fusion network for trusted pedestrian crossing intent prediction. *IEEE Transactions on Computational Social Systems*, 2025.

[45] Zhengming Zhang, Md Fazle Elahi, Joshua Domeyer, and Renran Tian. Driver temporal segmentation of pedestrian crossing intentions during negotiations. *Transportation Research Part F: Traffic Psychology and Behaviour*, 114:953 – 969, 2025.

[46] Jian Zhou, Minghao Yu, Yuan Guo, Bijun Li, Shen Ying, and Zhijiang Li. A high-definition map architecture for transportation digital twin system construction. *International Journal of Applied Earth Observation and Geoinformation*, 144, 2025.

[47] Yanyan Zhou, Lijun Yang, Tian Wang, and Jiali Peng. Vehicle trajectory hybrid prediction method integrating driver behavior characteristics. page 583 – 590, 2024.

[48] Hai Zou, Yongqing Guo, Fulu Wei, Dong Guo, Qingyin Li, and Jahongir Pirov. A pedestrian group crossing intention prediction model integrating spatiotemporal features. *Scientific Reports*, 15(1), 2025.