Transportation Research Interdisciplinary Perspectives 32 (2025) 101545

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: www.sciencedirect.com/journal/transportation-research-interdisciplinary-perspectives

Public trust and blame attribution in human-AI interactions: a comparison between air traffic control and vehicle driving

Peidong Mei [a,b,c,*], Richard Cannon [d], Jim Everett [e], Peng Liu [f], Edmond Awad [g,b,*]

[a] School of Psychology, University of Aberdeen, King's College, Aberdeen AB24 3FX1, UK
[b] Uehiro Oxford Institute, University of Oxford, Saint Ebbe's St, Oxford OX1 1PT, UK
[c] Department of Engineering, University of Exeter, The Queen's Dr, Exeter EX4 4QH, UK
[d] Research and Development, NATS, 4000 Parkway, Whiteley PO15 7FL, UK
[e] School of Psychology, University of Kent, Keynes College, Canterbury CT2 7NP, UK
[f] Center for Psychological Sciences, Zhejiang University, Yuhangtang Rd, Hangzhou 310027, China
[g] Department of Economics, University of Exeter, The Queen's Dr, Exeter EX4 4QH, UK

A R T I C L E I N F O

Keywords:
Human-AI interaction
Trust
Blame
Public perception
Automation levels
Air traffic control
Autonomous vehicle

A B S T R A C T

Artificial Intelligence (AI) has potential to address the increasing demand for capacity in Air Traffic Control (ATC). However, its integration poses several challenges and requires deep understanding of public perception. Insights from the context of Autonomous Vehicles (AVs), in which more studies have been done, can inform such understanding. In this article, we investigate how the public perceives the automated future of ATC in close comparison to AVs. We conducted two studies to examine public trust and blame attribution toward human and AI operators in different Human-AI Interaction (HAI) models, covering three levels of automation (*Level 0: AI tool, Level 3: AI trainee, and Level 5: AI manager*). We also explored their perceptions of ATC and vehicle driving (VD) by using ten task-related measures (*Familiarity, Expertise, Tech Awareness, Openness, Media Discourse, Stake, two measures of Uncertainty, Positive Safety, and Negative Safety*) and five agent-related characteristics (*Capability, Robustness, Predictability, Honesty and Cooperativeness*). The results showed greater trust and less blame attributed to humans in both ATC and VD, except in the Level 3 AI trainee model where humans were blamed more than AI. We also found both similarities and differences in people's perceptions of the two contexts. Our findings provide evidence-based insights into how the public attribute trust and blame to the operators in ATC and VD. These results will inform industries on the development and implementation of AI integration in aviation and advise policymakers in evaluating public opinion on AI regulation.

Introduction

Traditional transportation industries are undergoing a profound transformation with the advent of automation technologies, with road transportation being one prominent example of this change. The integration of automation, from driverless cars to autonomous trucks, is rapidly reshaping the way people and goods move on roads. The centre of this revolution appears to rely on the performance of the drivers. As these technologies advance, there is potential for machine drivers to support human drivers to dynamically optimise their driving patterns (Van Arem et al., 2006; Spieser et al., 2014; Gao, Hensley and Zielke, 2014; Waymo, 2024). It could lead to reduced traffic congestion, decreased fuel consumption and emissions, and improved efficiency and reliability in goods transportation (Olia et al., 2018; Zhao et al., 2022; Yao et al., 2021; Matin and Dia, 2022; Gueriau and Dusparic, 2020). In light of the significant benefits brought by automation, another critical sector of transportation, the aviation industry is also looking at the potential of automation to tackle their most significant change (Guleria et al., 2021; Perez-Castán et al., 2022): the growing disparity between the increasing demand for air travel and the limited capacity provided by existing air traffic control (ATC) systems (IATA, 2020; ATM, 2024; UK Department for Transport, 2022, 2024).

Abbreviations: ATC, Air Traffic Control; ATCO, Air Traffic Control Officer; VD, Vehicle Driving; AV, Autonomous Vehicle; HAI, Human-AI Interaction; Nat. Rep. Sample, nationally representative sample.

* Corresponding authors at: School of Psychology, University of Aberdeen, King's College, Aberdeen AB24 3FX1, UK (P. Mei). Department of Economics, University of Exeter, The Queen's Dr, Exeter EX4 4QH, UK (E. Awad).

E-mail addresses: peidong.mei@abdn.ac.uk (P. Mei), e.awad@exeter.ac.uk (E. Awad).

https://doi.org/10.1016/j.trip.2025.101545
Received 27 November 2024; Received in revised form 23 May 2025; Accepted 24 July 2025
Available online 2 August 2025

ATC officers (ATCOs) play a crucial role in ensuring safe and efficient air travel. They monitor and direct aircraft as they navigate through sectors of airspace. However, human cognitive capacity has its limits (Lee et al., 2012; Li et al., 2021), and the complexity of the operation means that simply hiring more ATCOs is not a sufficient solution to the challenges faced in optimising ATC systems. During peak hours, ATCOs are estimated to effectively manage up to 15 aircraft simultaneously, although this number can vary depending on factors such as airspace complexity, traffic density, and weather conditions (IOrganization, 2016, 2021; EUROCONTROL, 2023). This constraint poses a significant challenge as air travel continues to grow. To address this capacity gap, the aviation industry is exploring new technologies including Artificial Intelligence (AI) techniques to facilitate conflict resolution and support ATCOs (Van Rooijen et al., 2020; Guleria et al., 2024). However, slow adoption of AI technologies persists due to the safety–critical nature of ATC operations. Specifically, technical challenges in ATC—such as the shorter buffer window for decision-making during emergencies, the inherent uncertainty influenced by weather conditions (Erzberger et al., 2012), and the rigorous standards for testing, training, and certification—make AI integration complex and costly (Zhao and Rozier, 2014).

Beyond technical considerations, psychological perceptions play a pivotal role in this potentially risky but revolutionary solution (Morton et al., 2021; Borhan et al., 2019; Adnan et al., 2018; Zefreh et al., 2023; Tang et al., 2025). The transition to AI-powered ATC is not solely about system capability and safety; it also hinges on how the public and potential users think about the usage of AI in ATC. As end-users of ATC services, will people be willing to trust AI for critical decisions? The potential for AI to revolutionise ATC operations lies in understanding relevant psychological landscapes.

Therefore, in this study we aim to explore public perceptions and attitudes toward the automated future of ATC, using Autonomous Vehicles (AVs) as a reference. We will compare public trust and blame attribution to human and AI agents in these two contexts across different automation levels. Characteristics related to tasks and agents in these two contexts will also be examined. This study will inform the industry not only in understanding the public demand for AI development but also in aligning with their expectations for AI deployment. Moreover, shedding light on psychological attitudes in high-stakes areas facilitates future research in analysing the underlying mechanisms.

Background

Human-AI Interaction Models in ATC* vs. *VD. The journey towards automation involves multiple stages. Automation can range from covering the entire process to only affecting a small part of it. Given this, different 'levels of automation' were proposed. For instance, the Society of Automotive Engineers (SAE) introduced the Levels of Driving Automation for autonomous vehicles (AVs) (SAE, 2018) in 2014 and the automation taxonomy in air traffic control has also been established by the Single European Sky ATM Research (SESAR) in 2020 (SESAR Joint Undertaking, 2020). These levels are defined by the degree of automation (referred to as AI generally in this paper) and the extent of human involvement. Notably, the progression from lower to higher levels of automation, as proposed in both fields, exhibits a consistent pattern of alignment, which is briefly summarised as follows:

- *Level 0: No automation.*
- *Level 1: Decision support (in ATC)/ Driver Assistance (in AV).*
- *Level 2: Partial automation.*
- *Level 3: Conditional automation.*
- *Level 4: High automation.*
- *Level 5: Full automation.*

With AVs stand out as one of the most mature AI applications in industries, it provides comparable understanding for ATC development and allows us to anticipate potential hurdles and proactively address them. Just as AVs encountered obstacles related to trust, safety, and public acceptance during their adoption, similar issues may arise as AI permeates ATC operations (Langford et al., 2022; Kang et al., 2025). A comprehensive examination of public perceptions now will inform strategic decision-making and facilitate effective implementation for the future. In this study, we will focus on three of the six automation levels (due to resource constraints), including the lowest level (Level 0), the highest level (Level 5) and the middle ground (Level 3) to understand public perceptions of automation in the contexts of ATC vs. AV.

Trust and Multidimensional Trust in Human-AI Interaction. Successfully implementing AI systems requires more than technical proficiency and strong performance metrics, such as capability and robustness. It also demands incorporating human-centred attributes like predictability, cooperativeness, and honesty. These are crucial qualities that significantly influence the trust ascribed to agents in human-AI interactions (Hancock et al., 2011; Malle and Ullman, 2021), as frequently used in the literature and commonly voted by the public. A brief explanation of these qualities (operationalised mainly based on Malle and Ullman's multidimensional trust framework) is provided below:

- *Capability* refers to whether an agent can manage the assigned task in a safe, orderly, and expeditious way under all conditions.
- *Robustness* refers to whether an agent can minimize errors or maintain an "acceptable" behaviour in the presence of exceptional or unforeseen operating conditions.
- *Predictability* refers to whether an agent can act consistently in a way that aligns with its objectives and allows other agents to have a reasonable expectation or enough understanding of its responses in similar situations.
- *Cooperativeness* refers to whether an agent can work as a team with others to achieve common operational goals.
- *Honesty* refers to whether an agent can correctly report its decisions without deceiving or withholding information from the recipients.

While capability may significantly influence trust (Kenesei et al., 2022), its impact may vary between contexts or applications. In ATC, where the stakes are high and domain specific knowledge is essential, capability might be paramount, overshadowing other factors (Organization, 2016). Whereas in vehicle driving (VD), given tasks are more accessible to the public, cooperativeness could also play a vital role—consider the dynamics of road-sharing with driverless vehicles (Baber et al., 2005; Malik et al., 2021) and the mitigation of aggressive driving behaviours (Liu et al., 2020; Craig et al., 2021). Therefore, by examining public expectations of these unique qualities for agents in both tasks (ATC vs. VD), we can better understand trust requirements and foster confidence in their operational deployment.

Blame and Trust Damage after AI Failure. In an automated system, when an error occurs how do people attribute accountability? This is a question heavily studied in the field of AV. As summarised in (Bonnefon et al., 2024), previous research found that in single-driver situations machine drivers were blamed more than humans for the same mistakes. However, in a shared-driving context, humans are often assigned more blame than their AVs (Awad et al., 2018, 2020; Copp et al., 2023; Zhai et al., 2023). As the human-AI interaction varies across the levels of automation, this disparity in blame attribution becomes more pronounced: as automation increases, the blame gradually shifts from human drivers to AVs (Franklin et al., 2021; Zhang et al., 2021; Gill, 2020; Zhai et al., 2023). The collective blame placed on other stakeholders, including government bodies and manufacturers, also increases (Bennett et al., 2020). Nonetheless, the evidence is insufficient to draw parallel conclusions for ATC, where the public are passengers rather than operators ('ATCOs' in ATC; 'drivers' in VD).

Despite the critical role of ATC in ensuring aviation safety, blame attribution within this context remains understudied, particularly regarding the integration of AI systems. While previous research has identified a bias against airlines following incidents (Dean, 2024), the impact of AI on blame attribution remains unclear until these systems are involved in the decision-making process. Furthermore, the psychological stakes of Human-AI interaction failures go beyond the allocation of blame. For instance, a malfunction in Level 5 AVs can lead to an immediate loss of trust in the car, a reaction not observed with Level 3 AVs (Seet et al., 2022). Even minor errors, such as taking too long to park, can significantly damage trust levels (Yokoi, 2024). However, little is known about the extent of trust damage in ATC, particularly in an industry with stringent safety standards and a culture that emphasises "people create safety" (Kirwan, 2024). Thus, whether such damage will be irreparable for AI in ATC is another focus in this study. Moreover, by comparing the blame attribution and trust damage between human and AI operators across different automation levels, we will better understand if public bias, such as the blame asymmetry – where AI is blamed more than humans for similar mistakes (Porsdam Mann et al., 2023; Liu and Du, 2022; Pozharliev et al., 2023) – are influenced by the role AI plays in operations.

Public Perception of AI in ATC* vs. *VD Although ATC and AV share similarities in their automation trajectories within industrial development, they differ significantly in how the public perceives them as task domains. Public trust in AI adoption can be influenced by both psychological traits of the agents (such as the trust dimensions mentioned earlier) and task characteristics (Kuper and Krämer, 2024; Kaplan et al., 2023; Si et al., 2024).

- *Familiarity and Openness*. There are far more drivers than air traffic controllers (ATCOs), and people interact with driving more often, either by driving themselves or seeing others drive. This familiarity likely contributes to higher trust in AVs over ATC (Brannon et al., 2007; Alhakami and Slovic, 1994). For example, self-reported understanding of AVs is positively associated with higher levels of trust (DeGuzman and Donmez, 2024). In contrast, driving, as a common, everyday activity that most people are familiar with, allows people to rely on their personal experience and heuristic processing to form opinions about its risks (Finucane et al., 2000). The driving environment is also much more open to external factors, like unmanaged traffic or unpredictable drivers, which can influence how risky it feels (Filos et al., 2020; Etminani-Ghasrodashti, Kermanshachi, Rosenberger and Foss, 2023; Mondy et al., 2025). On the other hand, ATC operates in a highly controlled environment with strict procedures, making it seem safer but also more complex to the public (Mogford et al., 1995; McDougall and Roberts, 2008).
- *Media Discourse Influence*. Media priming effects and biases significantly shape public opinions (Wasinger and Mann, 2020). High-profile reports can have lasting impacts on public perceptions unless corrected (Lecheler and De Vreese, 2016). For example, when news outlets report crashes involving AVs, public trust in AVs diminishes.
- *Stakes and Uncertainty*. The perceived trade-off between risks and benefits can strongly influence public acceptance of new technologies, such as AVs (Li et al., 2024). The significant impact AI can have therefore raises the bar for how understandable it must be (Nussberger et al., 2022). Agent-related features, like explainability, become especially important in high-stakes decisions—such as when an algorithm recommends a medical treatment, compared to lower-stakes contexts like film suggestions (Lombrozo, 2006). Another key factor in decision-making is uncertainty: the more uncertain a task, the more careful we need to be (Gigerenzer et al., 2022).
- *Expected Safety and Expertise*. While safety is the top priority in both ATC and VD, particularly in ATC—an industry with strict safety protocols—psychological biases can lead to unrealistic safety expectations among the public (Fischhoff et al., 1978; Shariff et al., 2021). These expectations may inadvertently compromise the benefits that technology can offer society. Research has already shown that AVs need to be two to five times safer than human-driven vehicles across various conditions for people to accept their potential risks (Liu et al., 2019). Furthermore, compared to driving, a skill that most people can master in a few months, ATCOs undergo rigorous selection processes and years of specialised training to obtain their licences for operation.

Research questions

This study draws from the rich literature on perception in the AV context to investigate the safety–critical context of ATC. This includes not only establishing a reference point of trust and blame attribution—using their underlying similarities in the industrial trajectory of automation, but also examining the unique features of each that shape these attitudes. Building on this, we explore public perceptions of AI operations in both sectors, with the aim to provide contextually informed evidence. This evidence can guide industry stakeholders and government bodies in facilitating the safe deployment of advanced technology for the benefit of society. Therefore, with key study variables presented in Table 1, our research questions are listed as the following:

Research Question 1: How do the public attribute trust and blame in ATC compared to AV?

Research Question 2: How do agent types (Human vs. AI operator) and Human-AI Interaction models (3 Automation levels) affect these trust and blame attributions?

Research Question 3: How do public perception of agents' characteristics (5 trust dimensions: capability, robustness, predictability, honesty, and cooperativeness) differ between ATC and AV?

Research Question 4: How do public perceptions of tasks' characteristics (Familiarity and Openness, Media Discourse, Stakes and Uncertainty, and Expected Safety and Expertise) vary between ATC and VD?

Methods

Participants

Participants were recruited through the crowd-sourcing platform

Table 1
Study variables.

	Between-Subject Variable		
Human-AI Interaction (HAI) Models	Model a: Level 0 AI Tool (L0) Model b: Level 3 AI Trainee (L3) Model c: Level 5 AI Manager (L5)		
	Within-Subject Variable		
Task (Context) Types	Air Traffic Control (ATC) Vehicle Driving (VD) or Autonomous Vehicle (AV)		
Agent Types	Human Operator AI Operator		
	Dependent Variables		
	Attributions:	*Agent Characteristics:*	*Task Characteristics:*
Study 1 (Convenience Sample)	Initial Trust Restored Trust Blame	Capability Robustness Predictability Cooperativeness Honesty	Familiarity Expertise Tech Awareness Openness Media Discourse Stake Uncertainty Positive Safety Negative Safety
Study 2 (Nat. Rep. Sample)	Initial Trust Restored Trust Blame	N/A	N/A

Note. Operator: Air Traffic Control Officer (ATCO) in ATC vs. Driver in VD and AV; Nat. Rep. Sample: nationally representative sample.

Prolific and surveyed using the survey platform *Qualtrics*. Both platforms are widely recognised in research for their ability to facilitate effective testing with reliable samples (Palan and Schitter, 2018). Initially, we collected a convenience sample of 321 UK residents. This was followed by the collection of a nationally representative (Nat. Rep.) sample of 1,042 UK residents by age, gender and race, automatically weighted by Prolific based on census data from the UK Office of National Statistics (Prolific, 2024). The final size of the convenience sample was 301, with a passing rate of 93.77 % for attention check. The final Nat. Rep. sample consisted of 988 participants, with a passing rate of 94.82 %. Details of the sample demographics are provided in Table 2.

Study design and procedures

Study Design. This study employed a mixed design, combining a 3 (HAI models) × 2 (Context: ATC vs. VD) × 2 (Agent: human vs. AI) framework. Participants were first randomly assigned to one of three Human-AI interaction (HAI) models and then exposed to two contextual manipulations (randomly ordered) involving both human and AI operators.

The three HAI models: (1) Model a: Level 0–AI Tool, (2) Model b: Level 3–AI Trainee, and (3) Model c: Level 5–AI Manager, outlined task assignments for both human and AI operators. Participants first read the HAI description (as shown in Fig. 1); comprehension checks were then used to ensure the understanding of role assignment under each HAI model. In the comprehension checks, participants were asked to identify who has the control right in each HAI, which the correct answer is human in the AI tool and AI trainee models, and AI in the AI manager model.

Study 1 The convenience sample then rated trust, blame, restored trust and 5 other qualities for both human and AI operators on a 10-point Likert scale (1 [not at all] – 10 [very much]). Before answering the restored trust questions, participants were provided with additional information about the failure corresponding to each HAI model, which had been fixed. Under the AI-tool model, the failure was caused by the human operator's misjudgement based on the AI's miscalculation. Under the AI-trainee model, the human supervisor failed to intervene when AI trainee made a misjudgement. Under the AI-manager model, the AI manager made a misjudgement which resulted in a safety risk. The questions used are as the follows: *(1) how much do you trust the operator? (2) how much would you blame the operator for the failure? (3) After the failure, how much would you still trust the operator? (4) how capable, robust, predictable, honest or cooperative do you think the operator is?*

After the trust and blame attribution, participants were also asked about their general perception of the two tasks (ATC vs. VD, randomly presented), using the same scale (see Table 3).

Study 2 The Nationally Representative Sample were only asked three specific questions related to trust, blame, and restored trust, following the same experimental design as in Study 1.

During the debriefing, participants' general demographic information, including age, gender, ethnicity, education level, socioeconomic status was collected. Their experiences of flying and driving such as fly frequency and driving frequency were also asked to understand the sample profile better. Then a debriefing message was presented to provide more explanation of the study. An open text box was also provided for participants to enter their feedback or any other comments at the end of the study.

Prior to taking part in the study, informed consent was obtained from all participants. Participants were also debriefed at the end of the study, with opportunities to provide feedback. The average completion time of the study is approx. 7 min and £1 was paid to each participant for their time.

Table 2
Sample demographics.

Demographic Characteristics		*Study 1* Convenience (N = 301) n (%) = Number (Frequency)	*Study 2* Nat. Rep. (N = 988)
Age	Mean (SD)	42 (14.1)	46.59 (15.9)
	18–24	31 (10 %)	107 (11 %)
	25–34	94 (31 %)	187 (19 %)
	35–44	47 (16 %)	147 (15 %)
	45–54	56 (19 %)	161 (16 %)
	55–64	58 (19 %)	258 (26 %)
	65–74	13 (4.3 %)	103 (10 %)
	75–84	2 (0.7 %)	25 (2.5 %)
	85–100	0 (0 %)	0 (0 %)
Gender	Male	112 (37 %)	478 (48 %)
	Female	186 (62 %)	503 (51 %)
	Other	3 (1.0 %)	7 (0.7 %)
Ethnicity	White	258 (86 %)	844 (85 %)
	Black	16 (5.3 %)	38 (3.8 %)
	Asian	19 (6.3 %)	73 (7.4 %)
	Mixed	4 (1.3 %)	16 (1.6 %)
	Other	4 (1.3 %)	17 (1.7 %)
Education	High School or lower	51 (17 %)	186 (19 %)
	Vocational or College	76 (25 %)	206 (21 %)
	Undergraduate	109 (36 %)	382 (39 %)
	Postgraduate	64 (21 %)	210 (21 %)
	Other	1 (0.3 %)	4 (0.4 %)
Socioeconomic status (SES)	Mean (SD)	5.27 (1.51)	5.60 (1.53)
	LowSES	45 (15 %)	113 (11 %)
	MediumSES	248 (82 %)	805 (81 %)
	HighSES	8 (2.7 %)	70 (7.1 %)
FlyFrequency	Never	31 (10 %)	96 (9.7 %)
	Rarely	140 (47 %)	453 (46 %)
	Sometimes	89 (30 %)	303 (31 %)
	Frequently	38 (13 %)	124 (13 %)
	Veryfrequently	3 (1.0 %)	12 (1.2 %)
DrivingFrequency	Never	5 (1.7 %)	16 (1.6 %)
	Rarely	16 (5.3 %)	25 (2.5 %)
	Sometimes	30 (10.0 %)	87 (8.8 %)
	Frequently	54 (18 %)	180 (18 %)
	Veryfrequently	196 (65 %)	680 (69 %)

Note. Convenience: The Convenience Sample; Nat. Rep.: The Nationally Representative Sample; Other: participants use other more self-evident labels.

Results

Trust (damage) and blame attribution

To answer research questions 1 and 2, we employed regression models to examine the main effects of agent type (AI is compared with the baseline human) and Human-AI Interaction (HAI) (Level 0–AI tool as the baseline) on trust and blame attribution, using the Nat. Rep. sample from study 2. As shown in Table 4, AI is significantly trusted less and blamed more than human in both Air Traffic Control (ATC) and Vehicle Driving (VD). HAI also affects these attributions in a more complex way. Similar results were observed in study 1, using the convenience sample (detailed results can be found in the Appendices A and B).

Specifically, for initial trust, human is trusted more than AI in all three levels as displayed in parts A and C of Fig. 2. Moreover, trust is lowest for the AI Trainee compared to the AI Tool in ATC ($\beta = -0.26$, $SE = 0.01$, $P < 0.05$). Meanwhile, trust is lowest for the AI Manager than the AI Tool in VD ($\beta = -0.32$, $SE = 0.11$, $P < 0.01$).

For blame, as shown in the parts B and D of Fig. 2, human is blamed more than AI only in the AI-trainee model, while they are assigned less blame in the other two models. There is little difference observed between the AI Tool and the AI Manager model in terms of blame to human or AI. These patterns are consistent across ATC and VD.

Finally, restored trust, which is the newly rated trust after an initial mistake has been fixed, is lower than initial trust for all automation

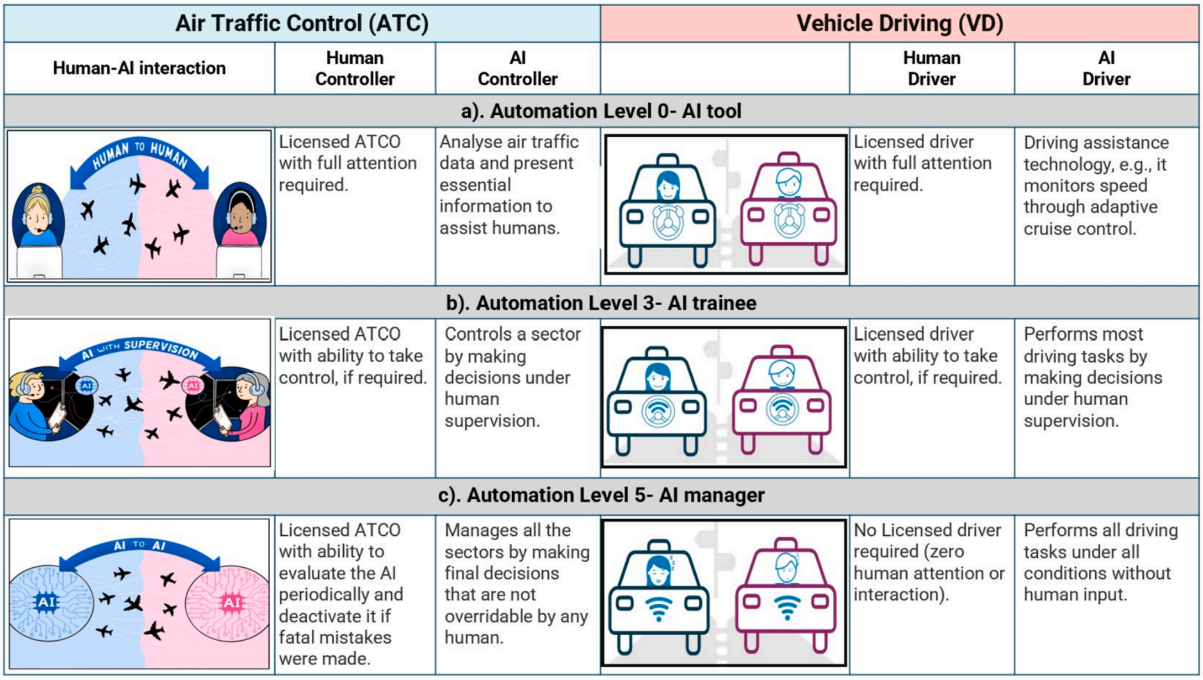


Fig. 1. Participants were exposed to one of three levels of Human-AI Interaction (HAI) in Air Traffic Control and Vehicle Driving. Note. These levels were (a) Level 0: No automation (AI Tool). The AI is used as a tool for information analysis. (b) Level 3: Conditional automation (AI Trainee). The AI takes actions under human supervision. (c) Level 5: Full automation (AI Manager). The AI is in control of everything with no human intervention.

Table 3
Task characteristics measurement.

Measure	*Scale*	*Description*
1a. Familiarity	1–10	How familiar are you with … (ATC vs. VD)?
1b. Openness	1–10	How open is the environment to external factors in … (ATC vs. VD)?
2a. Media Discourse	1–10	How likely do you think you'd hear about it on the news, if there is a fatal incident that happened about … (ATC vs. VD)?
2b. Tech Awareness	1–10	How aware are you of AI operating … (ATC vs. VD)?
3a. Stakes	1–10	For each task listed below (ATC vs. VD), imagine the worst thing that could happen and how it would affect everyone. How bad do you think the outcome would be?
3b. Uncertainty	1–10	1. How likely do you think that future events can be predicted from previous data, when AI is fully deployed in (ATC vs. VD)? 2. How easy do you think it would be to list or characterize all possible situations that an operator may encounter in (ATC vs. VD)?
4a. Positive Safety	0–100 %	How much safer than the average human would an AI operator need to be before you would choose to travel with it?
4b. Negative Safety	0–100 %	How much safer than the average human would an AI operator need to be before you would choose to travel with it?
4c. Expected Expertise	Years	How many years of experience do you believe are necessary to become a proficient (ATCO VS. Driver)?

Note. 1–10 (*not at all – very much*); ATC: air traffic control; VD: vehicle driving; In 4a, 0% means not safer than any human operator – 100% means safer than all human operators; In 4b, 0% means the AI operator has the same incident rate as humans – 100% means eliminating all the incidents that average operators make in the UK.

levels for both human and AI. Trust damage (Initial Trust minus Restored Trust) is higher for human than for AI, particularly in ATC. Interestingly, despite loss of trust in human, it is still higher than (initial) trust in AI across all three levels.

Table 4
Regression results for trust and blame attributions.

	Air Traffic Control			*Vehicle Driving*		
	Initial Trust	*Blame*	*Restored Trust*	*Initial Trust*	*Blame*	*Restored Trust*
Agent *[AI]*	−2.50 (0.08) ***	0.54 (0.11) ***	−2.30 (0.10) ***	−1.80 (0.09) ***	0.70 (0.11) ***	−2.00 (0.10) ***
HAImodel *[AItrainee]*	−0.26 (0.10) *	0.26 (0.14) p = 0.06	−0.23 (0.13) p = 0.09	−0.05 (0.11) p = 0.63	0.14 (0.14) p = 0.32	0.02 (0.14) p = 0.86
HAImode *[AImanager]*	−0.14 (0.10) p = 0.13	0.03 (0.13) p = 0.80	0.08 (0.13) p = 0.53	−0.32 (0.11) **	−0.30 (0.14) *	0.06 (0.14) p = 0.66
Constant	8.80 (0.08) ***	7.00 (0.11) ***	7.30 (0.11) ***	7.70 (0.09) ***	7.00 (0.11) ***	6.70 (0.11) ***
Observations	1,976	1,976	1,976	1,976	1,976	1,976

Note. ***,**,* significant at 0.001, 0.01, and 0.05 levels, respectively. The human operator is the baseline for Agent; the AI tool model is the baseline for HAI model; The Nat. Rep. sample from study 2 was used.

In general, humans are generally trusted more than AI across all three HAI levels. However, AI is often blamed more, except when humans are in supervisory roles, as in L3. When errors occur, even after they have been fixed, trust in humans suffers significantly, particularly in L3. For AI, the loss in trust is less pronounced given its inherently lower trust levels. Patterns of attribution in the VD context are consistent with those observed here, underscoring similar dynamics in human-AI interactions. Notably, the level of trust placed in ATC for humans tends to be greater than that in VD.

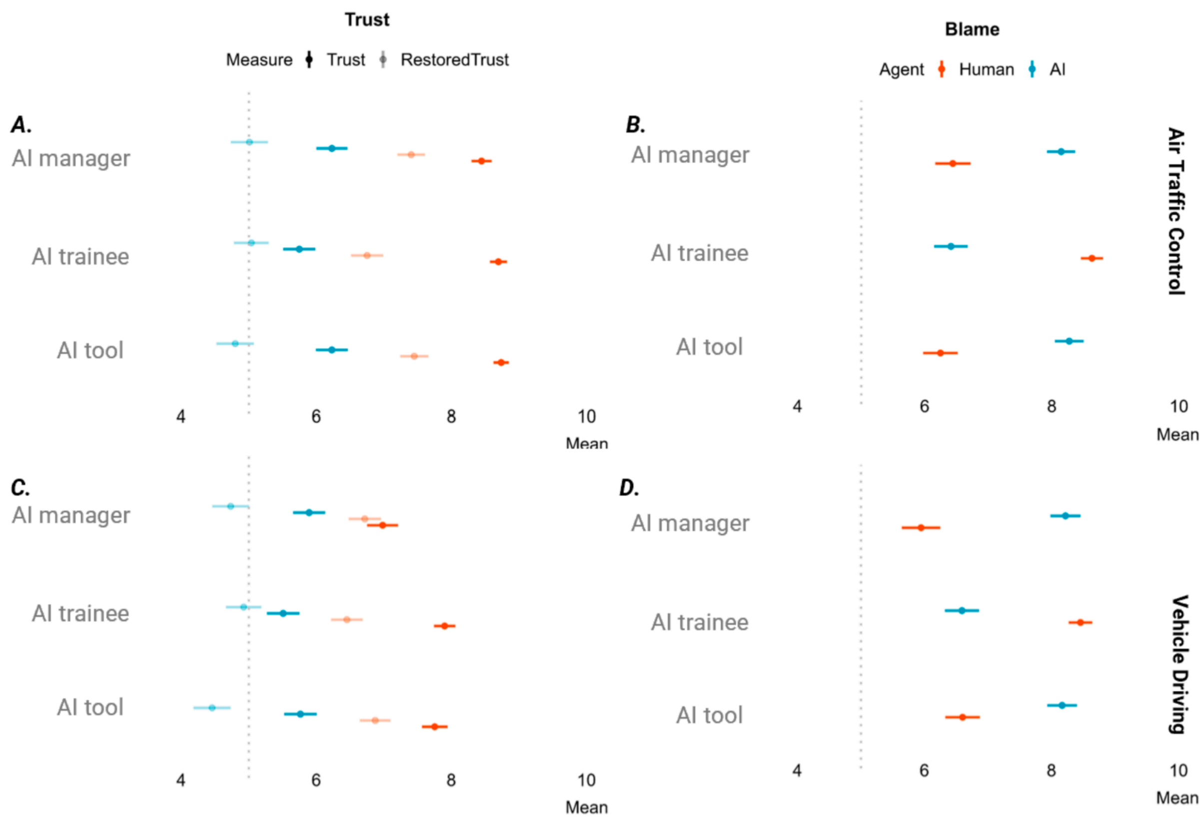


Fig. 2. Attribution (Study 2) of trust, blame, and restoration of trust after fixed mistake (using a 10-point Likert scale) in Human vs. AI operators across three Human-AI Interaction (HAI) models: (A) Trust and Restored Trust in ATC; (B) Blame in ATC; (C) Trust and Restored Trust in VD; (D) Blame in VD. Each plot includes three HAI models: L0-AI tool, L3-AI trainee, and L5-AI Manager, from the bottom to the top. Note. The dots on the solid lines show the mean difference with the accompanying error bars delineating the 95% confidence intervals for these means. The dashed lines show the average attributions assigned to human and AI operators combining the 3 HAI models. Mean ratings for human operators are shown in orange, while those for AI appear in green. The colours for restored trust (after a mistake) are intentionally lighter, emphasising the decrease in trust following errors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Perception of ATC and VD as tasks

In order to understand research questions 3 and 4, how do the public perceive Air Traffic Control (ATC) vs. Vehicle Driving (VD), we compare public perception of the two tasks in 10 task characteristics and 5 agent features, using the convenience sample from Study 1 (no relevant data was collected from the Nat. Rep. sample in Study 2 due to resource constraints). The ten task attributes assessed include familiarity, required expertise, technology awareness, openness, media discourse, stake, uncertainty, and positive/negative safety, as shown from top to bottom in the left panel in Fig. 3.

Regression results in Table 5 (also depicted on the left side of Fig. 3) indicate that significantly higher expertise ($\beta = 2.30$, $SE = 0.14$, $p < 0.001$) and safety are expected for ATC compared to VD, including both positive ($\beta = 3.10$, $SE = 1.10$, $p < 0.05$) and negative ($\beta = 3.40$, $SE = 1.10$, $p < 0.05$) safeties. ATC is also perceived to have a significantly higher stake ($\beta = 1.50$, $SE = 0.12$, $p < 0.001$), and uncertainties than VD ($\beta = 0.33$, $SE = 0.01$, $p = 0.001$; $\beta = 0.29$, $SE = 0.11$, $p < 0.05$). Although both contexts have high media discourse, ATC at-tracts more attention than VD ($\beta = 0.52$, $SE = 0.14$, $p < 0.01$). However, the public is much more aware of AI operating in VD ($\beta = -2.70$, $SE = 0.15$, $p < 0.001$) and considers it to be more open than ATC ($\beta = -1.40$, $SE = 0.12$, $p < 0.001$). Despite this, familiarity with VD is not significantly higher than with ATC ($\beta = -0.19$, $SE = 0.12$, $p = 0.14$).

The 5 agent attributes measured are capability, robustness, predictability, honesty, and cooperativeness for both human and AI operators (detailed in Table 6 and depicted on the right side of Fig. 3). In the context of ATC, human operators are perceived to have higher levels in most attributes compared to AI, particularly in capability ($\beta = -1.70$, $SE = 0.13$, $p < 0.001$) and robustness ($\beta = -1.40$, $SE = 0.15$, $p < 0.001$), with similar patterns observed in the VD context. Additionally, the overall attribution across the five dimensions is higher in ATC than in VD. Furthermore, the patterns for agent characteristic remain stable across the three HAI levels while there is a noticeable decline in these qualities at the L5 level, with Capability, Robustness and Cooperativeness being most affected (see Appendix C).

Overall, Fig. 3 reveals that while there are similarities in perceptions between ATC and VD, notable differences exist in areas such as expertise, technological awareness, openness and stakes. The expertise required for ATC, along with the associated media discourse and stakes, are markedly higher compared to VD. Conversely, greater levels of technology awareness and system openness are perceived in VD than ATC. Trust in ATC and VD shows agent-specific patterns, with humans generally perceived as more capable, robust, and cooperative, especially in ATC. Moreover, these attributes are invariably rated higher in ATC than in VD.

Heterogeneity analysis

In order to validate the robustness of the trust and blame attributions across different demographic groups, we conducted a heterogeneity analysis using the Nat. Rep. sample from Study 2. This sample was grouped by 7 demographic factors, including age, gender, ethnicity, education, socioeconomic status, fly frequency, and driving frequency. Across these 20 demographic subgroups (with more than 30 participants in each), we compared people's attitudes toward human and AI

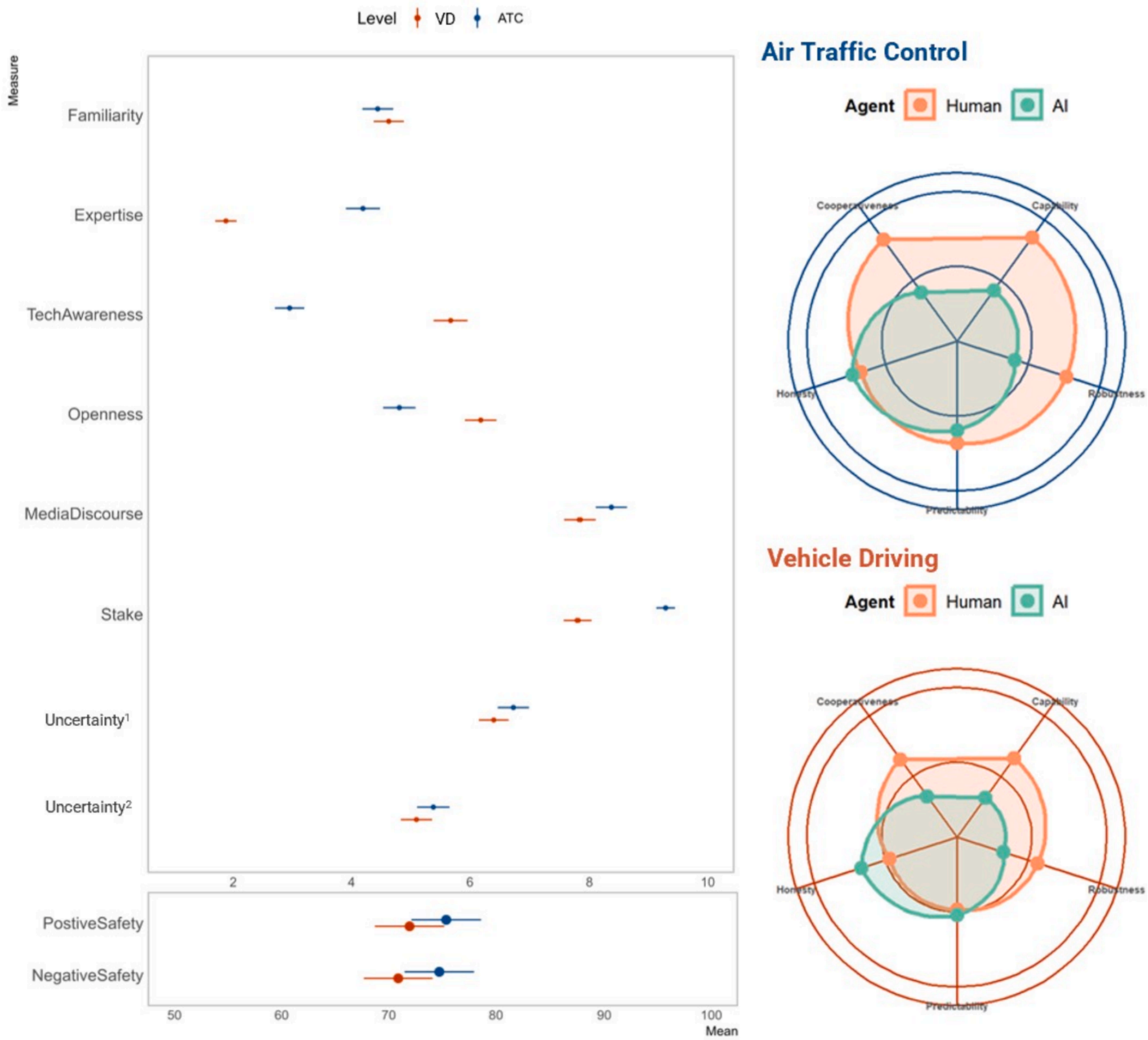


Fig. 3. Perception differences Compared between Air Traffic Control (ATC) and Vehicle Driving (VD). Note. Red colour represents VD and blue for ATC. The dots on the solid lines show the mean difference with the accompanying error bars delineating the 95% confidence intervals for these means. The spider plots reveal consistent patterns across three HAI levels, thus, only the contrasts between human and AI contributions are depicted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Regression results for task characteristics.

	Familiarity	*Expertise*	*Tech Awareness*	*Openness*	*Media Discourse*
Context [ATC]	−0.19 (0.12) p = 0.14	2.30 (0.14) ***	−2.70 (0.15) ***	−1.40 (0.12) ***	0.52 (0.14) **
Constant	4.60 (0.13) ***	1.90 (0.12) ***	5.70 (0.14) ***	6.20 (0.14) ***	7.80 (0.13) ***
Observations	602	427	602	602	602
	Stake	*Uncertainty1*	*Uncertainty2*	*Positive Safety*	*Negative Safety*
Context [ATC]	1.50 (0.12) ***	0.33 (0.10) **	0.29 (0.11) *	3.10 (1.10) *	3.40 (1.10) *
Constant	7.80 (0.10) ***	6.40 (0.13) ***	5.10 (0.14) ***	72.00 (1.60) ***	71.00 (1.60) ***
Observations	602	602	602	537	542

Note. ***,**,* significant at 0.001, 0.01, and 0.05 levels, respectively; Vehicle driving(VD) is the baseline for Context.

operators. Twelve additional subgroups (such as other genders, ethnicities, education levels, flying very frequently, never driving, etc.) were excluded from the analysis due to small sample sizes and limited clarity. Similar to Fig. 2 displayed above, Fig. 4 provides a detailed breakdown of trust, blame, and restored trust attribution, from the left panel to the right, within each subgroup.

In the context of Air Traffic Control (ATC), humans consistently receive higher levels of trust than AI across 20 subgroups (see the left

Table 6
Regression results for agent characteristics.

	Capability	*Robustness*	*Predictability*	*Honesty*	*Cooperativeness*
In the context of Air Traffic Control					
Agent [AI]	−1.70 (0.13) ***	−1.40 (0.15) ***	−0.35 (0.15) *	−1.70 (0.16) ***	0.24 (0.16) p = 0.13
Constant	8.40 (0.10) ***	8.10 (0.11) ***	7.70 (0.11) ***	8.30 (0.12) ***	7.70 (0.11) ***
Observations	602	602	602	602	602
In the context of Vehicle Driving					
Agent [AI]	−1.30 (0.14) ***	−0.97 (0.16) ***	0.14 (0.17) p = 0.39	−1.20 (0.16) ***	0.80 (0.17) ***
Constant	7.60 (0.11) ***	7.30 (0.12) ***	6.90 (0.12) ***	7.60 (0.13) ***	6.90 (0.12) ***
Observations	602	602	602	602	602

Note. ***,**,* significant at 0.001, 0.01, and 0.05 levels, respectively; the human operator is the baseline for Agent.

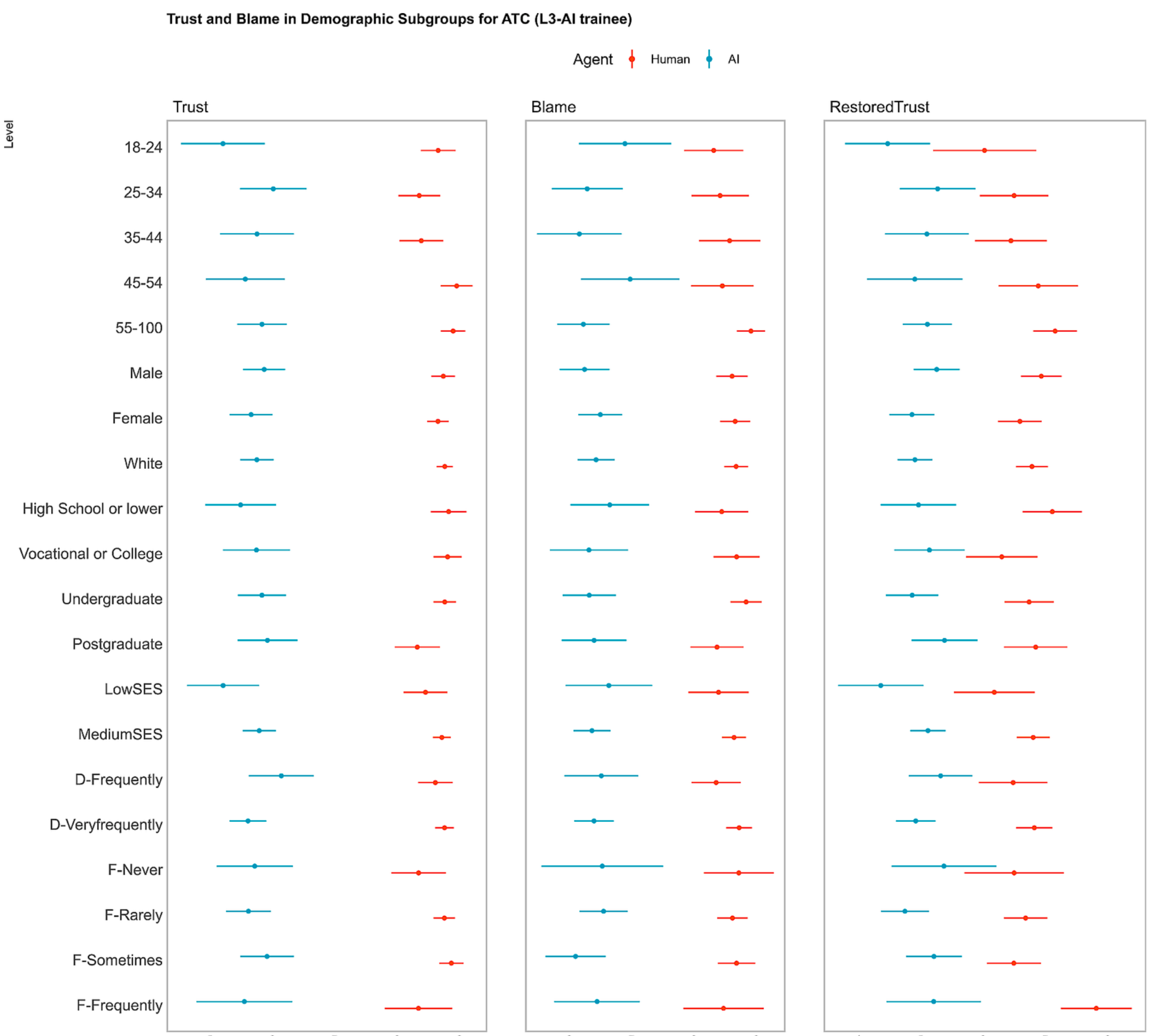


Fig. 4. Trust, Blame, Restored Trust Attribution in Demographic Subgroups for ATC in L3-AI trainee. Note. 'DiffTrust' refers to the difference between the Initial Trust and Restored Trust; 'D-' refers to driving; 'F-' refers to flying (as a passenger). For clarity in visualisation, only groups with more than 30 participants were included.

panel), even when considering restored trust after a mistake (see the right panel). However, blame attribution presents a more complex story, varying by HAI models. As illustrated in Fig. 2, AI operators are blamed more than human operators in the L0-AI tool and L5-AI manager models, but less so in the L3-AI trainee model. To clearly highlight this variation, Fig. 4's middle panel focuses exclusively on blame attribution within the L3-AI trainee model. It further supports this pattern, revealing that most subgroups only assign less blame to AI in L3. The greater blame assigned to AI in models L0 and L5 is demonstrated in the Appendix D (top panel). These attributions showed a very similar pattern in Vehicle Driving (VD) and are detailed in the Appendix D (bottom panel) too.

Discussion

In this paper, we employed a convenience sample in Study 1 and a Nat. Rep. sample in Study 2 to explore the public's trust and blame attribution in Air Traffic Control (ATC) compared to Vehicle Driving (VD). Results were consistent across both studies. Specifically, we focused on the effects of agent types and Human-AI Interaction (HAI) models. Additionally, the convenience sample from Study 1 was used to further investigate how task and agent characteristics differ between the two contexts. The results revealed that people tend to trust human operators more while assigning more blame to AI operators, in both ATC and VD. Interestingly, the Level 3 AI trainee model showed an opposite pattern. These results provide evidence-based insights into public perceptions of AI integration in the critical field of ATC compared to VD. This sheds light on the potential areas of concern and acceptance among the public, thereby guiding the effective implementation of automation in aviation for industries and policymakers.

Our findings reveal that the type of agent significantly influences public trust and blame, both in ATC and VD contexts. It supports the extensive literature documenting the human-AI asymmetry (Porsdam Mann et al., 2023; Liu and Du, 2022; Pozharliev et al., 2023)–human operators consistently receive greater trust but face less blame than AI operators in most conditions (Zhang et al., 2024). This preference for humans over AI also aligns well with the algorithm aversion effect, where people are biased against AI regardless of its performance (Burton et al., 2020). Furthermore, our results also expand this effect across various contexts and automation levels.

When considering how context affects people's trust and blame attribution, interestingly, algorithm aversion appears even more pronounced in more high-stakes and high-uncertainty contexts like ATC. Jussupow et al. summarised several algorithm and human characteristics that contribute to this effect, including agency, performance, perceived capability, expertise, and social distance (Jussupow et al., 2020). ATC in particular, underscores most of these factors. For example, expertise and performance significantly impact how agents are appreciated by the public (Hou and Jung, 2021; Bang and Frith, 2017; Yalcin et al., 2023). Indeed, domain-specific knowledge is a necessity for safe ATC operations, and licensed ATCOs clearly demonstrate this expertise, leading to greater approval by the public.

When participants were asked about scenarios where agents made errors but subsequently fixed them, they reported lower levels of trust for both humans and AI compared to their initial trust levels. AI, in particular, suffers more from this trust damage following the effect of algorithm aversion (Dietvorst et al., 2015). This reflects the stronger negative responses people tend to associate with algorithms (Renier et al., 2021). However, the variations in the trust and blame attribution reveals a more complex case under the scope of HAI models.

Human operators are not merely trusted more or blamed less than AI operators in a general sense. They are particularly preferred when AI is less capable or responsible for more advanced tasks (Bigman and Gray, 2018). For instance, in level 0 automation, where humans are responsible for everything and AI serves as a basic calculation tool, algorithm aversion is particularly pronounced, resulting in the low levels of trust in AI. This human-in-the-loop system places humans in control of decision-making and yields better performance from human operators (Agudo et al., 2024). It also leads to more blame being directed at the AI operator when it provides incorrect calculations to humans.

Similar attribution patterns appear at the other end of the automation spectrum, where AI acts as the manager, capable of all tasks and in charge of everything. In this model, humans are taken out of the loop. Yet, we observe greater trust in humans and more blame assigned to AI (Green and Chen, 2019). This preference for humans may emerge from completely different reasons than those seen in level 0 automation. While much of the appreciation for humans comes from perceived capability, it's inevitable that AI will eventually outperform humans at this full automation level. The aversion to AI at level 5 may instead stem from the greater social distance people feel toward AI operators (Trope and Liberman, 2010; Jussupow et al., 2020), especially when AI is seen as replacing humans entirely.

Only at level 3, where AI matches human performance as a trainee but requires supervision from human operators, does it receive less blame. This human-on-the-loop system highlights the oversight role of humans and limits AI's autonomy in decision-making process (Ivanov, 2023; Janotta and Hogreve, 2024). At this juncture, which bridges the two ends of the automation spectrum—from no automation to full automation—the public may move beyond gut reactions and engage in justice cognition. They consider humans as the ultimate decision-makers, and recognise their greater accountability for shared actions (Renier et al., 2021). This aligns well with the literature on AVs, where human drivers are often blamed more than machine drivers (Awad et al., 2020; Wotton et al., 2022; Bennett et al., 2020).

Furthermore, public perceptions of human and AI operators in the contexts of ATC and VD complement the two key effects related to agents and HAI models. Human operators are not only trusted more; they are also perceived as more capable and robust compared to AI in both ATC and AV. This multifaceted nature of trust underscores the complex factors contributing to the appreciation of humans, especially when interacting with AI (Lankton et al., 2015; Malle and Ullman, 2021). In addition to agent characteristics, public attitudes are influenced by task-specific features. For example, ATC is associated with higher expertise and stakes, whereas VD is more open to outsiders and visible to the public. Considering these dynamic tasks, these differences collectively position humans as all-around winners.

Implications

Our findings highlight the important role of public perceptions in the adoption of automation technologies. Although technological advances are central to this transition, how the public views and evaluates these technologies strongly influences their acceptance across different contexts.

First of all, for researchers and the wider designing communities of automation for ATC, technology functionality is essential for development (Vongvit et al., 2024), but it's not an isolated concept. Without good performance, trust in AI is delicate and can be easily damaged after a mistake. AI operators should match human performance not only in capability but also in qualities that support robustness and facilitate human-AI collaboration (Kaplan et al., 2023). Interestingly, AI sometimes is perceived as more cooperative than humans in AVs, highlighting an opportunity for improvement based on its inherent strengths.

Secondly, for policy makers, automation progression should not take place solely at the pace of technological development. Thoughtful assessment and consideration are necessary regarding the impact on human operators, industries, and society. As automation increases, new challenges will arise in human-AI interactions, making role assignment critical. A well-balanced arrangement, allowing AI to excel in its strengths while supporting humans in shared tasks, will promote technology adoption (Walsh and Feigh, 2021; Ivanov, 2023; Kumar et al., 2024). Additionally, public attitudes toward technology are not purely based on metrics. Broader social factors outside the lab also influence their judgement (Araujo et al., 2020). Industries should value openness and transparency in their operations to foster acceptance of new technology (Grant et al., 2023).

Limitations and future research

This study has several limitations that need to be address in future studies. Firstly, while we represent the two ends of the automation spectrum and the midpoint of this progression, our study only tested three automation levels due to resource constraints. The other three automation levels detailed in SAE's and SESAR's taxonomies for AV and ATC, respectively—namely, Level 1 for Decision Support, Level 2 for Partial Automation, and Level 4 for High Automation (SAE, 2018; SESAR Joint Undertaking, 2020)—were not included. Future research is required to fully explore the impact of the complete set of HAI models, identifying clear patterns in attitude changes associated with each model as automation progresses. For example, we need to investigate whether trust or blame consistently changes with higher levels of automation. Additionally, the AI failures explored in this paper are not exhaustive and reflect only one possible scenario based on the corresponding HAI model. We exemplified these failures by the key responsibilities assigned to human and AI operators at each automation level. However, other errors may occur during the process of human-AI teaming.

Moreover, although we recruited a nationally representative sample from the UK, our findings are still limited to this one country. In Study 2, the Nat. Rep. sample only answered three questions related to trust and blame. Thus, the results regarding public perceptions of task and agent characteristics should be further validated in future studies. Additionally, the Nat. Rep. sample was collected through Prolific with limited intervention from the researchers. While Prolific has been recognised as a reliable and effective platform for online human-subjects research (Douglas et al., 2023), it is also recommended for future research to consider multiple approaches and different demographic stratifications (e.g., incorporating political affiliation).

Finally, our study relied on a vignette and question-based method, which facilitated the presentation of future technology scenarios and data collection. However, this approach is limited to self-reported preferences. Future studies should explore other ways to capture trust and blame, such as behavioural experiments or neuroimaging (McEvily and Tortoriello, 2011; Ajenaghughrure et al., 2020). In addition, beyond the public as end users, industry (companies, regulators, etc.) and operators (e.g., ATCOs for ATC and drivers for AV) are also key stakeholders. It is therefore important for future research to explore their perspectives on AI in ATC and AV, and to broaden our understanding of its implications across other transport safety systems, such as railway (Jansson et al., 2023; Kusumastuti et al., 2025) and maritime management (Baum-Talmor and Kitada, 2022; Morariu et al., 2025).

Authors' contributions

P.M. and E.A. conceptualised and designed the study. P.M. acquired, analysed and visualised the data. P.M. drafted the initial manuscript. P. M., E.A., R.C., J.E. and P.L. reviewed and edited the final manuscript. All authors approved the final version for submission.

Funding source

The work described in this article is primarily funded by the grant "EP/V056522/1: Advancing Probabilistic Machine Learning to Deliver Safer, More Efficient and Predictable Air Traffic Control" (aka Project Bluebird) under EPSRC Prosperity Partnership between NATS, Turing, and Exeter.

CRediT authorship contribution statement

Peidong Mei: Writing – review & editing, Visualization, Investigation, Data curation, Writing – original draft, Conceptualization, Methodology, Formal analysis. **Richard Cannon:** Writing – review & editing, Funding acquisition. **Jim Everett:** Writing – review & editing, Methodology. **Peng Liu:** Writing – review & editing. **Edmond Awad:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition.

Ethics approval

The project has been approved by the Ethics Committee of the Faculty of Environment, Science, and Economy (FESE) at the University of Exeter, UK.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: E. A., P.M., and R.C. acknowledge funds by the grant "EP/V056522/1: Advancing Probabilistic Machine Learning to Deliver Safer, More Efficient and Predictable Air Traffic Control" (aka Project Bluebird) under EPSRC Prosperity Partnership between NATS, Turing, and Exeter. R.C.is an aerospace engineer and a Research & Development lead employed by NATS, a major air traffic control service provider in the UK. The remaining authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

For the purpose of open access, E.A has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission. The authors thank Marc Thomas, Ben Carvell, Andy Pacy and Tama Lwambula for providing valuable ATC knowledge that was helpful in the experimental design of the study. The authors also thank Tama Lwambula for providing valuable comments on an earlier version of the paper.

Appendix A. Regression for trust and blame attributions in Study 1

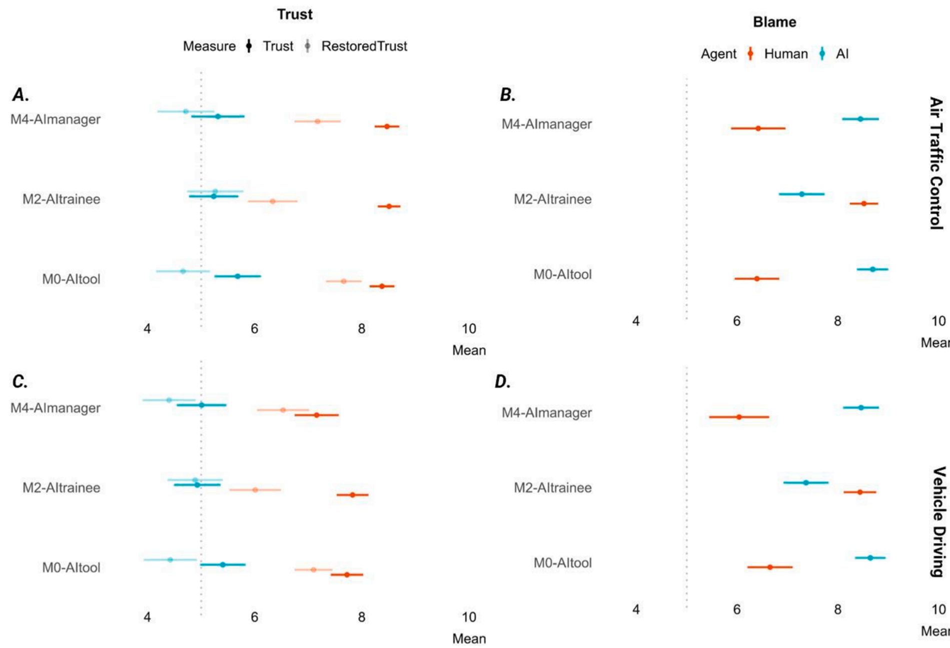
Regression results for trust and blame attributions

	Air Traffic Control			*Vehicle Driving*		
	Initial Trust	*Blame*	*Restored Trust*	*Initial Trust*	*Blame*	*Restored Trust*
Agent *[AI]*	−3.00 (0.15) ***	1.10 (0.18) ***	−2.20 (0.16) ***	−2.40 (0.16) ***	1.10 (0.19) ***	−2.00 (0.16) ***
HAI model *[AI trainee]*	−0.16 (0.19) $p = 0.40$	0.36 (0.22) $p = 0.11$	−0.36 (0.28) $p = 0.20$	−0.19 (0.21) $p = 0.36$	0.26 (0.23) $p = 0.27$	−0.31 (0.28) $p = 0.26$
HAI model *[AI manager]*	−0.14 (0.18) $p = 0.46$	−0.11 (0.22) $p = 0.63$	−0.22 (0.28) $p = 0.43$	−0.48 (0.20) **	−0.40 (0.23) $p = 0.08$	−0.30 (0.28) $p = 0.29$
Constant	8.60 (0.15) ***	7.00 (0.18) ***	7.30 (0.21) ***	7.80 (0.17) ***	7.10 (0.19) ***	6.80 (0.21) ***
Observations	602	602	602	602	602	602

Note. ***,**,* significant at 0.001, 0.01, and 0.05 levels, respectively; the human operator is the baseline for Agent; the AI tool model is the baseline for HAI model; the convenience sample from Study 1 was used.

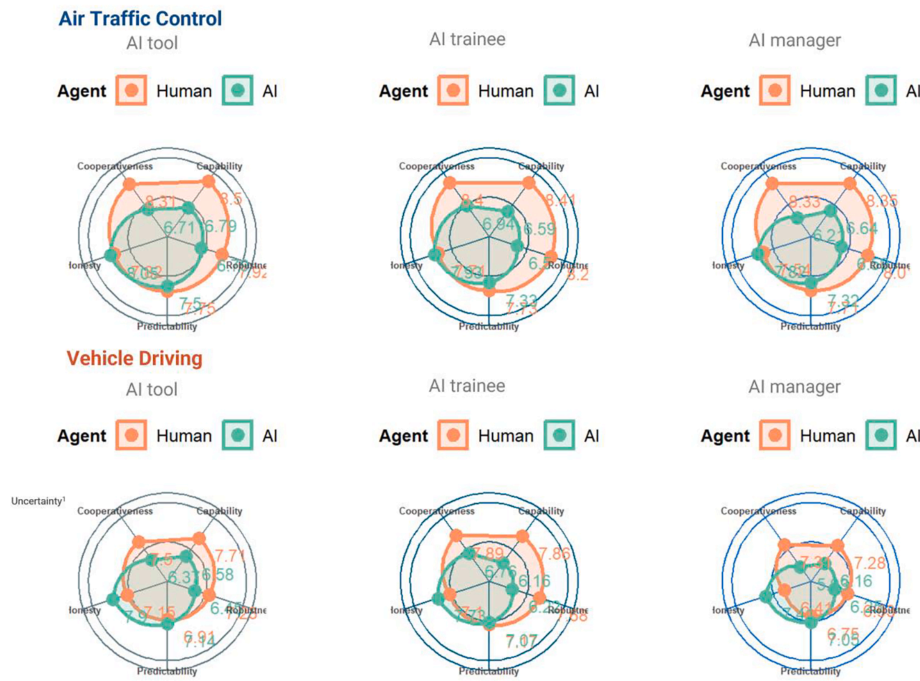
Consistent results were found in Study 1 (Convenience sample) and Study 2 (Nat. Rep. sample).

Appendix B. Trust and blame attributions in Study 1 using convenience sample



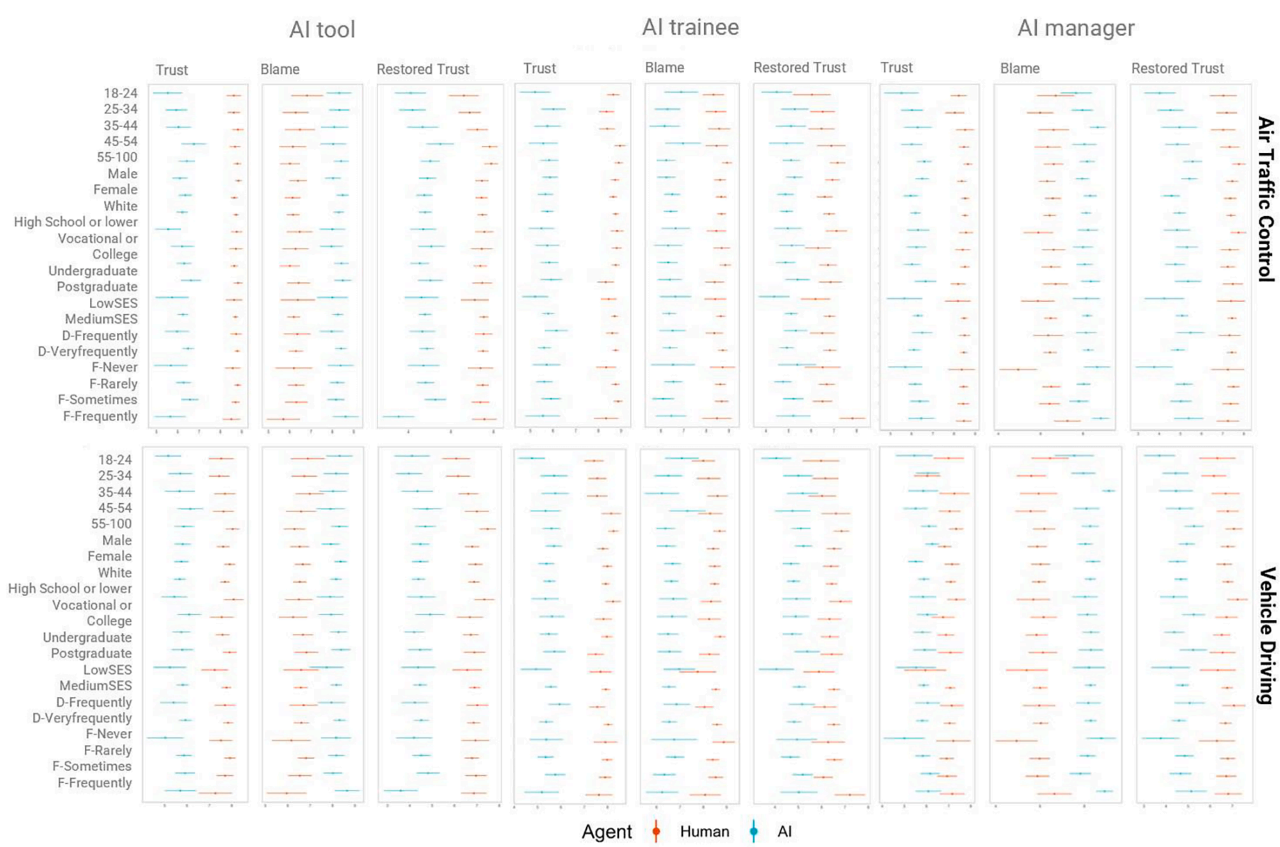
Comparative attribution of trust, blame, and restoration of trust after fixed mistake in Human vs. AI operators across three Human-AI Interaction (HAI) models (using a 10-point Likert scale): (A) Trust and Restored Trust in ATC. (B) Blame in ATC. (C) Trust and Restored Trust in VD. (D) Blame in VD. Consistent patterns were found across both Study 1 and Study 2. Each plot includes three HAI models: L0-AI tool, L3-AI trainee, and L5-AI Manager, from the bottom to the top. *Note. The dots on the solid lines show the mean difference with the accompanying error bars delineating the 95 % confidence intervals for these means. The dashed lines show the average attributions assigned to human and AI operators combining the 3 HAI models. Mean ratings for human operators are shown in orange, while those for AI appear in green. The colours for restored trust (after a mistake) are intentionally lighter, emphasising the decrease in trust following errors.*

Appendix C. Perception differences for ATC and VD in 3 HAI models



Perception differences Compared between Air Traffic Control (ATC) and Vehicle Driving (VD). Consistent patterns were found across 3 HAI models in both ATC and VD. *Note. The dots on the solid lines show the mean difference with the accompanying error bars delineating the 95 % confidence intervals for these means.*

Appendix D. Heterogeneity analysis for ATC and VD in 3 HAI models



Trust and Blame in Demographic Subgroups in 3 HAI models for ATC and VD. Consistent patterns were found across 3 HAI models in both ATC and VD. *Note. 'DiffTrust' refers to the difference between the Initial Trust and Restored Trust; 'D-' refers to driving; 'F-' refers to flying (as a passenger). For clarity in visualisation, only groups with more than 30 participants were included.*

Data availability

I have shared the link to my data at the attach files step.

References

Adnan, N., Nordin, S.M., bin Bahruddin, M.A., Ali, M., 2018. How trust can drive forward the user acceptance to the technology? in-vehicle technology for autonomous vehicle. Transp. Res. A Policy Pract. 118, 819–836.
Agudo, U., Liberal, K.G., Arrese, M., Matute, H., 2024. The impact of ai errors in a human-in-the-loop process. Cognit. Res.: Princ. Implic. 9 (1), 1.
Ajenaghughrure, I.B., Sousa, S.D.C., Lamas, D., 2020. Measuring trust with psychophysiological signals: a systematic mapping study of approaches used. Multimodal Technologies and Interaction 4 (3), 63.
Alhakami, A.S., Slovic, P., 1994. A psychological study of the inverse relationship between perceived risk and perceived benefit. Risk Anal. 14 (6), 1085–1096.
Araujo, T., Helberger, N., Kruikemeier, S., De Vreese, C.H., 2020. In ai we trust? perceptions about automated decision-making by artificial intelligence. AI & Soc. 35 (3), 611–623.
ATM, E, 2024. European atm master plan. Retrieved from https://atmmasterplan.eu/.
Awad, E., Levine, S., Kleiman-Weiner, M., Dsouza, S., Tenenbaum, J. B., Shariff, A., et al., 2018. Blaming humans in autonomous vehicle accidents: Shared responsibility across levels of automation. *arXiv preprint arXiv:1803.07170*.
Awad, E., Levine, S., Kleiman-Weiner, M., Dsouza, S., Tenenbaum, J.B., Shariff, A., Rahwan, I., 2020. Drivers are blamed more than their automated cars when both make mistakes. Nat. Hum. Behav. 4 (2), 134–143.
Baber, J., Kolodko, J., Noel, T., Parent, M., Vlacic, L., 2005. Cooperative autonomous driving: intelligent vehicles sharing city roads. IEEE Robotics Automation Magazine 12, 44–49. https://doi.org/10.1109/MRA.2005.1411418.
Bang, D., Frith, C.D., 2017. Making better decisions in groups. R. Soc. Open Sci. 4 (8), 170193.
Baum-Talmor, P., Kitada, M., 2022. Industry 4.0 in shipping: Implications to seafarers' skills and training. Transp. Res. Interdiscip. Perspect. 13, 100542.
Bennett, J.M., Challinor, K.L., Modesto, O., Prabhakharan, P., 2020. Attribution of blame of crash causation across varying levels of vehicle automation. Saf. Sci. 132, 104968.
Bigman, Y.E., Gray, K., 2018. People are averse to machines making moral decisions. Cognition 181, 21–34.
Bonnefon, J.-F., Rahwan, I., Shariff, A., 2024. The moral psychology of artificial intelligence. Annu. Rev. Psychol. 75 (1), 653–675.
Borhan, M.N., Ibrahim, A.N.H., Miskeen, M.A.A., 2019. Extending the theory of planned behaviour to predict the intention to take the new high-speed rail for intercity travel in libya: assessment of the influence of novelty seeking, trust and external influence. Transp. Res. A Policy Pract. 130, 373–384.
Brannon, L.A., Tagler, M.J., Eagly, A.H., 2007. The moderating role of attitude strength in selective exposure to information. J. Exp. Soc. Psychol. 43 (4), 611–617.
Burton, J.W., Stein, M.-K., Jensen, T.B., 2020. A systematic review of algorithm aversion in augmented decision making. J. Behav. Decis. Mak. 33 (2), 220–239.
Copp, C.J., Cabell, J.J., Kemmelmeier, M., 2023. Plenty of blame to go around: attributions of responsibility in a fatal autonomous vehicle accident. Curr. Psychol. 42 (8), 6752–6767.
Craig, M., Jamie, Nojoumian, 2021. Should Self-Driving Cars Mimic Human Driving Behaviors? In *Advances in Human Factors in Robots and Unmanned Systems,* pp. 213–225. doi: 10.1007/978-3-030-78358-7 14.
Dean, D.H., 2024. After the accident: is there a blame bias against the airline? J. Bus. Ethics 1–13.
DeGuzman, C.A., Donmez, B., 2024. Propensity to trust technology and subjective, but not objective, knowledge predict trust in advanced driver assistance systems. Transport. Res. F: Traffic Psychol. Behav. 107, 726–743.
Dietvorst, B.J., Simmons, J.P., Massey, C., 2015. Algorithm aversion: people erroneously avoid algorithms after seeing them err. J. Exp. Psychol. Gen. 144 (1), 114.
Douglas, B.D., Ewell, P.J., Brauer, M., 2023. Data quality in online human-subjects research: Comparisons between mturk, prolific, cloudresearch, qualtrics, and sona. PLoS One 18 (3), e0279720.
Erzberger, H., Lauderdale, T.A., Chu, Y.-C., 2012. Automated conflict resolution, arrival management, and weather avoidance for air traffic management. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of aerospace engineering*, *226*(8), 930–949.
Etminani-Ghasrodashti, R., Kermanshachi, S., Rosenberger, J.M., Foss, A., 2023. Exploring motivating factors and constraints of using and adoption of shared autonomous vehicles (savs). Transp. Res. Interdiscip. Perspect. 18, 100794.
EUROCONTROL, 2023. Performance review report (prr) 2023. Retrieved from https://www.eurocontrol.int/publication/performance-review-report-prr-2023.
Filos, A., Tigkas, P., McAllister, R., Rhinehart, N., Levine, S., Gal, Y., 2020. Can autonomous vehicles identify, recover from, and adapt to distribution shifts? In *International conference on machine learning,* pp. 3145–3153.
Finucane, M.L., Alhakami, A., Slovic, P., Johnson, S.M., 2000. The affect heuristic in judgments of risks and benefits. J. Behav. Decis. Mak. 13 (1), 1–17.
Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., Combs, B., 1978. How safe is safe enough? a psychometric study of attitudes towards technological risks and benefits. Policy Sci. 9, 127–152.
UK Department for Transport (UK DfT), 2022. *Jet zero strategy: Delivering net zero aviation by 2050.* Retrieved from https://assets.publishing.service.gov.uk/media/.
UK Department for Transport (UK DfT), 2024. *Future of flight: A comprehensive action plan.* Retrieved from https://assets.publishing.service.gov.uk/media/943b7679e9c8d921dfeeb/fof-action-plan.pdf.
Franklin, M., Awad, E., Lagnado, D., 2021. Blaming automated vehicles in difficult situations. Iscience 24 (4).
Gao, P., Hensley, R., Zielke, A., 2014. A road map to the future for the auto industry. McKinsey Quarterly Oct, 1–11.
Gigerenzer, G., Reb, J., Luan, S., 2022. Smart heuristics for individuals, teams, and organizations. Annu. Rev. Organ. Psych. Organ. Behav. 9, 171–198.
Gill, T., 2020. Blame it on the self-driving car: how autonomous vehicles can alter consumer morality. J. Consum. Res. 47 (2), 272–291.
Grant, D.G., Behrends, J., Basl, J., 2023. What we owe to decision-subjects: beyond transparency and explanation in automated decision-making. Philos. Stud. 1–31.
Green, B., Chen, Y., 2019. The principles and limits of algorithm-in-the-loop decision making. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–24.
Gueriau, M., Dusparic, I., 2020. Quantifying the impact of connected and autonomouś vehicles on traffic efficiency and safety in mixed traffic. In: In *2020 Ieee 23rd International Conference on Intelligent Transportation Systems (itsc)*, pp. 1–8.
Guleria, Y., Pham, D.-T., Alam, S., Tran, P.N., Durand, N., 2024. Towards conformal automation in air traffic control: Learning conflict resolution strategies through behavior cloning. Adv. Eng. Inf. 59, 102273.
Guleria, Y., Tran, P., Pham, D.-T., Durand, N., Alam, S., 2021. A machine learning framework for predicting atc conflict resolution strategies for conformal automation.
Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y., De Visser, E.J., Parasuraman, R., 2011. A meta-analysis of factors affecting trust in human-robot interaction. Hum. Factors 53 (5), 517–527.
Hou, Y.-T.-Y., Jung, M.F., 2021. Who is the expert? reconciling algorithm aversion and algorithm appreciation in ai-supported decision making. Proc. ACM. Hum. Comput. Interact. 5 (CSCW2), 1–25.
International Air Transport Association (IATA), I. A. T. A., 2020. *Vision 2050.* Retrieved from https://www.iata.org/contentassets/bccae1c5a24e43759607a5fd8f44770b/vision-2050.pdf.
Ivanov, S.H., 2023. Automated decision-making. Foresight 25 (1), 4–19.
Janotta, F., Hogreve, J., 2024. Ready for take-off? the dual role of affective and cognitive evaluations in the adoption of urban air mobility services. Transp. Res. A Policy Pract. 185, 104122.
Jansson, E., Olsson, N.O., Froidh, O.̈, 2023. Challenges of replacing train drivers in driverless and unattended railway mainline systems—a swedish case study on delay logs descriptions. Transp. Res. Interdiscip. Perspect. 21, 100875.
Jussupow, E., Benbasat, I., Heinzl, A., 2020. Why are we averse towards algorithms? a comprehensive literature review on algorithm aversion.
Kang, C., Lee, C., Zhao, X., Lee, D., Shin, J., Lee, J., 2025. Safety still matters: unveiling the value propositions of augmented reality head-up displays in autonomous vehicles through conjoint analysis. Travel Behav. Soc. 39, 100915.
Kaplan, A.D., Kessler, T.T., Brill, J.C., Hancock, P.A., 2023. Trust in artificial intelligence: Meta-analytic findings. Hum. Factors 65 (2), 337–359.
Kenesei, Z., Asv́anyi, K., Ḱok̈eny, L., J́aszbeŕenyi, M., Miskolczi, M., Gyulav́ari, T., Syahrí var, J., 2022. Trust and perceived risk: How different manifestations affect the adoption of autonomous vehicles. *Transportation research part A: policy and practice*, *164*, 379–393.
Kirwan, B., 2024. The impact of artificial intelligence on future aviation safety culture. Future Transportation 4 (2), 349–379.
Kumar, S., Datta, S., Singh, V., Datta, D., Singh, S.K., Sharma, R., 2024. Applications, challenges, and future directions of human-in-the-loop learning. IEEE Access.
Kuper, A., Kr̈amer, N., 2024. Psychological traits and appropriate reliance: Factors shaping trust in ai. Int. J. Human-Computer Interaction 1–17.
Kusumastuti, S.A., Kolkman, T.H., Lo, J.C., Borsci, S., 2025. Charting the landscape of rail human factors and automation: a systematic scoping review. Transp. Res. Interdiscip. Perspect. 30, 101350.
Langford, K., Kille, T., Lee, S.-Y., Zhang, Y., Bates, P.R., 2022. "in automation we trust"-australian air traffic controller perspectives of increasing automation in air traffic management. Transp. Policy 125, 352–362.
Lankton, N.K., McKnight, D.H., Tripp, J., 2015. Technology, humanness, and trust: Rethinking trust in technology. J. Assoc. Inf. Syst. 16 (10), 1.
Lecheler, S., De Vreese, C.H., 2016. How long do news framing effects last? a systematic review of longitudinal studies. Ann. Int. Commun. Assoc. 40 (1), 3–30.
Lee, Y.H., Jeon, J.-D., Choi, Y.-C., 2012. Air traffic controllers' situation awareness and workload under dynamic air traffic situations. Transp. J. 51 (3), 338–352.

Li, J., Liu, J., Wang, X., Liu, L., 2024. Investigating the factors influencing user trust and driving performance in level 3 automated driving from the perspective of perceived benefits. Transport. Res. F: Traffic Psychol. Behav. 105, 58–72.

Li, W.-C., Kearney, P., Zhang, J., Hsu, Y.-L., Braithwaite, G., 2021. The analysis of occurrences associated with air traffic volume and air traffic controllers' alertness for fatigue risk management. Risk Anal. 41 (6), 1004–1018.

Liu, P., Du, Y., 2022. Blame attribution asymmetry in human–automation cooperation. Risk Anal. 42 (8), 1769–1783.

Liu, P., Yang, R., Xu, Z., 2019. How safe is safe enough for self-driving vehicles? Risk Anal. 39 (2), 315–325.

Liu, Y.a.W. L. a. D. J., Peng and Du, 2020. Ready to bully automated vehicles on public roads? *Accident Analysis & Prevention*, *137*, 105457. doi: 10.1016/j.aap.2020.105457.

Lombrozo, T., 2006. The structure and function of explanations. Trends Cogn. Sci. 10 (10), 464–470.

Malik, S., Khan, M.A., El-Sayed, H., 2021. Collaborative autonomous driving—a survey of solution approaches and future challenges. Sensors (basel, Switzerland) 21. https://doi.org/10.3390/s21113783.

Malle, B.F., Ullman, D., 2021. A multidimensional conception and measure of human robot trust. In *Trust in human-robot interaction*. Elsevier, pp. 3–25.

Matin, A., Dia, H., 2022. Impacts of connected and automated vehicles on road safety and efficiency: a systematic literature review. IEEE Trans. Intell. Transp. Syst. 24 (3), 2705–2736.

McDougall, G., Roberts, A., 2008. Commercializing air traffic control: have the reforms worked? Can. Public Adm. 51 (1), 45–69.

McEvily, B., Tortoriello, M., 2011. Measuring trust in organisational research: Review and recommendations. J. Trust Res. 1 (1), 23–63.

Mogford, R. H., Guttman, J. A., Morrow, S., Kopardekar, P., et al., 1995. The complexity construct in air traffic control: A review and synthesis of the literature.

Mondy, C., Enz, M., Rust, D., 2025. An empirical analysis of factors affecting the commercial adoption of autonomous trucks. Transp. Res. Interdiscip. Perspect. 29, 101327.

Morariu, A.-R., Tsvetkova, A., Hellstrom, M., Bolbot, V., Virtanen, S., 2025. Investi-̈gating educational solutions in the field of autonomous shipping at finnish universities. Transp. Res. Interdiscip. Perspect. 31, 101402.

Morton, C., Mattioli, G., Anable, J., 2021. Public acceptability towards low emission zones: the role of attitudes, norms, emotions, and trust. Transp. Res. A Policy Pract. 150, 256–270.

Nussberger, A.-M., Luo, L., Celis, L.E., Crockett, M.J., 2022. Public attitudes value interpretability but prioritize accuracy in artificial intelligence. Nat. Commun. 13 (1), 5821.

Olia, A., Razavi, S., Abdulhai, B., Abdelgawad, H., 2018. Traffic capacity implications of automated vehicles mixed with regular vehicles. J. Intell. Transp. Syst. 22 (3), 244–262.

International Civil Aviation Organization (ICAO), 2016. Manual on air traffic controller competency-based training and assessment. Retrieved from https://www.icao.int/sam/documents/2016-cbt/10056_draft_en.pdf.

International Civil Aviation Organization (ICAO). *Manual on air traffic management requirements* [Computer software manual]. Retrieved from https://www.icao.int/airnavigation/imp/ documents/doc%209882%20-%20manual%20on%20atm%20requirements.pdf.

Palan, S., Schitter, C., 2018. Prolific. ac—a subject pool for online experiments. J. Behav. Exp. Financ. 17, 22–27.

Perez-Casť an, J. A., Ṕ erez-Sanz, L., Serrano-Mira, L., Sá ez-Hernando, F. J., Rodŕ ıguez Gauxachs, I., & Gomez-Comendador, V.F., 2022. Design of an atc tool for conflict detectioń based on machine learning techniques. Aerospace, 9(2), 67.

Porsdam Mann, S., Earp, B.D., Nyholm, S., Danaher, J., Møller, N., Bowman-Smart, H., et al., 2023. Generative ai entails a credit–blame asymmetry. Nat. Mach. Intell. 5 (5), 472–475.

Pozharliev, R., De Angelis, M., Donato, C., Rossi, D., 2023. Do not put the blame on me: Asymmetric responses to service outcome with autonomous vehicles versus human agents. J. Consum. Behav. 22 (2), 455–467.

Prolific, 2024. Representative samples. Retrieved from https://researcher-help.prolific.com/en/ article/95c345.

Renier, L.A., Mast, M.S., Bekbergenova, A., 2021. To err is human, not algorithmic–robust reactions to erring algorithms. Comput. Hum. Behav. 124, 106879.

SAE, 2018. Taxonomy and definitions for terms related to driving automation systems for onroad motor vehicles. Retrieved from https://www.sae.org/standards/content/j3016 201806/.

Seet, M., Harvy, J., Bose, R., Dragomir, A., Bezerianos, A., Thakor, N., 2022. Differential impact of autonomous vehicle malfunctions on human trust. IEEE Trans. Intell. Transp. Syst. 23 (1), 548–557. https://doi.org/10.1109/TITS.2020.3013278.

SESAR Joint Undertaking, 2020. Automation in air traffic management: Long-term vision and initial research roadmap. Retrieved from https://www.sesarju.eu/sites/default/files/documents/reports/automation%20atm%20vision%20roadmap.pdf.

Shariff, A., Bonnefon, J.-F., Rahwan, I., 2021. How safe is safe enough? psychological mechanisms underlying extreme safety demands for self-driving cars. Transp. Res. Part C Emerging Technol. 126, 103069.

Si, H., Duan, X., Cheng, L., De Vos, J., 2024. Adoption of shared autonomous vehicles: combined effects of the external environment and personal attributes. Travel Behav. Soc. 34, 100688.

Spieser, K., Treleaven, K., Zhang, R., Frazzoli, E., Morton, D., Pavone, M., 2014. Toward a systematic approach to the design and evaluation of automated mobility-on-demand systems: a case study in singapore. Road Vehicle Automation 229–245.

Tang, T., Guo, Y., Souders, D.J., Li, X., Yang, M., Xu, X., Qian, X., 2025. Moderating effects of policy measures on intention to adopt autonomous vehicles: evidence from china. Travel Behav. Soc. 38, 100921.

Trope, Y., Liberman, N., 2010. Construal-level theory of psychological distance. Psychol. Rev. 117 (2), 440.

Van Arem, B., Van Driel, C.J., Visser, R., 2006. The impact of cooperative adaptive cruise control on traffic-flow characteristics. IEEE Trans. Intell. Transp. Syst. 7 (4), 429–436.

Van Rooijen, S., Ellerbroek, J., Borst, C., van Kampen, E., 2020. Toward individualsensitive automation for air traffic control using convolutional neural networks. Journal of Air Transportation 28 (3), 105–113.

Vongvit, R., Maeng, K., Lee, S.C., 2024. Effects of trust and customer perceived value on the acceptance of urban air mobility as public transportation. Travel Behav. Soc. 36, 100788.

Walsh, S.E., Feigh, K.M., 2021. Differentiating 'human in the loop'decision process. In: In *2021 Ieee International Conference on Systems, Man, and Cybernetics (smc)*, pp. 3129–3133.

Wasinger, G., Mann, B., 2020. Twitter politics: the effect of language and source credibility on public perception. SACAD: John Heinrichs Scholarly and Creative Activity Days 2020 (2020), 91.

Waymo, 2024. *Waymo safety impact.* Retrieved from https://waymo.com/safety/impact/?utm_campaign=The%20Batch&utm_medium=email&_hsmi=324224770&utm_content=324223570&utm_source=hs_email.

Wotton, M.E., Bennett, J.M., Modesto, O., Challinor, K.L., Prabhakharan, P., 2022. Attention all 'drivers': you could be to blame, no matter your behaviour or the level of vehicle automation. Transport. Res. F: Traffic Psychol. Behav. 87, 219–235.

Yalcin, G., Themeli, E., Stamhuis, E., Philipsen, S., Puntoni, S., 2023. Perceptions of justice by algorithms. Artificial Intelligence and Law 31 (2), 269–292.

Yao, Z., Wang, Y., Liu, B., Zhao, B., Jiang, Y., 2021. Fuel consumption and transportation emissions evaluation of mixed traffic flow with connected automated vehicles and human-driven vehicles on expressway. Energy 230, 120766.

Yokoi, R., 2024. Trust in self-driving vehicles is lower than in human drivers when both drive almost perfectly. Transport. Res. F: Traffic Psychol. Behav. 103, 1–17.

Zefreh, M.M., Edries, B., Esztergar-Kiss, D., Torok, A., 2023. Intention to use privaté autonomous vehicles in developed and developing countries: what are the differences among the influential factors, mediators, and moderators? Travel Behav. Soc. 32, 100592.

Zhai, S., Gao, S., Wang, L., Liu, P., 2023. When both human and machine drivers make mistakes: Whom to blame? Transp. Res. A Policy Pract. 170, 103637.

Zhai, S., Wang, L., Liu, P., 2023. Human and machine drivers: sharing control, sharing responsibility. Accid. Anal. Prev. 188, 107096.

Zhang, Q., Wallbridge, C.D., Jones, D.M., Morgan, P., 2021. The blame game: double standards apply to autonomous vehicle accidents. In: In *International Conference on Applied Human Factors and Ergonomics*, pp. 308–314.

Zhang, Q., Wallbridge, C.D., Jones, D.M., Morgan, P.L., 2024. Public perception of autonomous vehicle capability determines judgment of blame and trust in road traffic accidents. Transp. Res. A Policy Pract. 179, 103887.

Zhao, B., Lin, Y., Hao, H., Yao, Z., 2022. Fuel consumption and traffic emissions evaluation of mixed traffic flow with connected automated vehicles at multiple traffic scenarios. J. Adv. Transp. 2022 (1), 6345404.

Zhao, Y., Rozier, K.Y., 2014. Formal specification and verification of a coordination protocol for an automated air traffic control system. Sci. Comput. Program. 96, 337–353.