Department for Transport

Statement of safety principles public perceptions research

March 2026

Department for Transport
Great Minster House
33 Horseferry Road
London
SW1P 4DR

This publication is also available on our website at
www.gov.uk/government/organisations/department-for-transport

Any enquiries regarding this publication should be sent to us at
www.gov.uk/government/organisations/department-for-transport

Table of contents

About

Authors: Balazs Csuvar, Edward Houghton, Lara Suraci, and Nick Reed.

Prepared by DG Cities and Reed Mobility for The Department for Transport

Executive summary

The Automated Vehicles Act 2024 established a benchmark requiring AVs to achieve a level of safety equivalent to, or higher than, that of careful and competent (C&C) human drivers, with the goal to improve overall road safety performance. However, sector-wide consultation and research have demonstrated a lack of understanding and clear definition of the C&C benchmark, and there remains a need for further insights on public attitudes and perceptions of it. Therefore, the DfT commissioned this study to conduct robust public research using deliberative methods on the Statement of Safety Principles (SOSP) for Automated Vehicles (AVs), specifically focusing on the "careful and competent" (C&C) driver benchmark. Engagement with members of the public within this project provides crucial insights into how the public will come to assess AV safety performance. Specifically, the study aimed to understand public perception of the C&C benchmark (Research question 1) and explore how people would expect AVs to demonstrate they meet it (Research question 2). These findings have several implications for the regulatory framework governing automated vehicles under the AV Act 2024. In particular, the findings provide insight into how the benchmark may need to be reflected in the regulatory requirements with which AV developers and operators will be required to comply.

Research question 1 - What are public attitudes and perceptions of the ‘careful and competent’ driver benchmark?

It was apparent that participants viewed the establishment of a set of safety benchmarks for AVs as a responsibility of government rather than industry. Participants were hesitant over any suggestion that AV developers should set their own standards or independently assess their own performance. This reinforces the importance of maintaining a regulatory model in which the safety benchmark is defined by the government and verified through independent assessment.

While participants broadly accepted the principle of the C&C benchmark, they also highlighted that the phrase itself is somewhat ambiguous and may not be easily understood without further explanation. Key terminology including ‘safety’ and ‘experience’ were highlighted as potentially more useful to convey as part of the definition. We found broad acceptance during the workshops happened once participants understood the framing of C&C as a benchmark higher than that of an average driver. This suggests that, although the benchmark may function effectively as a legal principle, it is unlikely to be sufficient on its own as a regulatory test without further explanation. Instead, regulators may need to translate the concept into more concrete and measurable performance

expectations. For AV developers, this would imply demonstrating how their systems exhibit behaviours consistent with C&C driving across a range of foreseeable driving scenarios that may be encountered within the operation design domain (ODD) of their AV operations. Such evidence to be produced by AV developers may include scenario-based testing, simulation results, and performance metrics relating to hazard detection, response times and compliance with traffic rules.

Another notable finding from the workshops is that participants tended to describe C&C driving in behavioural terms rather than statistical ones. Participants referred to behaviours such as anticipating hazards, maintaining awareness of surroundings, driving at appropriate speeds for conditions and interacting safely with other road users. They did not naturally frame the benchmark in terms of casualty statistics and on the whole demonstrated limited understanding of real-world collision rates, though intuited their importance. This suggests that the regulatory framework may need to incorporate both behavioural and outcome-based measures of safety. From the perspective of AV developers, this could mean demonstrating that systems behave in ways that align with recognised safe driving practices while also providing statistical evidence that overall safety performance is better than that of human drivers.

Whilst many participants in the driver cohort considered most humans to be C&C, this was not the case for some and especially did not align with the experience of many non-driver participants who highlighted that they feel unsafe on the road. This suggests that the perceptions of the C&C threshold vary. The workshops also indicated that participants often applied a very high standard when evaluating scenarios of road incidents, assuming that a C&C driver would have avoided them. As C&C is defined as being much safer than the average human driver, the findings highlight for regulators the importance of clearly defining the intended meaning of the benchmark, ensuring that regulatory guidance sets realistic expectations for what constitutes safe automated driving. Developers will likely need to demonstrate that their systems minimise risk and manage foreseeable hazards appropriately, rather than meeting an unrealistic expectation of eliminating all collisions.

Context was also shown to play an important role in how participants interpreted C&C driving behaviour. Participants in rural environments emphasised factors such as speed management, limited visibility and road conditions, while urban participants focused more on interactions with pedestrians and other road users. This finding supports the regulatory approach of defining AV capabilities in terms of ODDs. AV developers will therefore need to demonstrate that their systems meet the C&C benchmark within the specific environments in which they are intended to operate. This aligns with international regulatory practice, where automated driving approvals are typically linked to clearly defined operational conditions. Standards such as ISO34503:2023 and BSI PAS 1883:2025 provide useful framing for how ODDs can be characterised.

Participants also referred to concepts such as ‘*common sense*’ and ‘*intuition*’ when describing C&C driving. These qualities cannot be easily defined by objective measures but highlight the importance of demonstrating how AVs perceive and respond to complex driving situations. From a regulatory perspective, this suggests a growing importance for transparency and traceability in AV decision-making. Developers may therefore need to provide evidence explaining how their systems interpret the driving environment, predict the behaviour of other road users and determine appropriate responses. Such transparency is increasingly reflected in emerging regulatory discussions around safety

assurance and system accountability. The SAE J3237 Recommended Practice provides insights into the tools that might be used to achieve this in practice.

Research question 2 - How would the public expect ‘careful and competent’ driving to be demonstrated?

On the requirements to demonstrate AV safety, participants placed the most trust in ‘independent certification and audit evidence' and ‘real-world safety outcomes data’. There was a consensus that a variety of evidence should be presented in an easily digestible format and that information directly from AV manufacturers was explicitly distrusted. When discussing certification, the overarching view was that AV developers must provide structured safety evidence demonstrating how their systems meet the C&C benchmark and cannot rely on internally derived safety claims. This evidence should be reviewed and assessed by an independent body like DfT/Driver and Vehicle Standards Agency (DVSA)/Vehicle Certification Agency (VCA), organisations that do not have a direct commercial interest in the rollout of AV-based services. This also aligns with the emerging direction of international AV regulation, such as UNECE Regulation 157, which governs Automated Lane Keeping Systems (ALKS). Regulators set high-level safety objectives, require structured pre-deployment evidence, constrain deployments to defined operating conditions and expect ongoing in-service monitoring and data-sharing to ensure that systems continue to operate safely after approval.

The general view was that the more data available the better when looking at the demonstration of meeting the benchmarks, with highest importance placed on real-world evidence in the form of videos, simulations or case studies from international deployments. Transparency was a key factor and whilst participants suggested they might not look at all the data; its availability did build trust.
The discussions also outlined that people expected communication about the benchmarks and overall sector to be available in easy-to-digest formats, across a variety of channels to reach different audiences.

Finally, the findings suggest that public confidence in automated vehicles will depend not only on the authorisation process but also on ongoing oversight once vehicles are deployed. There were expectations that standards would continue to be enforced in practice continuously, with benchmarks that evolve over time. This points to the importance of post-deployment monitoring and regulatory supervision of automated vehicle performance. Developers and operators may therefore be required to provide ongoing operational safety data, enabling regulators to monitor real-world performance and investigate incidents where necessary. This approach is consistent with the direction of travel in both UK (as described in the DfT’s Developing the Automated Vehicles Regulatory Framework Call for Evidence, 2025 as the In-Use Regulatory Scheme) and international automated vehicle regulation, where continuous monitoring and regulatory review are increasingly seen as essential components of safety assurance.

Recommendations

The study concludes that whilst the C&C benchmark is a positive starting point, it must be accompanied by a standardised, concrete set of criteria that is less ambiguous and more measurable. These criteria would have to outline that the C&C benchmark is equivalent to

only the standard of the very best drivers and, by definition, behaving significantly better than average drivers.

We recommend that the language of the SOSP is tested on a representative UK sample, to ensure the expectations of people are meaningfully translated into AV driving behaviours. This will provide evidence to guide the basis of the SOSP and allow for related communications to be developed that effectively and meaningfully convey the requirements for AVs and the benchmarks for AV testing.

There was strong public confidence in independent regulatory bodies establishing and enforcing the safety standard. Creating the branding and public messaging around it will help to build trust in the system and ensure long-term confidence. As people interpret the C&C benchmark as a driving behavioural framework, it is recommended that it is tested and demonstrated as such, showcasing AV capability in the form of video evidence of either simulation or real-world trials for transparency.

Finally, future public engagement should prioritise clear, jargon-free explanations and use multiple channels to reach all groups including those who are digitally excluded. Robust survey research using a nationally representative sample is recommended to test messaging so to identify and explore behavioural dimensions of the benchmark related to adoption, acceptance and trust in different demographics. This will support the development of effective future communications approaches.

Introduction

Statement of safety principles

The Automated Vehicles Act 2024 (AV Act) sets the legal framework for the safe deployment of automated vehicles in Great Britain. It implements the recommendations of the four-year review of regulation for automated vehicles carried out jointly by the Law Commission of England and Wales and the Scottish Law Commission (the Law Commissions).

Section 2 of the AV Act requires that the "Secretary of State must prepare a statement of the principles that the Secretary of State proposes to apply in assessing, for the purposes of this Part, whether a vehicle is capable of travelling autonomously and safely." The Act continues that these principles should be framed to secure that:

(a) authorised automated vehicles will achieve a level of safety equivalent to, or higher than, that of C&C human drivers, and
(b) road safety in Great Britain will be better as a result of the use of authorised automated vehicles on roads than it would otherwise be.

The Centre for Connected and Autonomous Vehicles (CCAV) in DfT ran a call for evidence between June and September 2025 to gather evidence and opinions on how a 'careful & competent' (C&C) safety standard should be described, the outcomes that should be achieved in meeting the safety standard, and the appropriate measures for understanding if outcomes have been achieved.

Stakeholder responses to the call for evidence highlighted diverse opinions on the levels of performance, consistency and safety levels achieved by C&C human drivers. This aligns with wider evidence from published literature, as well as stakeholder engagement that CCAV has undertaken.
Whilst the call for evidence helps to build understanding of stakeholder views on how to define the safety standard, Government wanted to undertake further qualitative research to engage the public directly in order to sufficiently capture public opinion on a 'careful and competent' safety standard.

Deliberative methods were sought so as to provide an opportunity to engage the public with more detailed information, and to enable dialogue and exploration with topic experts ahead of discussion and questions, thus ensuring more informed responses. Findings from this research can then be used to ensure that the subsequent development of the SOSP is informed by public opinion, ahead of a public consultation on the draft SOSP in Spring 2026.

Purpose of this study

The Department for Transport (DfT) commissioned DG Cities, Reed Mobility and DJS Research to deliver a robust, inclusive public deliberation on the Statement of Safety Principles (SOSP) for AVs, focusing on the "C&C driver" benchmark. The objectives of this study were to:

- **O1:** Understand public interpretation of the SOSP and the extent to which it provides reassurance regarding AV safety;
- **O2:** Explore the public's perceptions of "careful and competent" driving behaviour by humans versus AVs, identifying key areas of alignment and divergence; and,
- **O3:** Generate actionable insights and communication recommendations to support DfT's future engagement, and policy development under the Automated Vehicles Act; including the development of the draft SOSP.

From these study objectives we have defined the following research questions (RQs) that guided the development of workshops and stimuli materials. The RQs were assessed across context (rural, semi-urban, urban) and driver/non-driver status:

- **RQ1:** What are public attitudes and perceptions of the 'careful and competent' driver benchmark?
    - **RQ1a:** What do the public believe are the behaviours that characterise a careful human driver?
    - **RQ1b:** What do the public believe are the behaviours that characterise a competent human driver?
    - **RQ1c:** What are public expectations of an AV that meets the careful and competent benchmark (e.g. which scenarios must it manage)?
    - **RQ1d:** What behaviours/actions of an AV would need to be demonstrated to meet the careful and competent benchmark?
    - **RQ1e:** Would the public assess a careful and competent human and an AV in the same way?
- **RQ2:** How would the public expect 'careful and competent' driving to be demonstrated?
    - **RQ2a:** How would the public benchmark a careful and competent human driver?
    - **RQ2b:** What outcome measures would be most trusted to confirm a careful and competent driving behaviour?
    - **RQ2c:** How would people expect careful and competent AV capability to be demonstrated before getting on the road?
    - **RQ2d:** What measure would people expect to be captured about AV driving to demonstrate careful and competent driving?

Methodology

Rapid review of AV safety terminology

A rapid review of road safety definitions was undertaken of published peer reviewed literature, grey literature, and selected internal DfT analysis documents to develop a terminology framework. The analytical frameworks included key standards, including BSI Standards and terms used across DfT/CCAV Innovate UK projects. A summary document detailing the outputs of the rapid review which detailed the implications for the research method was produced and used in synthesis and method design during the methodology development stage of the study.

Deliberative workshops

Deliberative workshops were used to explore the project RQs through an agenda of activities, reflection time and discussion, designed to support participant learning to enable informed engagement in discussions and deliberation. Workshops were facilitated by three facilitators and drew on a mix of data and information sharing, “ask the expert” expert Q&A, tasks using materials including scenario cards, and discussion.
The session plan is presented below. A full set of materials is shared in Annex 1.

1. Introduction to the workshop.
2. Discussion: Using the roads and road safety.
3. Video and expert Q&S: Introduction to Automated Vehicles.
4. Introduction to the careful and competent benchmark.
5. Activity 1: Group scenario sorting exercise.
6. Activity 2.1: Scenario exercise - travelling to the hospital.
7. Activity 2.2: Group evidence sorting exercise.
8. Conclusion discussion and close.

Locations and study sample

Three locations were selected: Leicester, Stirling, and Acle (Norfolk) and surrounding villages to ensure that public deliberation robustly reflects the diversity of road environments, population characteristics, and mobility challenges encountered across the UK. These locations offer contrasting transport contexts (urban, suburban/town, rural) while also capturing variation in demographics, modal reliance, safety concerns, and exposure to emerging mobility technologies. Two workshops were undertaken in each location, one with **drivers** and another with **non-drivers**.

Leicester provided a large, complex urban setting where mobility is shaped by density, diversity, and multimodal networks. It is a major city of over 388,000 residents with a notably young and ethnically diverse population, including 67.3% aged 16-64 and 43.4% identifying as Asian, far above national averages. Leicester demonstrates high public

transport coverage (93% of households within 400m of a bus stop) and extensive bus use. 33% of households do not have a car, well above the average for England which is 23.5%. This is comparable to larger urban centres like Manchester (39%), Birmingham (26%) and Liverpool (34%). Yet the city still faces high private car dominance, including 25% of all car journeys being under 2km. The city previously was host to AV-related activity, including the Project MuCCA safety demonstration in 2020.
Stirling provided a suburban and small-city environment, with a mid-sized population of 94,210 and travel patterns shaped by both commuting to larger cities and connectivity issues across surrounding settlements. Over 70% of households own one or more vehicles, but public priorities increasingly emphasise sustainable mobility, as highlighted through local consultation identifying gaps in bus reliability, active travel safety, and accessible transport options. Stirling also sits locally to the CCAV-funded GAMMA project testing autonomous public transport in Glasgow, providing proximity to a local case study and stakeholder networks.

Acle and surrounding villages are a rural community in Norfolk with a population of 2,788, provided perspectives from areas where mobility is constrained by limited transport choices and high car dependency; 57% of employed residents drive to work and only 3% use bus services, reflecting very low public transport provision and rural isolation challenges. With a comparatively older demographic (25.8% aged 65+) and higher proportion of disabled residents (22%) than the England average, Acle enables exploration of AV safety expectations among groups who may particularly benefit from, or feel vulnerable around, automated mobility. Norfolk's tourism-driven economy and seasonal transport pressures further add a unique dimension to deliberation on AV reliability, predictability, and trust in varied operating conditions. Whilst our focus for the workshops was on full-time residents, the tourism aspects provide a set of interesting local use cases for future AV services and questions related to perceived driving style of visitors.

Achieved study sample:

- Leicester (Urban): 15 Driver, 14 Non-Driver
- Stirling (Semi-urban): 15 Driver, 14 Non-Driver
- Acle (Rural): 15 Driver, 11 Non-Driver

Venues were selected in key urban centres with good public transport and road connections. The target 12-15 participants was achieved in 5 of 6 workshops. There were some difficulties in recruiting rural non-drivers for attendance from local villages - where the journey to a suitable venue in Norwich was a perceived barrier. However, attendance rates provided adequate data.

Pre- and post- workshop surveys

Participants undertook pre- and post-workshop surveys as part of project on-board and off-boarding processes. The surveys were deployed online, and shared with participants 5-7 days before or after sessions. Surveys consisted of the following topics, and full survey instruments are available in Annex 2:

- Attitudes towards automated vehicle safety benchmark
- Views on C&C human driver and AV behaviours
- Knowledge and awareness of automated vehicle
- Information communication preferences
- Workshop feedback (Post-measures only)

Survey responses were used during the analysis predominantly to measure how the deliberative exercises and information provided during the workshops shifted participants' opinion. Whilst the surveys do not constitute a statistically representative sample of the UK population, they provide useful insights in the context of the workshops.

Navigating this document

It is important to highlight that the focus of this study was to gather a broad range of views from participants, trying to cover the type of thoughts, beliefs, perceptions and concerns people might have about the C&C benchmark. The findings of this research are written in a manner that captures this rich feedback we have gathered, sharing with the reader the full range of immediate thoughts people might have, as well as conclusions upon deliberation that the public might reach. This was not a representative quantitative study and we did not try to establish what most people in the country would think; therefore, ratios are not given in the report. Across the report, we have indicated the scale of a response when it was significant and similarly made comparisons between findings in different workshops where relevant.

The Findings and Recommendations sections then distil the essence of the report, capturing, following our analysis, what the most important and relevant takeaways from the study are, and how they may help inform the SOSP.

Findings

RQ1: What are public attitudes and perceptions of the 'careful and competent' driver benchmark?

Across workshops, participants generally supported the benchmark, but there were instances of participants among driver and non-driver groups in all three locations who had some important reservations. These nuances and the underlying range of considerations are explored below.

Are human drivers careful & competent?

When asked if human drivers are careful and competent (C&C), participants' answers were varied. Most participants, across all **driver** cohorts, believed the majority of UK drivers are, with only some drivers acting irresponsibly.

Urban non-drivers had a different opinion, with no participant believing most drivers to be C&C. Participants expressed that they do not feel safe in their local areas, citing frequent crashes and the impatience of drivers who *'want to beat you to it, wanting to get there first'*. However, some participants reflected that these perceptions may be influenced by the tendency to remember negative driving behaviour and accidents more vividly. As in the urban non-driver workshop, several **rural drivers** noted a bias towards remembering poor driving more vividly than routine competent behaviours.

Deliberation on accident statistics mainly reinforced people's existing beliefs. Those participants who considered most drivers to be C&C often highlighted that, compared to the amount of driving that happens on UK roads every day, the number of killed and seriously injured (KSI) victims felt proportionate. In comparison, those who believed that drivers were not C&C deemed the number of KSI victims to be high.

Is human driving a relevant benchmark?

Semi-urban drivers considered the comparison to human drivers to be fair and reasonable for the benchmark and felt this was a necessary comparison for the benchmark to have utility, but some still considered C&C overly subjective. Participants also acknowledged that many human drivers are not up-to-date on the Highway Code, yet still saw the Highway Code as a useful benchmark for judging safe driving behaviour.

One **urban driver** suggested tying the definition to an advanced driver test, an established method of certification for human drivers that could be used as a benchmark.

Within the **rural non-drivers** there was clear push-back that highlighted a negative framing of C&C as a comparison with human driving. One participant noted: *"I'm halfway there - I think it should be a little higher benchmark. I don't think it should be based on humans and what humans do."* Analogies were made to other professions (e.g. surgeons), expressing the view that automated systems should exceed human standards rather than replicate them.

On a different point, **urban non-drivers** pointed out that; "We give human drivers a certain level of leniency. Otherwise, if one accident happened and everyone got banned from driving, everyone would be like 'That's unfair, that's unreasonable.' I almost think we do need to give automated cars that same chance, which is why I think it – I agree that looking at the benchmarks of human errors, to kind of, transfer it over does make a lot of sense."

What are the first impressions of the C&C benchmark?

Positive aspects of the benchmarks:

Many participants across all cohorts viewed the C&C benchmark as a reasonable and sensible standard for judging automated vehicle (AV) safety. The majority continued to endorse the benchmark after deliberative discussions and tracking questions, though with important reservations regarding the clarity of language.

Many considered that the standard was good enough to help people understand: *"it's the best we have out of a bad bunch."* **Semi-urban drivers** also noted that the difficulty in defining the standard was a major barrier due to the uncertainty it presents to Government and the public: "*we wouldn't know if we had chosen the right criteria*".

One participant noted that after discussion it was clear that 'careful' was a term they'd encountered in various public transport contexts - and on reflection it was a term they would expect to see in relation to an AV on the roads.

Ambiguity in terminology:

Participants who did not agree with the benchmark by the end of deliberation stated it was because they felt the language was not specific enough.

"(This) discussion and scenarios has helped me come to the conclusion that being careful and competent is too vague and subjective to be a reliable benchmark". (Semi-urban driver)

Others felt it was: "quite woolly - everybody has a different interpretation"; and vague:
"I don't think it's tight enough. There are 15 different opinions about what it means and that brings me insecurity. I need a definition to make me feel secure" (Semi-urban driver)

Rural drivers who expressed uncertainty in the benchmark recognised that there wasn't enough information available to make an informed decision:
"I need more detail on careful and competent before I can decide - there are a lot more parameters I would need before I judge that it is safe".

"I strongly feel the word 'careful' is a little ambiguous and open to opinion … I feel 'safe' would be a better word as 'safe' to me is a more measurable goal and would have clearer definitions/standards it needs to achieve. 'Safe' and 'competent' both have benchmarks that would be clear to then make an informed decision.… The word 'careful' conveys to me a goal that is based more on a human opinion and therefore would not likely be a consistent benchmark if assessed by the population and is not a measurable standard." Rural Driver.

The **rural non-driver** discussion highlighted there was varied interpretation of C&C, which was considered a barrier for a precise benchmark for AVs, given that precision is key for a safety standard designed to improve safety.

Rural drivers discussed the distinction between "careful" and "competent", suggesting that careful relates more to behaviour and attitude, whereas competence relates to knowledge and rule adherence and was thus preferred as an objective benchmark by some.

Among those who didn't support the C&C benchmark there were concerns that over time companies will not be held to this high standard:
"That's the challenge – when new things like this normally come about, with companies especially, they'll normally put all of their efforts into making it all amazing to get people hooked and then things start getting diluted over the years." (Urban Driver)

How do participants define the C&C benchmark?

A key detail discussed during deliberation was regarding the capabilities of a C&C driver, what it means in practice and how C&C compares to average drivers. Most participants had a clear expectation for AVs to be better than human drivers and expected the C&C benchmark to meet this standard. A range of views on how participants have interpreted this discussion can be read below.

Need for clarity of wording:

One **urban driver** stated that the benchmark is just the starting point of safety, like a driving test is for humans: "*What I'm saying is, careful and competent is where it starts from, where it begins.*" It is the minimum level from which improvements should come. There was a consensus that '*getting from A to B without incident*' is not enough to qualify as C&C if rules are broken along the way. Being driven by a C&C driver was described as a journey that results in no incidents or near-misses and where the passenger feels comfortable, but upon prompting, people agreed that C&C is a good descriptor for the type of journey that will likely lead to the outcome of arriving safely at the destination. Upon further deliberation, **urban drivers** suggested that C&C needs a definition -specifically

additional words to explain what it means - though they noted that throwing in too many descriptors can make the definition '*fuzzy*'.

Regarding the wording of the standard, one **urban driver** argued for a higher threshold: *"It doesn't need to be careful and competent, it needs to be exceptionally careful and competent."* Some **urban drivers** suggested a third-C, *'considerate'* as a necessary addition, involving respect for other road users (e.g. letting people out at junctions). *'Common sense'* was another C-word suggested by **urban drivers and non-drivers** alike. "*Awareness*" was also mentioned as a crucial behaviour, whilst highlighting that it is a very human attribute, potentially difficult to for an AV to demonstrate. A similar conclusion was drawn with the word "*intuition*", as C&C humans can understand that something might happen based on small behaviours of other road users around them.

Participants in the **semi-urban drivers group** thought the benchmark lacks the concept of "*experience*":

"What has been learned over 30 years of driving, compared to a learner driver."

"I think it is important to include human experience to the careful and competent (benchmark) as a 17-year-old who has passed their test recently could be careful and competent but an experienced driver of 30 years will have so much more experience of dealing with a huge variety of situations in a careful and competent way."

Some **semi-rural non-drivers** noted that the benchmark had to make them feel responsible for others on the roads; '*I am taking care of other road users'* and, *'we are doing our best".*

Overall, people expressed that the C&C wording by itself does not cover all the requirements they might have for a driving benchmark, with a range of nuances and options explored.

Comparison with average drivers:

The **semi-urban non-driver** group noted that setting a 'careful and competent' standard for AVs would mean a standard for human driving would also need to be agreed - but that was considered difficult as some human drivers find ways to '*always break the rules*'. Some considered careful and competent to be adequate but that the benchmark should be "*at least better than… or the top 5% of drivers. Has to be better than the top 5% of drivers*", however this was rejected by others, who considered the careful and competent benchmark to better describe the average driver - "*50% are careful and competent*". This was further highlighted by a **semi-urban non-driver**: *"AV safety should definitely be above careful and competent".*

There were some concerns that the benchmark may be interpreted as an average driving quality by the public. Among the non-driver group, this also meant that experience played a part in attitudes:
"I think people would like to think that they are careful and competent but I keep remembering the (graph illustrating average drivers) …and I think it depends as well on your experience just generally of cars and the road." Rural non-driver

When an alternative benchmark against an average driver was discussed by illustrating the standards of driving through a graph of driving quality, from dangerous driving, through to C&C driving – the standard of the "average driver" was deemed unacceptable by most: *"It needs to be better than average driver - it doesn't have human error".* One **rural driver** participant did note that they currently use public transport - and therefore trust other human drivers who may be average.

Some participants explicitly noted that applying the C&C benchmark to AVs effectively implies a higher standard than that achieved by the average human driver; this framing appeared to increase support for the benchmark among other, more sceptical participants.

In short, people only accepted C&C as a benchmark when it was clear that it was a higher standard than that of an average driver.

Requirement for a very high C&C benchmark:

One **rural non-driver** required AVs to be benchmarked against a higher standard than C&C, accepting very little deviation from near-perfect driving. There was clear agreement that a benchmark against an average human driver was not acceptable as it was too low a standard, and against the reason for automating driving:

"It needs to be better than that"; "We're trying to improve safety and we're trying to reduce the number of fatalities, so it's got to be the best." Rural non-driver.
"I would like a higher benchmark than careful and competent, perhaps "excellent" to raise it just a little." Rural non-driver.

In conclusion:

Overall, participants indicated that while C&C initially felt vague, the visual explanation illustrating C&C as the *"golden standard"* which is higher than the average human driver, helped participants understand how the benchmark functions as a reference point for judging safety. This comparison appeared to increase confidence in the benchmark and may have encouraged support beyond what the term alone generated.

RQ1a: What do the public believe are the behaviours that characterise a careful human driver? | RQ1b: What do the public believe are the behaviours that characterise a competent human driver?

Across all workshops, there were several common behaviours cited by drivers and non-drivers that they considered to characterise a careful and competent human driver. The full set of responses across factors are included in Table 2.

Drivers across all locations tended to think that a careful and competent human driver is characterised by the following behaviours:

- Considerate / respectful of other road users
- Rule compliant
- Situationally aware and anticipatory

- Predictable
- Communicates well with other road users
- Adaptable

Non-drivers across all locations tended to think that a careful and competent human driver is characterised by the following behaviours:

- Focused/attentive to the road
- Aware of driving conditions and vehicle capability
- Situationally aware and anticipatory
- Calm - steady emotional state
- Considerate / respectful of other road users
- Rule compliant

Communicates well with other road users

Table 2: Behaviours that describe careful and competent drivers, by location and driving status.

1. Urban

Careful

- Driver:
  - Respectful of other drivers, aware of, thinking about and considerate towards others on the road.
  - Polite (not too polite though to confuse others).
  - Too careful is possible, a competent driver knows how careful one needs to be.
- Non-Driver:
  - Calm, observant.
  - Aware of other people, cyclists.
  - Leaves space.
  - Not chatting with friends, not on the phone.
  - Watches for blind spots.

Competent

- Driver:
  - Able to respond to situations in the 'grey' areas - unusual events.
  - Always up to date with the Highway Code.
  - Able to assess different types of situations.
  - Plan ahead.
  - Adapt to different driving conditions.
  - The driver might be competent but can sometimes drive in a non-competent / non-careful way.
- Non-Driver:
  - Drive like a driving instructor.
  - Follow the Highway Code, traffic rules.
  - Signals.
  - The car is 'generic', not tuned up.
  - Choose a lane ahead of time, and knows where they're going.
  - Thinking ahead, proactive.

2. Semi-Urban

Careful

- Driver:
  - Observant
  - Considerate / thinking about others
  - Disciplined
  - Patient
  - Nice
  - Inclusive
- Non-Driver:
  - Focused
  - Alert
  - Aware of surroundings

    - Staying calm
    - Drives to the right speed
    - Drives away from the pavement

Competent

- Driver:
    - Obedient
    - Responsible
    - Observe the rules of the road
    - Anticipation of others on the road.
    - Use indicators
    - Drive to the expected speed
    - Demonstrate basic driving behaviours
    - Be predictable to drivers
    - Make clear / “eye contact” with other drivers
- Non-Driver:
    - Experienced
    - Mindful
    - Following the law
    - Fit-to-drive
    - Good decision making
    - Vehicle is kept safe

3. Rural

Careful

- Driver:
    - Consideration for other road users (e.g. leaving more than enough space when passing other users)
    - Awareness of surroundings
    - Adherence to road laws / compliant
    - Able to park
    - Able to reverse
    - Recognition of physical fitness and cognitive ability to drive
    - Politeness & informal communication norms (e.g. thanking other drivers)
- Non-Driver:
    - Not speeding
    - Hazard perception
    - Anticipating
    - No road rage / avoiding driving while angry
    - Patient
    - Braking with plenty of time
    - Not using handsfree calls if this reduces attention (though not everyone agreed on this)
    - Not shouting at or intimidating pedestrians

Competent

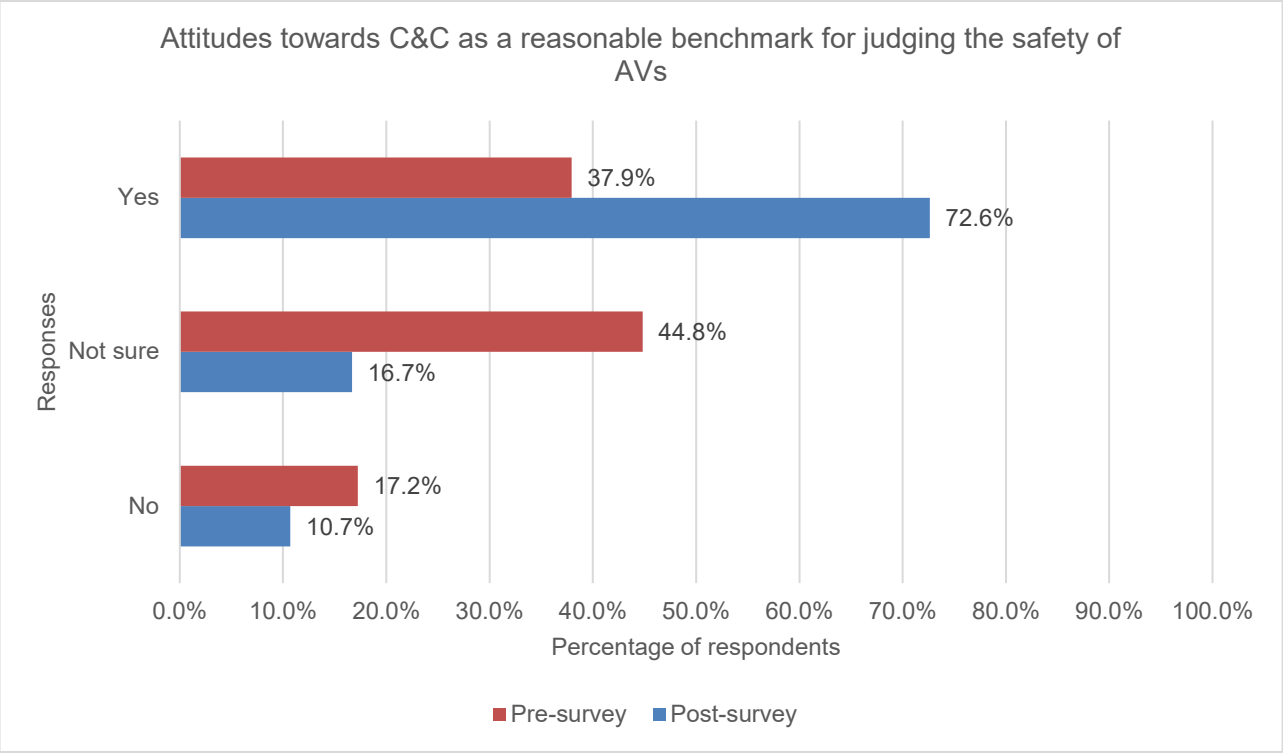
- Driver:
    - Able to take pre-emptive action.
    - Fully-focused.
    - Ability to drive safely in all conditions - judging when not driving.

    - Making the choice to not drive if conditions are poor.
    - Need to be safe to drive.
    - Observant.
    - Understanding what other road users can and can't do (e.g. understanding how cyclists use the roads - where they can and can't drive).
    - Smoothness of driving (not creating unnecessary disruption)
    - Managing blind bends and narrow lanes safely
- Non-Driver:
    - Not speeding
    - Not being under the influence
    - Mindful of others
    - Communicating through clear hand-signals
    - Aware of the vehicle’s ability
    - Good reaction times
    - Experienced on the road
    - Evolving / learning
    - Being able to park properly (specifically not mounting pavements)
    - Driving according to road conditions rather than strictly to the speed limit
    - Recognition that overconfidence can reduce competence

RQ1c: What are public expectations of an AV that meets the careful and competent benchmark (e.g. which scenarios must it manage)?
RQ1d: What behaviours/actions of an AV would need to be demonstrated to meet the careful and competent benchmark?

Participants throughout workshops discussed the value of the benchmark to understanding the safety of AVs. Activities in sessions explored changing opinions over the course of the workshop, as participants received different information on road safety, AV technology, and potential use-cases for the technology in the form of scenarios. These discussions illustrated a need for a benchmark that is clearly defined, with agreed measures and indicators, and which can be applied readily to both human drivers and AVs. Workshop participants regularly cited the need for clarity of terminology; often noting that C&C includes overlapping/interacting concepts that are often subjective in nature and thus difficult to measure. This presented an issue to participants who were concerned about the practicality of using C&C as a benchmark for automated vehicles. Pre- and post- workshop testing of views on the benchmark are illustrated in Figure 1:

Figure 1: Do you think C&C is a reasonable benchmark for judging the safety of automated vehicles? (Percentages, %)



Base n: Pre-workshop survey = 87; Post-workshop survey = 8

A variety of behaviours were highlighted by participants that describe the expected behaviours of C&C automated vehicles. Key expected behaviours of AVs that meet the C&C benchmark for drivers were:

- Considerate, human-like driving behaviours - although in discussion a small number of participants referenced the need for obviously automated behaviours to distinguish between human and automated driving.
- Legally compliant with driving standards.

- Regularly maintained and with fail-safes.

Key expected behaviours of AVs that meet the C&C benchmark for non-drivers were:
- Adaptive to complex environmental conditions.
- Legally compliant / adherent to the rules of the road.
- Transparent, clear communication to all road-users.

Expected behaviours are further outlined in Table 3.

Table 3: Expected behaviours of AVs that meet the C&C benchmark

To meet the C&C benchmark AVs will need to…

1. Urban

- Driver:
  - Follow driving rules.
  - Communicate with people around and inside vehicles.
  - Be “exceptionally” C&C.
  - Be maintained to the right standard - the condition of the vehicle is expected to be good, MOT.
  - Provide confidence to users that the vehicle is in good condition.
  - Manage potholes / poor road surface conditions.
  - Read human driver behaviour and body language.
  - Manage moral decisions.
  - Be considerate.
- Non-Driver:
  - Observe and judge body language, see inside the car
  - Help other road users feel that AV will be able to react when they interact with it / it is to behave as expected.
  - Explain to passengers what it is doing, that it is seeing what’s around it
  - Maintain a high standard against the benchmark, and ensure that service providers do not deteriorate in quality once approval is gained.
  - Demonstrate common sense decision making.

2. Semi-Urban

- Driver:
  - Drive as if to demonstrate it has the “human touch”; e.g. drive as a safe driver would.
  - Acknowledge other drivers and their movements; and those merging to roads from junctions.
- Non-Driver:
  - Drive safely at night.
  - Drive safely in difficult weather conditions, including snow.
  - Be aware of other drivers, particularly cyclists - driving too close to cyclists was an issue raised.
  - Drive slower than human drivers due to perceived unpredictability
  - Brake well in time e.g.no sudden braking.
  - Be recognisable as an AV, but not too recognisable as to distract e.g. cause rubber-necking from other road users.
  - Drive into roundabout or junction without being indecisive (e.g. roundabout merging scenario)
  - Anticipate hazards and drive to the situation (e.g. child in the road scenario)

3. Rural

- Driver:
  - Understand road signs.
  - Identify different types of road users, including horses and motorcyclists.

    - Pull over for emergency vehicles.
    - Adapt to all conditions, e.g. in fog use correct lights and speed.
    - Acknowledge other drivers (e.g. merging from a junction).
    - Be able to navigate complex situations when safety is compromised.
    - Be able to operate safely with safeguards in case of software corruption or WiFi drop-out.
    - Have clear rules in place for maintenance responsibility (e.g. keeping sensors clean)
    - Recognise informal road communication (e.g. hand signals, headlight flashes).
    - Cross road markings where necessary for safe manoeuvres.
    - Not be overly cautious to the point of frustrating other road users
- Non-Driver:
    - Park in the correct place.
    - Stick to the laws, unless avoiding an accident.
    - Follow the Highway Code very closely (no deviation).
    - Operate with fail-safes which include human operator intervention if needed.
    - Proportional reactions to hazards (e.g. no overcautious responses to hazards like birds).
    - Mixing safely with human drivers in unpredictable scenarios.

RQ1e: Would the public assess a careful and competent human and an AV in the same way?

There was some variation in workshops as to how the public assesses C&C for human and AVs:

Urban drivers often expected AVs to exceed human capabilities; particularly in terms of hazard perception and situational awareness. For instance, some suggested that because AVs have 360-degree vision and faster reaction times, they should be able to detect and respond to hazards more effectively than human drivers. Furthermore, some highlighted that AVs could avoid behavioural limitations such as stress or poor decision-making under pressure. **Urban non-drivers** noted that certain competencies are expected of an AV, even if not from a C&C human driver; for example, being able to correctly react to a road sign even when it's obscured. There was a sentiment that they need to have capabilities people know vehicle computers currently have e.g. access to up to date traffic maps. This cohort also believed that AVs should learn from each other and communicate readily sharing key information about the road environment (e.g. information about roadworks).

Semi-urban non-drivers' attitudes frequently changed between acceptance or rejection of the benchmark over the course of the workshops, as participants reflected on the new information discussed. Some non-drivers noted that automated vehicles "*have to be better than human drivers for (adoption) to make any sense."* Participants also noted that "*careful"* might be considered to be too 'blase', 'subjective' or 'too human' - and not a good description for a technology or process. 'Competent' for this group was not the correct word, with participants stating there is likely to be a better word.

Some **rural drivers** suggested that "safe and competent" would be a better benchmark than C&C for measuring automated vehicles (AVs), as "safe" is less ambiguous and would help gain public confidence. There was also some consideration to what participants described as a challenge with strict Highway Code compliance and real-world driving norms. Some suggested that rigid rule-following by AVs could generate frustration among human drivers, potentially triggering unsafe behaviours. There was also discussion about whether Highway Code rules (e.g. stopping distances) might require review given recent technological advances. **Rural drivers** noted that assessing AVs against a benchmark would require a set of standards that are higher than that required of human drivers. Some frequently expressed the expectation that AVs should "*make no mistakes*" whilst acknowledging that in mixed traffic environments, complete elimination of risk may not be possible due to interaction with human drivers.

Contextual variation in the driving scenarios and perceived ambiguity in the C&C benchmark meant that some participants believed more consideration of other terminology is important:

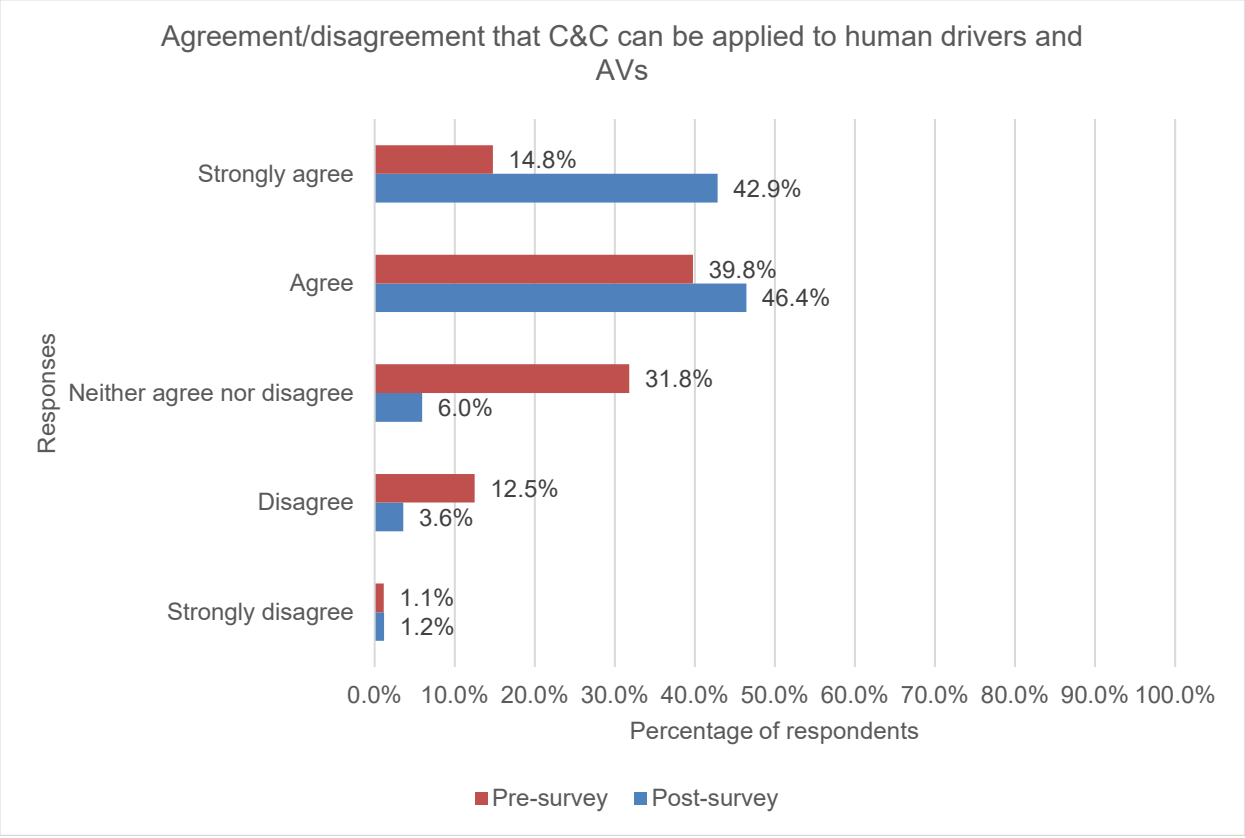
"I feel careful and competent is up for debate, as was evident in our discussions. There needs to be a deep definition of what this means on a situation-by-situation case." Semi-urban driver

However, we found other participants were happy with the benchmark and its application, noting it met the needs required - with the value of prompting industry to work towards achieving a higher safety standard overall:

“I think that “careful and competent” is pretty effective, and I couldn't think of anything better. I would still expect the company running automated cars to hold themselves to a higher standard than an average human road user and they should be closely monitored and audited by themselves, a government agency and hopefully a neutral/independent agency too.” Semi-urban driver

Participants were asked to consider whether the C&C benchmark can be applied to both human drivers and AVs in the pre- and post-workshop survey. The findings highlight that following the workshops participants were more likely to state that the benchmark was applicable to both human drivers and AVs (Fig 2):

Figure 2: Participants’ agreement/disagreement with the idea that the “careful and competent” standard can be applied to both human drivers and automated vehicles. (Percentages, %)



Base n: Pre-workshop survey = 88; Post-workshop survey = 84.

RQ2: How would the public expect ‘careful and competent’ driving to be demonstrated?

Urban drivers described C&C driving as involving strong situational awareness, adherence to the Highway Code, and consideration for other road users. When assessing AVs, they expected this standard to be demonstrated through robust evidence of safety

performance, emphasising "*data, data, data*"; including video evidence of complex scenarios, simulation testing, independent safety certification, and aggregated user feedback.

Urban non-drivers associated C&C driving with observation, calm decision-making, and awareness of others on the road, but noted that judging AV behaviour can be difficult because people lack the clear framework that they have for human drivers as a result of their day-to-day experiences. They placed strong emphasis on clear and independent assurance, viewing regulator certification, defined Operational Design Domains, and video evidence of AV behaviour as essential, with participants noting that "*seeing is believing*".

Semi-urban drivers considered testing and trialling to be important, and that public demonstrations of AVs are needed to allow people to gain firsthand experience of the technology. Demonstrators and testing were considered one of the ways through which the disagreements over the C&C standard could be clarified.

We found instances where participants recognised C&C as in their current experience of mobility. For example, **semi-urban non-drivers** noted that public transport displays 'careful driving' through posters and stickers on vehicles - which one participant believed had '*conditioned*' them to consider safety important when using buses as a passenger. One participant noted that they assume all public transport is safe.

Rural drivers' expectations for C&C driving on rural roads included awareness of wildlife and farm animals, farm machinery, and the ability to perform manoeuvres e.g. reversing on single-track lanes. The importance of vehicle maintenance was also noted for vehicles used on rural roads.

Rural non-drivers expressed a broad belief that AVs should adhere to the highest safety standards, e.g. adhering to the Highway Code and following all rules given that AVs have less risk of human error, and expressed that multiple sources of evidence would be needed to demonstrate this. For example, multiple sources of video evidence of AV driving quality, beyond that which is provided by the commercial service provider. They emphasised that video evidence is convincing because it shows the capabilities and limitations of the AV. Participants also considered testing and trialling to be important, with communications and trials that were specific to Norwich and driving in Norwich. Live demonstrations and in-person events were preferred by some participants over video evidence alone, alongside communication across multiple channels to reach digitally excluded groups. Clear, jargon-free explanations of how AVs would handle uncertain or unusual situations were seen as essential to building trust.

RQ2a: How would the public benchmark a careful and competent human driver?

Feedback across workshops highlight various factors of consideration for benchmarking a C&C human driver. Across workshops, participants were able to make comparisons between different styles of driving, and under different conditions, highlighting where they think C&C behaviours were apparent. Workshops explored the benchmark in comparison to existing approaches to driving, which are further outlined below:

Urban drivers referenced the need for AVs to drive to the same standard as a driving instructor e.g. *"As a driving instructor of 22 years, the benchmark for AVs should be the equivalent of an experienced driving instructor"* (Urban driver). This was deemed adequate so they could demonstrate skills expected to be tested during an advanced driving test, something the public could trust. Aggregated user feedback - such as 5-star ratings on Uber - were also viewed positively, as participants felt broader user experience data would be more reliable than individual anecdotes. Urban non-drivers also recognised driving instructors and the driving test as an existing benchmark reflective of what it means to be C&C.

The **semi-urban** workshop highlighted differences in opinion as to the standard in relation to the information shared as a driving quality spectrum- from dangerous driving through to C&C driving. **Semi-urban drivers** thought that the comparison to a benchmark is important, but questioned whether the C&C benchmark is the right comparison for human drivers: "*When you pass your test you are competent, but not careful?*" and; "*If you let someone out of a junction you are polite …but is that careful or competent?*". We found variation in the level at which participants believed C&C should be set. Among the **semi-urban non-driver** cohort some participants noted that the C&C benchmark should be considered to be equivalent to the driving standard for the top 5% of UK drivers, whilst others noted that "*C&C*" must signify a higher standard of human driving - for example one participant noted that 50% of UK drivers are C&C and considered the 'average driver' and 'C&C driver' to be equivalent.

Rural drivers considered C&C driving to be important in the rural context, and noted that there were varied driving skills that were needed to be considered C&C in the rural context. One participant noted that, as a motorcyclist, they considered C&C drivers that were aware of motorcyclists and able to accommodate them on rural roads. Another **rural driver** noted the importance of being able to anticipate farm machinery or wildlife when driving on rural roads. Within the **rural non-drivers** group there were clear examples of careful and competent driving being important for driving in small rural villages and towns, with **non-drivers** noting that C&C described a smaller proportion of drivers on roads.

RQ2b: What outcome measures would be most trusted to confirm a careful and competent driving behaviour?

There were similarities across groups regarding the outcome measures that would best demonstrate C&C human driving behaviour, with participants recognising the value of standard measures such as KSIs to describe road safety. Key points from each location are highlighted below:

Urban drivers and non-drivers both recognised that traditional road safety outcomes, such as a reduction or lack of incidents and lower Killed or Seriously Injured (KSI) rates, provide a useful indication of C&C driving but emphasised that outcome measures by themselves do not confirm whether someone is a C&C driver. Instead, **urban drivers** highlighted the importance of observable behaviours and technical performance, with aggregated reviews from human users or personal experience with AVs considered a way to build trust in them. **Urban non-drivers** similarly stressed the need for independent and credible evidence, such as regulator certification or video demonstrations of AV behaviour in complex scenarios.

Participants in the **semi-urban driver** workshop were mindful of the importance of KSI information for demonstrating road safety, but were aware of the diversity of responses across their group when prompted, and reflected that this may point to low levels of awareness of road safety among the public and that the public sees road safety differently according to various factors: *"this conversation has shown that people drive differently but all believe they are competent - they believe they are safe…clearly none of us knew the statistics earlier..."* As a result, KSI data was considered an important outcome measure of the C&C benchmark. The **semi-urban non-driver** cohort also recognised the importance of the KSI measure.

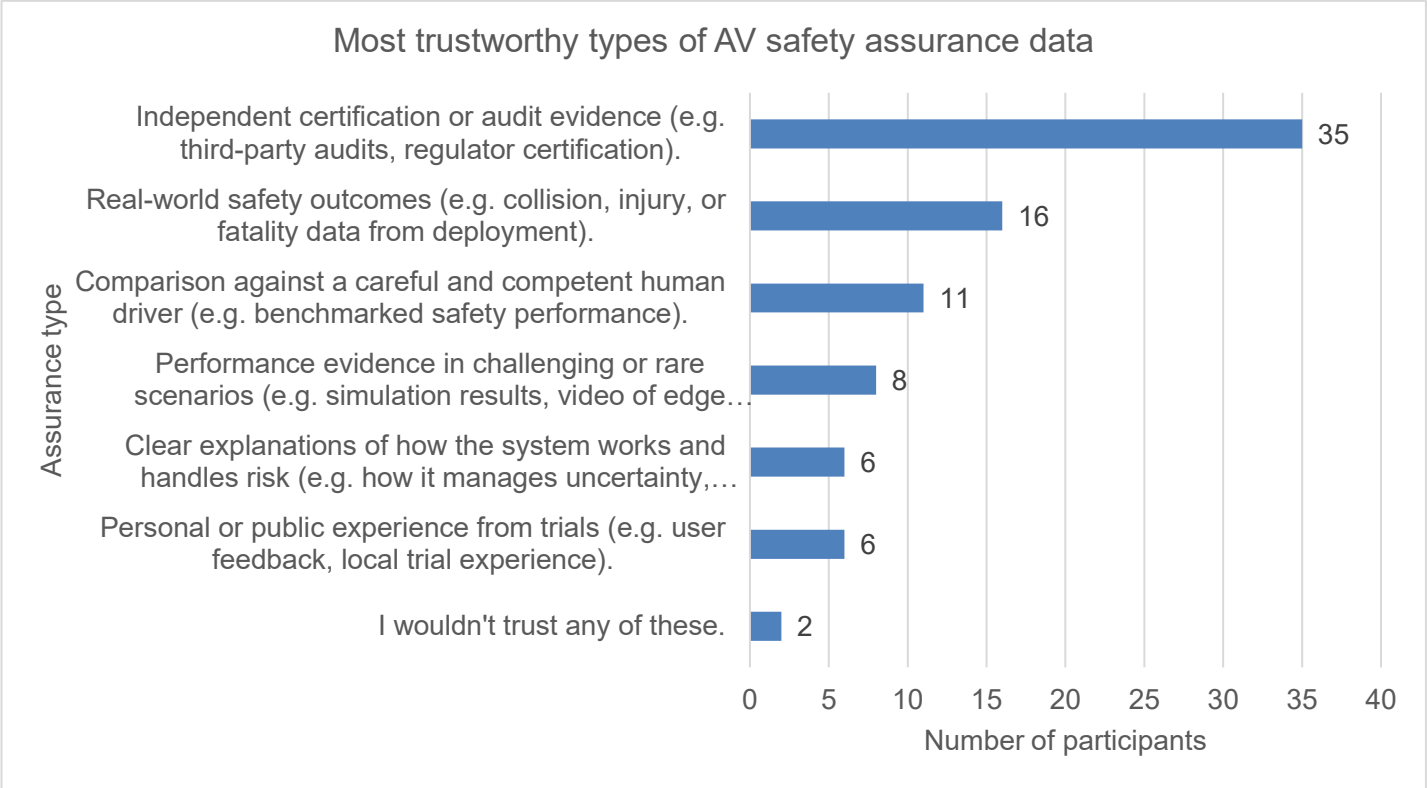
The importance of KSIs as an outcome measure of C&C was further confirmed by **rural participants**. We found that several **rural drivers** remained 'moderate' in their position of C&C as a description of most human drivers today, with one person noting that the figures only reinforce their belief that continuous re-education is necessary for all drivers. We also found that a minority of **rural non-drivers** did not change their opinion throughout the discussion on C&C behaviours of human driving, and noted that specific driving behaviours in rural villages highlighted that most behaviours are not C&C: *"I think mostly around my village there are too many people speeding, cutting corners and not looking out for pedestrians".* Views were further cemented during discussion following the KSI discussions whereby some participants changed their initial "*yes*" answer to *"no*", citing the frequency of accidents implied by the high KSI numbers and observing that accidents are a common occurrence on A roads in their local area:

"I've changed my mind because seeing those numbers and then actually thinking about it, it's almost on a daily basis you hear there's an accident on the A47 … it actually does happen quite a lot and it is mainly on those dual carriageways…obviously people are not being careful if there's so many accidents".

RQ2c: How would people expect careful and competent AV capability to be demonstrated before getting on the road? | RQ2d: What measure would people expect to be captured about AV driving to demonstrate careful and competent driving?

We found participants were interested in measures and methods that build trust through third-party assessments and supported independent assurance of AV technologies. In the post-survey independent certification and audit evidence was selected by over two-fifths (41.2%) of participants, whereas real-world safety outcomes data was chosen by just under one-fifth (19.1%). Whilst this was not a quantitative study, the selection in the surveys aligns with the selection people made during the deliberation process, suggesting that this type of information may be more likely to enhance public trust in the technology (Figure 3).

Figure 3: Which type of information would you trust most to show that an automated vehicle is 'safe'? – Select one option (frequencies).



Base n: Post-workshop survey = 84

Across workshops, there was agreement that data demonstrating AV safety performance was important and should be actively shared with the public. **Urban driver and non-driver participants** preferred to have 'all the evidence' presented to them, and to be able to then make a decision using that which was important to them - noting the need for the public to have evidence which meets their information preferences. Participants in the **urban non-driver** group also reflected that part of the reason for asking for all the data is because the technology is new. "*When something's just starting out you want to know a lot more about it as opposed to, like, with a plane since it's already been established for so long. You just trust in all the past generations and everything that's happened beforehand.*" Participants explicitly recognised that their answers to this activity might change considerably with experience or familiarity with AVs and that their standards for proof of safety would be much higher before deployment than after. We also found that measures that track safety performance over time in a consistent manner were rated higher than others when pressed to prioritise. One **semi-urban non-driver** did note that they would not need to actively seek AV safety data and information, and instead would trust any communication from the government about AV safety - assuming that all testing had been done, and that regulations signalled that approval was given.

Independent/third-party sources:

Independent/third party sources of information and validation were highly rated by participants in several cohorts. **Rural non-drivers** were positive towards independent agencies and third-parties:

"We just felt that it needs a third party…you can't have one company say they can come and do it - you have to have someone else authorising."

Semi-urban drivers shared these views and required clear information ahead of deployment, with safety assessments from independent third-party sources being particularly important. One **semi-urban driver** noted that certification from one of these sources was useful to overcome the “fear factor” of adoption. **Semi-urban non-drivers** were of a similar opinion, with both groups in the workshop activity coalescing on independent information which showed some level of rigor or analysis. This was considered important by the **semi-urban non-driver group** as a pre-deployment measure, which should include clear certification and outcomes from audits. This group considered this type of information to be basic and should form the minimum requirement before deployment. The **urban driver** group highlighted that safety certification needs to be independent and not come from the manufacturer. **Urban non-drivers** mentioned they would trust the Government, DfT, or DVSA, but would explicitly distrust information coming directly from the AV companies who were seen as having a strong commercial incentive to accelerate AV deployment.

Regulators were also considered to be important in establishing the C&C benchmark as a standard:
“C&C” should have a standardised, concrete set of criteria agreed upon by regulatory bodies internationally” Semi-urban non-driver

Service performance measures and comparison data sets

Participants in several workshops noted that they would value having access to performance measures of services as they are deployed, with post-deployment data on service performance which would be actively shared by service providers and regulators which demonstrate key indicators. **Semi-urban non-drivers** cited the value of ongoing data on incidents, and vehicle malfunctions or errors, as useful and supportive in building trust. **Rural drivers** also noted that information on vehicle malfunctions was useful to understand pre-deployment, for instance considering digital connectivity. They specifically highlighted potential issues with software updates for AVs, asking how the system would respond to connectivity loss in rural settings. This group also agreed that a comparison of AV safety performance against a C&C human driver was essential and convincing. As an opposite view, an **urban driver** felt they didn't need this as ‘*a computer is far more intelligent than a human,*’ but acknowledged others might, noting the dependence on prior attitudes:
"For me, I would be quite happy to get in an AV now, knowing very little about them. I don’t necessarily need to see the analytics because I think that the AV criteria is far better than a human. Like, a computer is far more intelligent than a human and therefore I trust that more – so that would be useful but not essential, but other people who need convincing might be different."

AV behaviour videos and simulation examples

Rural drivers considered video evidence of AV behaviour through virtual simulations, for example in complex and edge-case scenarios, essential. **Rural drivers** emphasised that deployment videos must clearly distinguish between careful and competent behaviours. This is a key distinction because participants often conflated the two; for example, a vehicle making a complex, planned decision might be viewed simply as 'competent' rather than deliberately 'careful.' Participants sought visual evidence to help them understand

these nuances. Video evidence of AV behaviour in complex and edge-case scenarios was also considered essential by **urban drivers** during discussion who highlighted that *"seeing is believing"*.

User ratings and reviews

Across the workshops there were various preferences for the different data sources for building trust and demonstrating AV capabilities. For example, testimonials from experts or friends and family were considered less useful due to the potential for bias, whilst aggregated ratings scores from past users held value. This was demonstrated by some **urban participants** who considered individual user testimonials non-essential but placed some importance on high service-ratings. The source of ratings and reviews were important to participants. For example, we found differences of opinion on the importance of testimonials among the **rural non-driver** cohort. The value of testimonials from friends, or from media and magazines was questioned: *"(For magazines) I would question what they're being paid".* **Semi-urban non-drivers** considered post-deployment ratings of services to be a basic expectation for service operations.

Testing and trialling in context

Rural non-drivers considered testing and trialling to be important, with communications and trials that were specific to driving on rural roads in their local context. Live demonstrations and in-person events were preferred by some participants over video evidence alone. Related to this, the **rural drivers** and **urban non-drivers** considered information about the Operational Design Domain (ODD) as important pre-deployment: suggesting that it would help local residents to understand whether the AV would be able to operate within their local road network:

"(Our group liked) one card that mentioned showing evidence of the area that the cars would come into that they could deal with the area and deal with certain situations in Norwich." Rural driver

Some people in the **urban driver** group noted that real world evidence from other countries is a convincing piece of evidence for them for deployment in the UK.

Pre- and post- deployment communication preferences

Preferences for plain language explanations of the technology, how it operates, its potential benefits, and the way in which rollout is being managed came up across workshops. Participants noted the lack of clarity in terminology meaning that communications will need to be clear regarding definitions.

Communication channels were also important. **Urban participants** were interested in a mix of engagement methods to share information about AV services in their area. They noted the need for major campaigns, coupled with '*bite size info*' such as short-form video content, so people can understand the technical details about AV safety at key points. They also noted that real world examples of AV use in other countries (e.g. rollout in San Francisco) would be beneficial to build their trust. **Urban drivers** requested clarity from industry and the government on the protocol of '*what happens after an accident*' for both

AV users and other road users. **Rural non-drivers** noted the value of communication across multiple channels to reach digitally excluded groups. Clear, jargon-free explanations of how AVs would handle uncertain or unusual situations were seen as essential to building trust. **Rural non-drivers** stressed that post-deployment monitoring should be ongoing and transparent, with system-wide reviews triggered if faults were identified. Some suggested that regular safety checks should be publicly reported. There was a strong emphasis on ensuring that communications were accessible to digitally excluded populations and that multiple independent voices were involved in public messaging.

Additional findings

Attitudes towards road safety, AVs and intention to use an AV service

Deliberation throughout the course of the workshop appeared to have a couple of effects on participant attitudes:

- Discussion, data and information on KSIs and scenario-based testing appeared to elicit responses that highlighted participants were more aware of road safety and current driving standards as a result of the deliberation process. A minority of participants in both driver and non-driver cohorts who were less aware of road safety noted increased awareness of the importance of improving road safety.
- Knowledge of AVs and their potential road safety benefit also appeared to improve, as a result of information sharing and access to a knowledgeable expert. Participants tended to recognise the net-safety benefit of AV, but noted the importance of transparent, accessible and safe rollout of the technology. In no workshop were AVs expected to generate negative or unsafe road conditions e.g. decreasing the current safety standard of human drivers.

When asked if they would be willing to use an automated bus service, some **urban drivers, most semi-urban drivers** and **rural drivers** expressed reluctance to adopt immediately stating that they would wait to see sufficient data before trying out the service. However, the potential for AV public transit to solve unemployment issues for individuals who cannot drive due to various conditions, such as those with various health conditions, was recognised as a positive factor. **Semi-urban non-drivers** were also mixed - with one noting that they would not wish to ride alone but may be persuaded if they were able to use it with friends or family. **Rural non-drivers** differed - when asked several indicated they would use an automated bus service, citing the belief that the systems would be thoroughly tested and a desire to embrace future technology. Concerns were raised about who would manage incidents or passenger related disruption or safety issues on the bus. **Urban drivers** and **urban non-drivers** were mostly positive and indicated that they would be willing to try a service.

Conclusions

The safety argument for AV adoption resonated with participants across workshops. Road safety statistics including trends across modalities elicited surprised and concerned emotional responses, and appeared to clarify to participants the safety potential of AVs:

- Deliberation across the workshop series highlighted road safety concerns with drivers and non-drivers in all settings broadly agreeing that current driving standards present some level of risk to safety, with mixed responses across groups regarding whether current driving standards meet the C&C benchmark. Most drivers believed current driving standards to meet the C&C benchmark, whilst non-drivers believed current driving standards were below the C&C benchmark.
- There was a clear difference between driver and non-driver groups in the level of safety risk they perceived when using roads today. Whilst drivers tended to recognise an acceptable driving quality in most cases, there were many instances of non-drivers noting high-risk and unsafe driving behaviours.
- Over the course of deliberation, there was an increased belief that AVs present a potential net benefit to road safety; driven largely by an increased understanding of technical capabilities of AVs. Deliberation also appeared to influence those who were less concerned about road safety in the first instance to better appreciate the severity of the issue.

The C&C benchmark received moderate to positive support, but with notable reservations about its clarity and clear feedback that a robust standard is needed with measurable requirements to ensure safety.

- Firstly, participants understood the need to establish a benchmark for AVs, saw that this was the role of government and outlined their confidence in the system establishing such a benchmark. No participant suggested that private companies should set the standards or '*mark their own homework'*; rather, participants explicitly and often without prompting stated that they are against the involvement of manufacturers or others who stand to financially benefit from AV deployment to be involved. This is a strong starting point when looking at the C&C benchmark, as there was no argument against its core ideal.
- Moderate support highlighted that the benchmark is the best available today, but that there are likely better terms that can be used to make it clearer, less ambiguous and more measurable. Whilst most people did not question the relevance of either careful or competent as descriptors, there was a suggestion that these two words do not cover everything people would expect to hear. Specifically, participants requested the addition of further characteristics or for the existing ones to be supplemented with concrete behavioural explanations. Participants noted the importance of establishing a benchmark of a higher standard than 'average driver',

with most reflecting that C&C must show an improvement on current driving standards.
- Negative feedback on the benchmark often pointed towards two key issues. First, some participants felt that 'C&C' did not go far enough to meet the higher safety standards they expect from automated vehicles. Second, many initially did not interpret 'C&C' as representing a safety level above the average driver. The subjectivity of C&C was a barrier to some and pointed to the potential for different understanding across the population. Furthermore, non-drivers were more likely to believe that most drivers on the road do not meet the C&C standard, illustrating differential perceptions of risk based on their lived experience. Deliberative exercises highlighted that further explanation or visual aids were needed to define the benchmark as a high driving standard. Once participants understood this the benchmark received more support, showing that the C&C definition alone was not self-explanatory.

Participants across workshops were readily able to describe C&C in terms of behaviours rather than expected driving outcomes, and could outline their expectations of AVs meeting them. There was however clear challenge to the broad definition of C&C:

- Workshop participants defined C&C driving in terms of human driver or AV behaviours, and were able to list many expected behaviours for both. The behaviours cited provided insights into how the benchmark might be defined. However, the significant number of behaviours that participants offered as examples highlights the perceived breadth of the term which may act as a challenge to defining comprehensive criteria that capture how it is achieved.
- Whilst the majority of behaviours cited by participants may be easily translatable to an AV, and described within the SOSP, we found that some may be more difficult to formalise - such as displaying common sense or intuition. These behaviours were important to several participants in describing C&C in their context - and therefore consideration will be needed as to how best to convey these important driving attributes through the SOSP and any complementary materials associated with it.
- One participant gave the concrete example of the ‘advanced-driving exam’ as a good benchmark, where a human driver is doing their utmost to pass whilst being supervised and scored. This was positively regarded by other participants as a comparable test.
- Participants consistently held AV behaviour to high safety standards. When presented during deliberation with a road incident scenario, participants rarely found sufficient mitigating circumstances and noted that a C&C driver would have avoided the incident. Participant expectations therefore aligned closely with the definition of C&C which was tested, and sets a very high benchmark for the SOSP. We must note that these activities were delivered once participants understood that C&C was a high benchmark.
- Participants described an expectation of AVs to be better than human drivers in aspects where computers traditionally outperform humans, such as reaction time, constant 360-degree vision, communication with other AVs and knowledge of the road infrastructure. These requirements were in addition to AVs meeting the C&C benchmark.
- Context plays a role in C&C application and understanding. Behaviour attributes gathered across the two population groups and three locations highlight that there are commonalities and differences worth considering. This difference was most apparent when considering rural drivers and non-drivers who recognised different

driving behaviours related to speed, and driving conditions, compared to urban drivers. This also reinforces the sentiment that expectations differ across segments of the population who may have different expectations of which C&C behaviours are necessary. Participants therefore consider vehicle ODDs intuitively and when presented with the term it was something they understood and agreed on its importance for AV rollout.
- Discussions of KSIs and C&C behaviours supported participant engagement and dialogue on the benchmark, how it may be applied, and how its use may be measured. Participants did not readily make associations between C&C and incident statistics or other quantitative measures: it was predominantly considered to characterise a range of driving behaviours.

Demonstrating adherence to the benchmark is important to the public, and across workshops the role of third-party and independent actors was salient to building trust and assurance that AVs are being introduced and operated in line with requirements.

- Participants placed the most trust in independent and third-party certification and audit, with information provided by private companies generally considered unreliable. People felt that government organisations like DfT, DVSA, VCA would be relevant organisations to carry out a function of certification and providing information to the public.
- Real-world safety outcomes and video evidence of driving behaviour in complex scenarios were considered another reliable source of evidence. However, participants highlighted a desire for AV performance and safety data to be openly available, enabling those seeking reassurance to examine objective statistics for themselves.
- There is an expected time dimension to building trust. It was highlighted that over time the public would likely be less concerned about AV technology as a result of exposure, familiarity, and direct or indirect experience, and therefore would require less performance and outcome data or evidence. However, there were also expectations that the safety benchmark would be maintained or improved over time to ensure AV companies do not scale back on their ambitions and must continuously improve safety performance.
- Participants discussed a need for clear, relevant and easily understandable communication of the safety benchmarks. There was no clear priority in communication content or channel, instead preferences were towards the broadest possible coverage of media channels to ensure accessibility across different demographics. More insight is required on the most effective ways to communicate and update on safety performance, with clear evaluation of the effectiveness of those channels in building acceptance and enabling the public to make informed decisions regarding the adoption of AVs.

Recommendations

Prior to deliberation we found wide variation in public views as to current driving standards. However, deliberation which included access to KSI data and road safety experts, generated a broad consensus that current human driving standards present a significant safety risk, and as participants better understood the technical capabilities of automated vehicles (AVs), they increasingly viewed AV adoption as a potential net-benefit for road safety. In this context, the "C&C" (C&C) benchmark received moderate-to-positive support as a foundational starting point, closely aligning with public expectations of a high, "advanced-driving exam" level of safety that would be sustained and monitored over time. However, participants expressed notable reservations regarding the benchmark's subjectivity and ambiguity. To be effectively utilised as a regulatory standard, the public emphasised that C&C must move beyond the abstract – particularly in relation to human traits such as ‘common sense’ and ‘intuition’ – and be defined by concrete, measurable behavioural attributes that account for varying contextual expectations that are apparent for drivers and non-drivers when using the UK’s roads.

Crucially, this research highlights that public support for the benchmark is contingent on the visible role of strict independent oversight and transparent communication from key stakeholders across the AV ecosystem – including developers, service providers, and the media. Participants explicitly rejected the idea of industry "marking their own homework," and instead looked to government bodies (such as the DfT, DVSA, or VCA) to manage third-party certification and auditing. For such an organisation to ensure certification and hold the industry accountable, the benchmark must be not subjective but measurable.

While participants acknowledged that trust in AVs will naturally grow over time through exposure and familiarity, they concurrently expect continuous improvements in safety standards and open access to real-world performance data. In order to achieve broad acceptance, the public expects measures of the C&C benchmark, and particularly KSIs to show marked improvement on current levels. Doing so will ensure that public expectations around AV safety performance are met and they are reassured that the developers and operators of this new technology are being held to relevant, rigorous, and publicly accountable standards.

The following recommendations are made as a result of this study:

Recommendation 1: Ensure behaviours form the basis of the C&C benchmark and SOSP.

We found moderate-positive understanding of the benchmark when initially introduced and participants were readily able to describe behaviours for the C&C benchmark for human drivers, and extend these to expectations for AVs. The SOSP should therefore be framed so that the required behaviours of AVs are clearly specified, with safety outcomes

understood as the result of those behaviours rather than as evidence from which C&C behaviour is inferred. In practice, this means the principles should illustrate what C&C driving looks like via clear examples of how AVs should behave in common manoeuvres or situations, as well as complex and challenging circumstances.

Recommendation 2: Further research is required to test elements of the C&C benchmark and SOSP across key demographics and explore the behavioural dimensions of the benchmark related to adoption and acceptance of AVs and trust of different parts of the AV ecosystem.

Findings from the deliberative workshops suggest that while the C&C benchmark is broadly acceptable in principle, the terminology alone does not consistently convey to the public that it represents a safety standard above that of the average human driver. Participants across all groups articulated a wide range of behaviours associated with C&C driving, which highlights the need to determine which are most meaningful and important to the general public. Further research with a larger, socio-demographically representative sample would allow testing of different behavioural framing and examples that could supplement the benchmark and make its meaning clearer. In particular, identifying a small number of highly recognisable behaviours that illustrate what constitutes “careful” versus “competent” driving could help reduce confusion and bridge the gap between the abstract benchmark and concrete driving behaviour; reinforcing that the standard reflects a high level of safety.

Such research would also allow for the exploration of how these behavioural expectations vary across different road users groups and demographics. The deliberative workshops indicated that drivers, non-drivers and other road users (such as cyclists) may prioritise different behaviours, but the non-probability sample method and sample size did not enable us to make statistically robust conclusions. A further area for exploration relates to participants’ expectations that safe driving includes elements often described as ‘common sense’ or ‘intuition’ which cannot be as readily translated into technical or regulatory requirements within the SOSPs. This research could therefore also examine how such qualities can be meaningfully translated into AV driving behaviours, which may help ensure that the SOSPs and related communications effectively convey that AVs are capable of managing the kinds of nuanced situations that human drivers often handle through experience and judgment.

Recommendation 3: Outcome measures for deployments should be detailed to support adoption across various ODDs, and support industry to rapidly deploy C&C measures and report on progress to Government and the public.

There is clear support and acceptance of the need for a safety standard for the deployment of AVs. However, without defined and robust outcome measures participants expressed concerns that standards would not be upheld. We recommend that further work is undertaken as part of the development of the SOSP to define robust outcome measures and indicators of success regarding C&C of the SOSP.

The research suggests that whilst improved road safety is a must, most people will judge AVs based on the way they behave so the requirements and testing protocols of AVs should reflect this. Certification by an independent body involving the collection of video and other evidence would help to increase trust in the verification process and support the public to make an informed decision to adopt AVs. This is likely to be crucial during the early stages of deployment, whilst the public gets familiar with the technology. Over time, our research suggests that the key requirement is to maintain transparent records of AV performance statistics and ensure certification is an ongoing process with a continuously updated methodology.

Recommendation 4: Further research is required to understand the most effective way to communicate the SOSP and C&C benchmark.

This study has highlighted the high safety standard the public expects for the introduction of AVs, and the value of deliberation on road safety to build awareness of the safety case of AV technologies. Large scale deliberative exercises can support exploration and awareness building of AVs, and the potential value they can bring to communities by recognising the importance of local context to buy-in and acceptance. As such further work should be undertaken to explore how the public can shape and build local buy-in to AV adoption in their community.

We found that people often struggled understanding the general context of AVs until some of their very specific questions were answered. For example, what a passenger in an AV would have to do in an incident or how an AV would behave if there was heavy snow. This represents a challenge as there is large variety of scenarios that an AV might encounter on public roads, but it also might be an opportunity to not only focus on holistic messaging but also to determine the selected set of examples that are most persuasive in terms of demonstrating C&C behaviour.

Market testing of messaging and information should be undertaken to support public engagement campaigns, with the roles of trusted organisations across the road safety ecosystem being well defined to support the government to bring the public up to speed on work underway to ensure safety in this market.