In cars (are we really safest of all?): Interior sensing and emotional opacity

McStay, Andrew; Urquhart, Lachlan

International Review of Law, Computers & Technology

DOI:
10.1080/13600869.2021.2009181

Published: 01/09/2022

Publisher's PDF, also known as Version of record

International Review of Law, Computers & Technology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cirl20

In cars (are we really safest of all?): interior sensing and emotional opacity

Andrew McStay & Lachlan Urquhart

To cite this article: Andrew McStay & Lachlan Urquhart (2022): In cars (are we really safest of all?): interior sensing and emotional opacity, International Review of Law, Computers & Technology, DOI: 10.1080/13600869.2021.2009181

To link to this article: https://doi.org/10.1080/13600869.2021.2009181

INTERNATIONAL REVIEW OF LAW, COMPUTERS & TECHNOLOGY
https://doi.org/10.1080/13600869.2021.2009181

In cars (are we really safest of all?): interior sensing and emotional opacity

Andrew McStay[a] and Lachlan Urquhart[b]

[a]Digital Life, School of Languages, Literatures, Linguistics and Media, Bangor University, Bangor, UK; [b]Technology Law, School of Law, University of Edinburgh, Edinburgh, UK

ABSTRACT

This paper analyses expert and regulatory perspectives on car driver-monitoring systems that measure bodies to infer and react to emotions, fatigue, and attentiveness. Developers of driver-monitoring systems promise increased safety on the road, alongside comfort for cabin occupants through personalisation and automation. The impetus is three-fold, namely: (1) European road safety policy seeks to vastly reduce road deaths using computational surveillance; (2) there is a growing interest around in cabin safety solutions that sense emotion and affective states of drivers and passengers; and (3) autonomous driving trends are changing the nature of interactions between vehicle and driver. Safety led applications are of special interest because they are backed by policy and standards initiatives including the European Union's Vision Zero policy and the industry led New Car Assessment Programme (NCAP). Informed by 13 interviews with experts working in and around in-cabin sensing technologies, this paper first identifies and explores features of emergent in-cabin profiling through emotional artificial intelligence (AI) and biometric measures. It then examines how in-car sensing should be regulated by analysing data protection laws and the proposed EU AI Act. A deep ambivalence emerged from our participants around the emergence of emotional AI in cars, and how best to regulate these technologies.

ARTICLE HISTORY
Received 11 June 2021
Accepted 18 November 2021

KEYWORDS
cars; biometrics; safety

Introduction

This paper considers the implications of car driver-monitoring systems that measure bodies to infer and react to experiential and affective states. The impetus behind the paper is three-fold, namely: (1) European road safety policy that seeks to vastly reduce road deaths using computational surveillance; (2) interest in the role of safety solutions based on in-cabin sensing of emotion and affective states of drivers and passengers and (3) wider autonomous driving trends that are changing the nature of interactions between vehicle and driver. Through interviews with experts in mobility, human-factors engineering, smart cities, data protection, insurance, privacy, biometrics and emotional

CONTACT Lachlan Urquhart lachlan.urquhart@ed.ac.uk

artificial intelligence (AI), we explore the merits and problems raised by cars that pertain to track affective states and emotions. We then discuss the issues raised by expert interviewees in relation to applicable regulations for the car industry and the European Commission's Proposed Artificial Intelligence Act (European Commission 2021). These regulations address AI systems placed on the European Union (EU) market that affect people located in the EU and, importantly, are the first proposals in the world to directly address the regulation of emotion recognition technologies.

The 13 in-depth interviews (Miller and Crabtree 1992) were designed to uncover technical, industrial, legal and ethical insights regarding in-cabin sensing. With informed consent and approval from the Bangor University board of ethics, the interviews were conducted between March 2020 and July 2020, with an average interview length of an hour. Interviewees were happy to be identified and speak of their own views, but noted that these views should not be taken as representative of their organisations. Table 1 depicts the range of experts consulted regarding developments and implications of in-cabin sensing.

Analysis of the 13 interview transcripts followed an adaptive approach (Layder 1998) balancing deductive with inductive insights from data. With the authors already having considerable experience in critical approaches to emotion recognition technologies and relevant governance, a number of critical concerns were pre-empted regarding system accuracy, discrimination in training data, privacy, anonymisation processes and suppositions baked into systems regarding the nature of emotions (McStay and Urquhart 2019). Yet, the mobility sector was relatively new to us, so we were keenly attentive to inductive elements. Using thematic analysis (Cresswell 1994; Miles, Michael, and Johnny 2014) and Nvivo, the first attempt at coding and theme development was undertaken by Urquhart, then separately by McStay, who both then debated and agreed the following five key themes of safety, data, context, trust and governance. These themes comprise:

- Theme 1: *Safety in the city* (which was supported by the discussion of positive uses of autonomous human-state sensing, in-cabin nudging and naturalising human–car interaction).
- Theme 2: *Data flow* (which was supported by interest in identifiability, scope creep, privacy and interaction via data with other cars and systems).

Table 1. Interviewees asked about in-cabin sensing

Name	Organisation
Angelo Ferraro	Expert in electrical engineering (Academic) and P7014 developer
Ann Cavoukian	Privacy and smart cities expert
Ben Bland	IEEE P7014 Chair (IEEE 2019)
Duncan Minty	Insurance ethics expert and consultant
Gawain Morrison	Founder of Sensum
Anon	EU policy-maker
Gary Burnett	Expert in automobility (Academic)
Gilad Rosner	Expert in privacy and networked identity management
Karen Bennet	AI systems developer, consultant and P7014 developer
Ken Bell	P7014 developer
Olivier Janin	CEO of Emotional AI firm
Randy Soper	P7014 developer
Seth Grimes	Emotional AI expert

- Theme 3: Concern with *context* (which was supported by understanding the limitations of emotional AI sensing, and discussion of driver and passenger experience).
- Theme 4: *Trust* as a key social factor hindering adoption (which was supported by discussion of concern about data accuracy, industrial secretiveness of how systems work and safety concerns).
- Theme 5: *Governance* principles and technical interests (which was supported by diverse views on belated regulation of emotion profiling and AI, perspectives on beneficial coexistence between technology and society, ethics-washing, the scope for discrimination in sensing and processing and liability, perspectives on corporate moral responsibility and exploration of how to connect theorisation of ethics with design practices).

Profiling interiors

In the name of safety and enhanced occupant experience, cars are increasingly featuring cameras and sensors that point inwards, as well as outwards. Inward sensors seek to detect specific *states* such as fatigue, drowsiness, intoxication and stress; affective states (such as excitement and relaxation) and expressions of emotions (such as fear, anger, joy, sad, contempt, disgust and surprise). Given that popular discussion of 'smart' driving has been framed in terms of six levels of autonomous driving, ranging from no support for drivers up to the ability for a vehicle to itself drive in all conditions (SAE 2018), the premise of in-cabin systems for human drivers may seem counter-intuitive compared to popular coverage of emerging car markets. We explore this apparent contradiction further below, but for now want to highlight that it is driven in part by practicalities of safety legislation, standards and desire for in-cabin personalisation. This has led Ford, Porsche, Audi, Hyundai, Toyota, Volkswagen and Jaguar (amongst many others) to deploy affect and emotion tracking systems to: (1) assist with safety and (2) profile the emotional behaviour of drivers to personalise in-cabin experience. These legacy car makers are being joined by numerous start-ups, each promising to create value from data about in-cabin occupants. The facial coding company, Affectiva, for example, teamed with Nuance (conversational AI and voices) 'to deliver the industry's first interactive automotive assistant that understands drivers' and passengers' complex cognitive and emotional states from face and voice and adapts behaviour accordingly' (Automotive World 2018). Affectiva again, in 2019, teamed with the mobility company Aptiv 'to unobtrusively identify, in real time, complex cognitive states of vehicle occupants' (Aptiv 2019), and then in 2021 Affectiva was acquired by Smart Eye, a company that focuses on driver monitoring and interior sensing. This acquisition itself is significant for the emotional AI industry that has long sought a stable revenue stream beyond media and advertising research. The car sector is seen as a realistic and lucrative market, albeit a competitive one also including firms such as BeyondVerbal, B-Secur, Cerence, eyeSight, Eyeris, Guardian Optical, Vayyar Imaging, Seeing Machines, Sensay and Xperi.

Those looking to deploy in-cabin sensing are mainly firms classed as either Original Equipment Manufacturers (OEMs), Tier 1 (direct suppliers of components) or Tier 2 (indirect suppliers of services and equipment). They face trade-offs: for example, a driver-facing camera has limitations, but it is relatively inexpensive, compared to multimodal measures.

Coupled with computer vision, Affectiva (2018) shows how an in-car camera can track for occupancy in seats, distraction of driver, liveness and emotion state, presence of child seats and drowsiness through sensing of the presence of faces and 'key body points', smartphone use and micro-expressions. Other measures to sense in-cabin could include heart rate variability, respiration, infrared, motion detection, voice and touch-sensors on sites such as the driving wheel and seats (including dedicated child seats). For developers, the cost is a leading factor, and whilst sensor costs are falling, there remain costs and complexity of installation and lifecycles of component systems to consider. This makes the car sector a mature example of the rollout of affect-sensitive systems, yet also one at odds with growing governance concerns in European policymaking about the role of affect and human-state profiling in everyday life, a theme we will return to in the latter stages of the paper.

Part I

Theme 1: safety, computational super-egos, mobility and the city

Crystalising Theme 1 of safety through in-cabin sensing of drivers, Ben Bland (IEEE P7014 Chair) points out that the 'moral pressure of saving lives strikes us as a pretty strong driver for market change, compared to a vague aim to just make the experience of gaming or whatever better'. This is echoed by Seth Grimes (a consultant and technologist with expertise in sentiment and text analytics), who also served as a city council member of his home city which is also relevant for considering wider smart city dimensions of smart mobility. He sees scope for pro-social use of data about people's feelings so as to be responsive to civic needs. Yet, in cars, he is highly sceptical, remarking that,

> You're going to detect when someone is angry or drowsy in order to prevent road rage or someone falling asleep at the wheel? Frankly, I think this is dumb. I think it is ridiculous. It's a small high-end segment of the market they're targeting, and it is a solution that is trying to discover a market.

Ken Bell (a consultant and IEEE P7014 standards developer) sets out a more optimistic view, highlighting pro-social interactionist possibilities for technologies that respond to physiological behaviour. In cars, this includes factors such as anger and scope for systems to try to regulate and calm driver states. Whilst recognising hype around these possibilities, Bell maintains there is an inevitability in systems that attempt to sense, label and process data about physiology and behaviour. Citing the late Clifford Nass, an expert on communication studies and human–machine interactions, Bell points to the process by which people tend to anthropomorphise when given appropriate cues, thus enabling interactionist possibilities with systems that passively gauge and judge human behaviour. Randy Soper (data scientist, ex-military intelligence and P7014 standards developer), similarly sees a general potential for emotional AI systems to track and coach people to make better decisions (such as in emotively led work-oriented behaviour or decisions made when driving a car). This leads Soper to speculate empathetic systems are a sort of 'conscience' (or perhaps super-ego), watching the body to mediate where necessary.

Nudging, sludging and social contract

Ken Bell offers a useful early binary definition approach to ethics and use of in-cabin AI systems: is it being employed for the benefit of the person being sensed and affected (nudging), or the system trying to manipulate a person for the system owner's ends (sludging)? Another challenge identified by Gawain Morrison is that of paternalism and freedom in relation to manufacturer branding. What if some cars are equipped with in-cabin sensing, and others not? Morrison says:

> If we've all bought into the social contract that that's an okay thing for it to do, well then that's what it does, we've all bought into that. If we have not, and the car still does it, the brand is likely to be the one that suffers, not anybody else. Because the brand's going to be, "Oh, we think it should pull over," and people stop buying that car because it takes control, takes your freedom away'.

Given the longstanding connection between mobility and freedom, Morrison points out that this is requiring brand repositioning from the semiotics of freedom to 'lifestyle branding' (such as increased safety and hyper-personalisation) adding that 'Some are just really having serious difficulties with it'. In addition to branding difficulties, and technical questions of fatigue, sleepiness and emotion profiling, Gary Burnett (expert in human-factors engineering and transport systems) highlights a safety and controls issue in relation to Levels 2 and 3 of autonomous driving. Describing modern cars as 'computers on wheels', Burnett says, 'Level 3 is the really controversial one because that's where you could theoretically do other activities when the vehicle is driving itself, but if there's a problem, the vehicle will chuck control back to you'. Consequently, even for Level 3 systems that are popularly associated with full autonomous driving, there is need for a reliable driver-monitoring system. Further, in the name of safety, such systems would not let drivers take back control if they're too emotionally aroused or at risk of being panicked by having vehicle control passed back to them, potentially going into minimum risk mode and pulling over to the side of the road.

Burnett highlights that for cars and emotion, rage and anger are the predominant emotions to have been studied, but also flags that the goal is not simply to profile, but to encourage positive driver emotions. Yet, in Burnett's account, even this required caution in that systems and options that may nudge to a more positive emotional state also run risk of distracting drivers. Consequently, although this paper is focusing on sensing and human-led driving, in-cabin sensing is present in all emergent understandings of automobility. For Burnett, affect and emotion sensing will be part of the solution of how to re-engage drivers with an appropriate mindset to negotiate the vehicle and environment. This is echoed by Karen Bennet (data scientist and systems developer), who has worked with North American car manufacturers, autonomous vehicles developers and mobility services including Uber. She also highlights the role of in-cabin and driver-monitoring measures when taking control away from the driver, or in the self-driving car situation when a driver might need to take over vehicle control because 'AI is getting confused or does not have appropriate levels of accuracy that we set up in the software'. Bennet also remarks on usability studies, seeing that 'emotion data are going to take off very quickly, because that's a set of data that makes it a better experience, a user interface for humans'. This is a subtle point in that emotion itself becomes the

medium, a means of affecting objects and systems. Asked about the nature of interest by the large technology corporates, Bennet says,

> I've had the luxury of actually working with them, and they saw, a long time ago, that it was a natural extension to go into AI and machine learning up and over the Cloud. It was how they were going to gain more money, and now they can see having emotion as the next step in expanding their Cloud services.

Theme 2: data flow

With Theme 2 on *data flow*, we refer to the movement of data between computers on a network. As outlined, a key claim by industrialists is that data collected from in-cabin profiling stay local in that personal data are not transferred beyond the device it was originally collected on. Many interviewees were sceptical, with Ann Cavoukian (privacy expert and ex-board member for Toronto's Sidewalk Labs project) noting her longstanding experience with autonomous vehicles development, stating that, 'The big problem is they're connected, not just in terms of the technology required to drive the vehicles, but all kinds of personal information relating to the drivers, the owners, and that this will happen without consent of the drivers, the data subjects'. Ann Cavoukian likens information about emotion with health. One consequence of this, beyond the scope of this paper, is the overlapping of data with bio-ethics. Cavoukian states:

> I put [emotional AI] in that cadre of health information because it's your personal perceptions, your emotions, and this should have the strongest protection possible and none of it should be conveyed or disclosed without your positive consent, like all other personal information and health-related data.

Beyond bio-ethics, for a legal context, this observation would classify data about emotion and human states as special category data (in General Data Protection Regulation 2016 terms), strengthening governance requirements around its use to ensure lawful processing (ICO 2021). Changes in transportation and shifts towards smart mobility are often part of larger shifts to create smarter cities, where data play a key role in management of infrastructure, movement and everyday life (Kitchin 2014). For Cavoukian, de-identified data are key to an ethical smart city, where value can be gained but the risks to individuals are arguably minimised. In relation to Sidewalk Labs, she recounts,

> So I said, "You need to de-identify all data at source, meaning that the minute it's collected, you need to scrub all the personal identifiers, remove them. That way you'll still have a lot of data, but it won't be privacy invasive, because there won't be personal identifiers linked to it".

The lessons here for limiting data flow across complex supply chains would also be relevant for the automotive sector, where minimising scope to identify may align with the secretiveness in the sector, a point we return to later.

Identity: measuring the meat-bag in the seat

Car usage is changing, with one key development, particularly in urban areas, being a reduction in personal vehicle ownership and rise in on-demand hire or ridesharing schemes (Enoch 2018; Department for Transport 2020). Identity in relation to in-cabin sensing has two overlapping modes: (1) identity through authentication, potentially to

access the vehicle and its systems and (2) knowing who the individual is to profile behavioural attributes to. Gilad Rosner (a privacy and identity management expert) flags that this reduction in ownership raises questions about authentication, data mobility, identity, selective sharing and who has permission to collect and access data in relation to specified situations. As to whether cars will seek to identify 'the meat-bag in the seat', Rosner does not know, but posits that it is in car manufacturers' interests to be able to claim that driver and collected data are not linkable. On ownership and ridesharing, Gawain Morrison (CEO of Sensum) also sees changes from personal car ownership to subscription based or Uber-like public mobility. Driven by safety legislation, for Morrison there is an inevitability towards in-cabin sensing, and this will tie into in-cabin retail as well as safety. He says, 'The best people who will do it with the correct opt-in requirements. "I don't want this, don't sell it to me, thank you very much." Maybe it puts a 10% premium on your ride'. This links to issues of scope creep once sensing infrastructure is available, a point we return to below.

Industry self-justification: 'sensing not storage'

Ben Bland points out that the car industry is struggling with technology cycles because it prefers to lock a suite of technologies in for a given period (Bland notionally suggests 5 years), but also that 'They're trying to speed that whole thing up, recognising at the same time that it puts them at risk'. The other existential threat is recall of a product/service/entire vehicle, potentially caused 'by mistakes coming from third-party providers who have sold you a camera or whatever'. Yet, the pressure legacy firms face is pressure from new market entrants, from dedicated firms such as Tesla, but also non-automotive companies such as Amazon and Google that have expertise focused on computation and user experience design. Bland argues, 'I think the biggest risk, I mean obviously it's doing something that endangers lives and they want to avoid that, but more sort of commercially speaking, recall is like the biggest fear in the industry as far I understand'.

The sensitivity of privacy concerns around in-cabin sensing is keenly felt by industry and governance actors looking to deploy these human-state analytics. Sensitivity became apparent in a series of webinars (Affectiva 2020a; 2020b; 2021) hosted by the emotional AI firm, Affectiva, featuring a range of industry experts, representatives from New Car Assessment Programme (NCAP), and academics working in transport and mobility. Echoing EU legislation discussed below, Richard Schram of European NCAP (Euro NCAP) was keen to assure webinar delegates who asked questions about privacy that the car is sensing rather than storing data. Yet, it is also worth noting that whilst sensing data may not be stored, it may inform other systems. For example, if a system senses the beginning of fatigue, it might suggest upbeat music. The biometric data (e.g. face and heart rate) may be discarded instantly but the information (that a person may have become tired) lives on by other means (a recorded uptick from a music recommendation or cooling of the in-cabin temperature).

We observe that there is clear scope for the contradiction between claims of 'sensing not storage' versus the strict EU rules on data processing, especially if consent is not going to relied upon as the lawful basis by which to process personal data. This in theory might be quite often, such as when borrowing or leasing a car, or buying a used car. On this point, the EU EDPB (2020) is clear: consent requirements should not be bundled with

the contract to buy or lease a new car. Further, given that in-cabin sensing is meant to be ambient and operate under the radar of conscious awareness, this creates further difficulties in people being aware of data they are communicating in the first place.

One outcome of the de-identification processes, raised by Ann Cavoukian, is scope for third-party inferential analytics that operate just outside of the General Data Protection Regulation 2016 (GDPR). On car insurance, Duncan Minty explains that this overlaps with the longstanding practice of affinity underwriting, 'in that you will get, say, an affinity group of 10,000 people underwritten as a group. So that logic is actually firmly embedded in insurance as a way of doing business with chunks of individuals'. This has two implications: (1) on sector level care for aggregated data and (2) on acceptance of inaccuracy. Minty says:

> I don't think the sector has been that strong on caring about how they handle anonymised data, and also feeling "actually we can read you from that data and make decisions from that. And as long as our portfolio management is good enough (that's where actuaries come in then) then we don't care if we're a little bit inaccurate." So yes, I've certainly seen synthetic data being used by insurers, and this was one who was doing a lot of good things in relation to their data.

For Minty, this raises the question of how usable insights may be achieved through synthetic and inferential analytics (properties of populations), remarking 'Look at those people over there, they did this, and you seem to be similar to them, so I'm going to assume you're going to do the same'. The role of inferential analytics is also raised by Randy Soper (data scientist, ex-military intelligence and P7014 standards developer) in that emotional behaviour may be gauged without sensing the body, potentially through real-time vehicle telemetry tracking and GPS location data to provide context to vehicle behaviour. Gawain Morrison amplified this point in relation to fatigue, noting external-pointing sensors that track white lines of the road and telematics for acceleration and deceleration patterns. For example, high-engine revving in an area without hills or high traffic might be claimed to indicate 'aggressive' driving. This allows an organisation to gauge emotional states without the use of biometrics, yet also make similar judgements about the driver as if data were being collected about blood flow and heart rate.

Scope creep

Despite policy insistence on privacy-friendliness and industry claims of 'sensing-not-storage', Gary Burnett sees clear scope for data to be shared with manufacturers, potentially for improving and optimising products, yet also for marketing activities by first-party OEMs. Citing both driver-facing cameras and natural language interfaces in cars, he recounts discussions regarding the value of this data, not only for the data that comes from these different types of sensors, but also the broader inferences about drivers and disposition. Gawain Morrison also highlights the scale of investment that car manufacturers need to see returns on, not only for upfront costs, but also ongoing storage and processing of data being collected from cars. Consequently, there is pressure to generate revenue from in-cabin sensing.

All interviewees with direct knowledge of the car industry flagged high levels of company secrecy and desire to withhold data for themselves. Burnett observes that manufacturer self-interest will likely play a positive role for privacy because manufacturers do

not like the premise of connectivity with parties that they are not part of (such as road infrastructure, other businesses, councils and other vehicles). Nevertheless, Burnett thinks they *do* like the idea of sharing the information with themselves, for marketing, product development and for understanding their customers. Yet, third-parties, such as insurance companies, are interested. Insurance ethics specialist Duncan Minty grounds the technical discussion in how third-party usage of data might work in practice, noting that 'there has long been a relationship between the insurance industry and the motor industry through group schemes', meaning that 'if you buy a (specific brand) of car, you will get access to an insurance scheme set up by an insurer just for those type of people, be it new cars or second-hand cars'. It should not be missed either that there is a longstanding connections between insurance and telemetry, especially for young drivers. In relation to data, Minty adds that 'over the last few years, certainly within Europe, there has been a campaign by insurers to at stop vehicle manufacturers feeling that they own that data'. This is significant in that the insurance industry is likely to lobby EU policy-makers and that there is a commercial incentive for all in the in-cabin sensing production chain (OEMs, Tier 1 and Tier 2 firms) to find a way to monetise this data. Minty sees a further issue in that it is 'quite straightforward for motor manufacturers to use their own in-house insurance company, which most of them will have, to start underwriting it themselves'. An implication of this is clear opportunity for manufacturers to move into insurance due to increased datafied proximity to drivers and consequent intimacy with subjective and physiological states. Further, on scope creep, especially when seen in terms of in-cabin sensing that is not solely about safety (but also adaptive in-cabin environments, comfort, entertainment and marketing to occupants), one begins to see the emergence of the automotive sectors as a data industry. For Minty, such profiling is inextricable from concern about location and other areas of insurance practice, such as health. He observes: 'So, if they're driving to the gym or pharmacy a lot, health underwriters will be moving underwriting dials to say, "It's a life and health implication here"'.

Asked about controversial usages of in-cabin data, Gary Burnett queries usage in law courts, such as in relation to road rage incidents. Yet, in reference to the use of driver-facing cameras, especially when Levels 3 and 4 autonomous vehicles begin to crash, he identifies future questions about whether the human-in-the-loop driver was asleep, or that they did not respond to a 'resume control' request, e.g. when the car was in a dangerous situation requiring human oversight. This overlaps with earlier sections on safety in that whilst extension of scope into legal question is a creeping of sorts, it is also a defence for drivers. For example, the manufacturer of a Level 3 or 4 autonomous vehicles also has a responsibility to the driver occupants and people and objects in proximity of the vehicle, creating complex consumer protection and liability questions. Gawain Morrison raises a similar point, speculating on the tired driver that chose to keep driving, despite having been warned by the car, and who consequently caused a crash. Morrison sees that this information will be available, raising difficult privacy and safety questions.

Theme 3: context and accuracy [A]

This refers to the ability of systems to make sense of behaviour and emotion expressions in relation to a person and the situation from which computational inferences are being

derived. Against the context of critiques of coded bias (Buolamwini and Gebru 2018), systemic lack of sensitivity to people with protected characteristics (notably here people with disabilities, as well as cameras that do not see darker skin tones well), and instability of the psychological models of emotion that underpin emotional AI, we were interested in practitioner responses to these critical matters. Randy Soper opens the topic well by pointing out the role of cost in relation to social equality. He says,

> So, the first problem with cost is if you want that parity, you've got to design the dataset yourself. A lot of times, it goes against the whole point of big data AI, which is to reuse somebody else's work to save yourself some cost.

On the quality of training data, Gawain Morrison highlights that even large companies will struggle to create their own reliable training data (mentioning Google and Alibaba as potential exceptions), which means, 'I think you do need to have a process of evaluation to be able to do some A/B comparisons with datasets that you own and trust, where you can then cross-references on A/B different capabilities'. Yet even that will only prepare systems for lab-based conditions, leading Morrison to recommend a secondary wave of testing when driver-monitoring systems are deployed in the real world.

On training data for emotion labelling in supervised machine learning, Bennet recognises disproportionate social representation in emotion training data, recounting an alarming gender-based case:

> Sadly, I was part of a review team on one of those situations, they brought in some people to look at the data and what had happened and how it had been trained, and it had not been trained with enough female data, that it didn't recognise that it was a female walking. And so, in the moment, the software thought it was just an object, and so it chose to hit it, not realising it was a human.

On emotion (especially facial recognition coding of expressions), Bennet observes that accuracy is low, and the pressure from groups such as AI now to ban it, also saying that 'In my heart of hearts, I would love to see it banned until it meets a certain level', but adding 'unfortunately, technology will keep moving forwards. So, I'm not a supporter of banning, I'm a supporter of "we need to start having measurements to know what is good and what is bad out there"'. On the question of measurement and accuracy, Seth Grimes introduced a valuable distinction between absolute values and change in measures over time. He explains:

> We need to understand the tolerances, the accuracy, and we also need to think about how to best use the data that's produced perhaps by looking at changes rather than just absolute values.

This begins to add scope for local variation, especially when gauged over an extended period. With the Emotional AI Lab, that McStay and Urquhart participate in, we have been studying cross-cultural questions regarding emotion sensing (for UK and Japan), and we were interested in internal developer debates about the socially contingent nature of emotion labelling. Bennet comments that in her experience these conversations are not taking place. Responding to questions about how internationally diverse manufacturer research and testing is, Gary Burnett observes that this is driven in part by market conditions and where a car company is based. He points out, 'General Motors in Detroit, or for BMW and Audi in Munich, or Jaguar Land Rover in the UK, that's

where most things will happen, but they will have offices elsewhere, but it just won't have the resources to be able to do huge amounts. Cultural diversity in human-factors engineering for cars is a recognised problem, especially with China being such a large (and poorly understood) market for Western car manufacturers (Braun et al. 2020). This raises concerns about homogenous application of emotion profiling, versus how systems can account for socially heterogenous experiences and articulations of emotional life. Industry interviewees were mostly alert to criticism of facial coding of expressions (Barrett et al. 2019), although Morrison argued that greater accuracy *is* possible. He states:

> It's been a good decade and more of work on the marriage of psychology, biology and technology to get us to a point of looking at single or multiple data streams, being able to understand different human states and classify human states and emotions from single and multiple data streams.

In relation to cars, he posits this as paramount as, 'It's not just about analysing historical data sets for research or interactive campaigns, which aren't life or death, but we're moving into products that now actually have to properly live beside us and interact with us'.

Privacy expert Ann Cavoukian is careful to separate data protection concerns, arguing that accuracy is a pre-condition for crossing privacy. She highlights that,

> It actually goes, of course, beyond accuracy, but we have to start at accuracy. That's why I always want to start there, because if the conclusion is that this information is not accurate, then there's no point in delving beyond that in terms of intrusions of privacy, which are obviously very strong.

Cavoukian's insights are especially relevant in that whilst she is well known for her role and criticism of Alphabet's Sidewalk Labs and her background as a psychologist. In reference to race and discrimination, she states:

> You can't do an accurate read of emotions. Even people like psychologists, that's my background, who are skilled in this area often incorrectly reach conclusions associated with emotion. So now you're saying you're entrusting the police to make the correct conclusions with respect to the emotional cues they're getting? Forget it. It's going to be a disaster.

For Duncan Minty, an expert in data analytics and insurance, he recognises interest from the insurance industry in emotional AI, because new technology allows them to 'dominate the market or to take control of the market'. For Minty, the concerns are two-fold: (1) the premise of actuarial fairness, that insurance and risk are pooled, and that insurance does become done on a per-person basis; (2) that questions of emotion and human-state profiling are fraught with potential errors and poor decisions. Yet, despite concerns, Minty also observes that accuracy might not be a terminal problem. He distinguishes between accuracy requirements at underwriting and claim stages. For the former, accuracy is not paramount and 'You can be reasonably accurate and that's fine', although 'you want to be accurate enough not to have a poorly performing portfolio'. At a claim stage, however, greater accuracy is needed because a claim is based on the premise of a contract, which means that 'the level of quality has to increase if you're going to, for example, reject a claim based upon emotional data, which has been interpreted in a certain way'. Minty adds that an additional factor to accuracy and actuarial fairness is the extra need for the insurance industry to conduct due diligence, in that 'the outcomes

experienced by customers are fair and accurate, honesty and so on'. Here one could easily see a split between arguably more reliable affective states such as fatigue and sleepiness, versus emotion and driver intention. On motive for automated emotion assessment, Minty observes that insurance firms are 'entranced by the sheer size of the AI bubble' and that there is constant pressure to reduce costs. The concern for Minty is less about statistical ability of the insurance industry, but that the topic of emotion is a human sciences topic, something that the insurance industry has less expertise in.

Theme 4: trust

Theme 4 is about *trust*, broadly conceived in reference to sociological understanding that sees trust in modern life as involving double-edged characteristics in that complex technology infrastructures may offer greater comfort, yet they come with social, physiological and environmental risks (Luhmann 1979; Lupton 1999). All interviewees, regardless of background, recognised that use of data about emotion is an ethically debatable topic, with Emotional AI firm CEO, Olivier Janin stating, 'We need to allow the dignification of this future business and to [be] very, very careful of how it is used by and by who'. Janin alludes to emotional AI as a sector that is not yet fully dignified, and one where trust and reputation is paramount if it is to grow in coming years. The comment echoes wider interviewing by McStay (2018) that found start-ups having to manage the reputation of the technologies not only amongst the public, but potential large clients too. This applies well to automobile manufacturers, especially given the role of branding, trust and the long cycle of new product development in the sector.

Hindering trust: secretiveness

This paper was originally motivated by an interest in supply chains of data about emotion and affective states, but in the course of interviewing and attending industry webinars it quickly became apparent that this is a highly secretive sector, in large part due to expensive development stages and need to guard advances to enable competitive advantages. Indeed, Sensum's Gawain Morrison goes as far as calling it a 'paranoid industry' that 'does not trust itself' because 'it's a very small tune of humans across a small number of companies that make an awful lot of money'. Gary Burnett observes that willingness to be open about training data for in-cabin sensing will vary between manufacturers, depending on the politics of their organisation and the relationships they have with suppliers. He also adds that, 'I think they will do quite a bit of the sort of development work themselves, but then the car companies like to see their own data'. This takes on extra significance given the ambiguity of what emotions are, the lack of academic or industry-agreed measures on expressions said to indicate an emotion, and that there is set of standards for manufactures or third-party providers to take recourse to.

Grey zones: trust and business culture

In interviewing, we were interested in the normative space between egregious abuse and harmless uses of technology, where judgements are not straightforward. Angelo Ferraro (Expert in Electrical Engineering (Academic) and IEEE P7014 standards developer)

addresses this by pointing to post-development rationalisation (an antithesis of 'ethics-by-design'), remarking, 'it's rationalised by saying the net good is better than the harm it causes', adding that 'How can you ask someone at a start-up company, a small company that worked hard, that is starting to see some success […] and their livelihoods, their entire careers, are based on developing this company, to make these grey zone judgements?' The developer operations side is referred to again by Karen Bennet who, in reference to data about emotion and her time associated with multinational firms working in the sector, says, 'what I see is the engineers who I work with, or I manage, are constantly asking me, "Ethically, is this correct? Should we be doing this?"' Although recognising the role of law, meta-principles and human rights she adds, 'when it gets down to actually tweaking an algorithm, they don't really help you on the development side', urging for technical and operational advice'. She sees this in part as an intra-organisational issue in that executive-level management buy-in is required, but also that high-level goals, be these ethical and/or legal, need grounding in a common language all can understand work with. Further, she adds that because of the lack of clarity on ethics (and arguably law) on use of data about emotions, she notes there is greater willingness by executives to listen to all parts of an organisation regarding the merits (or not) of using data about emotion. The observation that technologies are ahead of regulation and that restrictions are afterthoughts was a recurrent one, with Seth Grimes adding that they are 'often imposed from outside by people who don't really learn about the impact and importance of these technologies for several years', a point that we argue lends well to systems that make judgements about emotion.

Theme 5: governance

We focus here on systems of accountability and how rules regard the deployment and management of in-cabin sensing. All interviewees recognised the role of governance, some to restrict and control the deployment of emotional AI in cars, others to enable it through good standards and social trust. Angelo Ferraro has professional developer expertise in AI, reaching back to the 1970s. From this perspective, he spoke of missed governance opportunities to address modern harms posed by AI. He remarks:

> Sometimes I think that Pandora's box has been opened, and we let the bad out a little bit too soon, because back in the '70–80s', when it was truly a nascent technology, we had the opportunity to actually put forth some good workable standards, but it was so far-fetched at that point to do even simple tasks using AI.

With an understandable lack of appreciation for a future that had not come to pass, Ferraro reflected on absence of commercial dominance of AI in the 1970s and 1980s, adding that. 'Here we were, trying to teach a baby on how to feed itself, and the thought of having to come up with elaborate standards, it was just so far out in the future, it never took place'. An interviewee from the European Commission, charged with modern governance responsibility, and who worked directly on the GDPR, flagged the role of the EU High-level expert group on AI in navigating harms and benefits from emergent decisional technologies involving emotion and affects. Whilst seeing risks, they also cautiously observed that emotion and affect measures may have scope for social good. Their general perspective was also noteworthy, articulating a 'synergetic

approach' to technology and society, speaking of co-evolution and a 'beneficial coexistence'. This co-shaping worldview had recognisable policy origins in the social shaping of technology theses (Williams and Edge 1996) that rejects determinism, yet recognises that technologies do impact on society.

Part II

Vision Zero, trust and the proposed European Commission AI Act

To recap, industry-oriented interviewees saw the potential for in-cabin human-state measures to deliver safer cars; yet all recognised concerns about data and privacy, coded bias, ethnocentric questions about the constitution of emotion, trust (and lack thereof) and the difficulties in how to govern the deployment of emotion and affect profiling in cars. Broadly, we see some positive motivations, but also training data and hardware opacity problems, questionable 'sensing-not-storage' claims, industrial secrecy, scope for harms to users and normative grey zones about appropriate ethical and legal standards in this emerging area. With legal clarity needed, this section turns to the governance and legal frameworks that will contribute to the emergence of in-cabin sensing. We provide novel analysis of the recently proposed European Commission's AI Act (European Commission 2021), where we: (1) argue that emotional AI in cars used for safety critical functions should be deemed a High-Risk AI System (HRAIS); (2) reflect on the implications of this classification for design and operation of these systems; (3) consider the proposed transparency led approach to regulation of emotional AI in the current proposal.

In Europe, emerging AI ethical policy and regulations are seen as a way of instilling trust in AI (High-Level Expert Group on AI, 2019; European Commission 2021). The belief is that the region's citizens may enjoy benefits of systems that judge and make decisions for them, without the harm. This paper, focused on cars and in-cabin sensing, finds this to be a problematic premise. As found at the interview stage, of all potential use-cases for profiling of emotion and psycho-physiological states, in-cabin systems are unusual in that they are receiving policy backing. Indeed, as of mid-2022 in the EU, all new cars put on the EU market will have to be equipped with advanced safety systems due to the implementation of the EU Vehicle Safety Regulation 2019/2144 (European Parliament and Council 2019). This entails numerous factors, such as in-cabin alcohol testing (where a sample of alcohol-free breath is required before the vehicle will turn on), driver drowsiness and attention warning systems, advanced driver distraction warning systems and event data recorders. The Vehicle Safety Regulation seeks to limit the use of collected data (in line with existing EU GDPR obligations) and includes measures in Article 6 around limiting third-party access to data, restricting repurposing of data and requiring immediate deletion of data after processing.

Vision Zero

Regulation and governance are important to understand the industrial interest outlined above in emotional AI in cars, because this is being driven in large part by policy and standards initiatives. In Europe, for example, the EU's Vision Zero initiative is based on the Swedish approach to road safety thinking where no loss of life is acceptable. It recognises

that people make mistakes, and that traffic needs to flow, so urges mitigation measures to reach Vision Zero (TRIMIS 2021). The working document to reach Vision Zero by 2050 is the EU Road Safety Policy Framework 2021–2030, which recognises that reducing road death is a multifaceted problem and technological solutions are prioritised 'foremost in connectivity and automation' (European Commission 2020, 7) to reduce the role of human errors by 'taking the physics of human vulnerability into account' (European Commission 2020, 11). Reminiscent of the 'synergetic approach' noted in the European Commission interview above, the EU Road Safety Policy Framework 2021–2030 also encourages fitting of 'state-of-the-art advanced safety technologies' in reference to the Euro NCAP.

Originating in the United States, the NCAP provides consumers with information regarding the safety of passenger vehicles, publishes safety reports on new cars and awards 'star ratings' based on the performance of the vehicles. Its goal is to improve vehicle safety standards beyond those found in European law. This is in relation to all likely collisions a car and its passengers (of varying ages and vulnerability) might be involved in. In addition to collision standards, Euro NCAP Advanced was set up as a reward system launched in 2010 for advanced safety technologies. Since 2020, Euro NCAP requires driver monitoring for five-star vehicle ratings. This is 'to mitigate the very significant problems of driver distraction and impairment through alcohol, fatigue, etc.' (Euro NCAP 2018, 2).

From the point of view of privacy, there are three key aspects to the EU Road Safety Policy Framework 2021–2030 document that are relevant to flag: (1) it states that EU Member States should be able to access in-vehicle data to determine liability in the case of an accident; (2) it recommends consideration regarding the collection of anonymised data about car safety performance; (3) it shows interest in 'more complex human–machine interfaces (HMI)' (European Commission 2020, 18). Factoring for legal liability, the 'sensing-not-storage' privacy argument is diminished. The EU Vehicle Safety Regulation also adds detail on integrating matters of usability, systems architecture and data protection into one legal provision for cars (European Parliament and Council 2019). It legally mandates distraction detection, particularly for warning of driver drowsiness or distraction. Its Recital 10 states that 'any such safety systems should function without the use of any kind of biometric information of drivers or passengers, including facial recognition', which can be read to mean that human-state data that is used should not identify a person, because biometrics are considered as a more protected category of personal data (in GDPR terms). Rather, systems should use edge and fog computing where data are processed on the device locally and discarded, never leaving the car. Recital 14 is explicit, stating that personal data, such as information about the driver's drowsiness and attention or the driver's distraction, should be carried out in accordance with the EU GDPR and that 'event data recorders should operate on a closed-loop system, in which the data stored is overwritten, and which does not allow the vehicle or holder to be identified'. Further, systems should not 'continuously record nor retain any data other than what is necessary in relation to the purposes for which they were collected or otherwise processed within the closed-loop system' (European Parliament and Council 2019). To this end, Art. 6(3) further states that not only should such systems be 'designed in such a way that those systems do not continuously record nor retain any data other than what is necessary in relation to the purposes for which they were collected or

otherwise processed within the closed-loop system', but also 'data shall not be accessible or made available to third parties at any time and shall be immediately deleted after processing'. The significance of both Vision Zero and the EU Vehicle Safety Regulation for emotional AI, and in-cabin technical feeling-into of fatigue and drowsiness, is that new forms of sensing and surveillance are permissible, as long as no data are retained that identifies a person. Assuming this is feasible, the given industrial interest in inferences for in-cabin personalisation identified earlier, attention should be paid to whether traces of affect and emotion inferences remain in non-safety-specific systems.

EU proposed AI Act and emotional AI in cars

Another major governance shift for in-cabin sensing comes from the recent European Commission's Proposed AI Act (European Commission 2021). These set out some drastic changes in how certain AI applications are to be formally regulated in Europe. Article 2 of the AI Act ensures it has relevance beyond Europe, as it seeks to regulate classes of AI systems put on the EU market. Thus, OEM and Tiers 1 and 2 firms from third countries like the US, Japan or UK that are seeking access to EU consumers and business users will need to comply with these requirements. It signals regulatory efforts from the EU in defining how to build a more trustworthy ecosystem around AI innovation, building on ethics focused work from the HLEG on AI (High-Level Expert Group on Artificial Intelligence 2018) and earlier policy shifts indicated in the EU White Paper on AI (European Commission 2020b). It is a comprehensive framework providing a risk-based approach to regulate AI applications and provides principles to guide future governance on access to the market. The AI Act introduces a risk-led tiered approach to different AI applications ranging from: (1) prohibited systems to (2) high risk AI systems; (3) limited risk systems interacting with humans; (4) minimal risk systems. It encourages voluntary codes of conduct for minimal risk AI in Art 69. Most relevant to the automotive sector are high risk systems and limited risk systems using emotional AI, so we focus on those below.

High risk AI systems

We argue that use of emotional AI in cars should be deemed a HRAIS on the basis of two provisions. Firstly, Art 6(1) AI Act states a system is HRAIS if it is a *safety* component of another system (e.g. if that system fails it causes risk to health and safety) and if it requires ex ante third-party conformity assessment (under Art 19 and 43 of the law). With the first condition, our interviewees stated that emotional AI in cabins will include driver-facing cameras for advanced driver distraction warning systems. Importantly, industry interviewees such as Sensum's Gawain Morrison believe that emotion profiling accuracy *is* improving through multimodal measures and that these will be used in situations where death is a possibility. As discussed above, the EU Vehicle Safety Regulation covers driver-facing cameras using AI which enables detection of valence, attentiveness, fatigue and thus these could be seen as a safety critical system, as if they fail, a driver would not be notified of inattentiveness and potentially crash. Furthermore, many of our participants argued that as vehicles become more automated, there will be an increasing role for awareness detection systems when giving back control to a driver.

Thus 'knowing' the state of the occupant in a safety critical situation, before giving control of the vehicle back at high speed further shows the safety criticality of these types of systems. In terms of the second condition of requiring a third-party ex ante conformity assessment, the AI Act Annex II states this is needed in the context of AI systems in motor vehicles.

Secondly, Art 6(2), states that if a standalone AI system is listed on Annex III, then it is also a HRAIS. This list includes: '*Biometric identification and categorisation of natural persons. (a) AI systems intended to be used for the 'real-time' and 'post' remote biometric identification of natural persons*'. Whilst law enforcement use of biometrics in this way is prohibited by Art 5 AI Act, non-law enforcement use remains HRAIS. Biometric data are defined in Art 33(3) AI Act as '*Personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data*'. There are debates around to what extent emotional AI systems use biometric data to single out and thus identify an individual (George, Reutimann, and Tamò-Larrieux 2019). However, as McStay and Urquhart (2019) have argued, the shift towards contextual forms of emotion sensing, where facial action coding is combined with context specific data to determine where emoting is ongoing, it is likely that the mosaic of data available in the cabin context will enable singling out. The car is a sensor rich environment meaning personal data will be implicated by different systems, from pairing of mobile devices to in-car entertainment, to more mundane registration of users for eCall and car service notifications or even registration of vehicle. These are distributed around different ID management services for different purposes ranging from customer satisfaction to safety to taxation. Thus, as our participant Rosner mentioned, automotive manufacturers could create risks for themselves in becoming identity management providers, especially as in-car biometrics emerge, such as galvanic skin response and heart rate measures through steering wheels that could feed into ID management. Furthermore, the European Data Protection Board has recently highlighted data protection risks of coupling of biometric identification and emotion sensing systems (EDPB 2021) and even if personal data are deleted quickly upon collection, to enable privacy preserving processing measures, this limited processing could still be subject to GDPR, raising questions for EAI systems that work in this way (George, Reutimann, and Tamò-Larrieux 2019).

The importance of being a HRAIS is that it requires AI systems to adhere to a set of design requirements, documented in Title 3/Chapter 2 of the AI Act. It also sets requirements for different actors in the AI supply chain to adhere to, a point returned to below. Non-compliance with the data governance provisions results in significant sanctions, for example, non-compliance with Art 10 AI Act can attract fines up to 6% of the previous year's turnover or €30M and for breach of other provisions, Art 71 AI Act states it can still be 4% of turnover or €20m can be levied. Articles 9–15 cover the range of system design requirements that need to be adhered to for HRAIS. These range from: establishing and maintaining a *risk management system* to identify risks, possible misuse, mitigation and testing measures (Art 9); ensuring there is *record keeping* and *automated logs* that can be accessed by authorities when needed (Art 12); drafting of *technical documentation* before putting a system on the market (Art 11) and ensuring there are measures to achieve *robustness, accuracy* and *cybersecurity* across the AI lifecycle (Art 15). We will

focus on Articles 10, 13 and 14 here, as these have the most relevance to automotive uses of emotional AI.

Art 10 focuses on data governance by establishing quality criteria around training, validation and testing datasets. The requirements on data governance and management practices in Art 10(2), for example, focus on the required dimensions of compliant datasets. This paper was originally motivated by understanding the nature of the training data supply chain in the automotive sector. Our interviews revealed there was a high level of opacity around how data are sourced and used in training AI systems. This provision puts in place the tools to prevent such opacity being able to continue. It goes down to the level of requiring documenting of design choices, identifying gaps and shortcomings in data, documenting data preparation and sourcing mechanisms including around 'annotation, labelling, cleaning, enrichment and aggregation', alongside reflecting on the formation of assumptions and examining biases in data too. Whilst this is a system requirement, there are questions about how these obligations will materialise in practice between providers and users of HRAIS, especially as users will further train a system in use, beyond initial training from the provider.

Art 10(3) focuses on ensuring datasets have further quality features such as being 'relevant, representative, free of errors, complete' and have specificity to groups that will be subject to HRAIS. These raise concerns around ethnocentric and cultural sensitivity, and appropriateness of training processes, which could perpetuate harmful norms around gender, race or social interactions, which follow discussions from participants above. Art 10(4) highlights these datasets should, for the intended purpose, take into account 'characteristics or elements that are particular to the specific geographical, behavioural or functional setting within which the HRAIS is intended to be used'. This push to understand multiple forms of context is interesting to see in the law but also for application in the automotive sector. Our participants have painted a picture of how, at meta-level, the sector deals with actuarial risk, regional research and development processes and secretiveness between OEMs, Tier 1 and Tier 2 service providers. This also goes down to the context of use where cars will have varying levels of automation, different technological capabilities and sensing functions, and lack of interoperability between proprietary systems which leads to a highly heterogenous sensing environment, and need to address any impacts this has on training data governance.

Art 10(5) and (6) cover a range of technical details but broadly address when it is legitimate to process special category personal data, namely for data bias management and also requires establishment of safeguards such as privacy preserving measures and security. The shift to edge processing highlighted by some participants is interesting insofar as this might support requirements in Art 10(5) and enable privacy preserving measure, such as the use of trusted execution environments on hardware or personal information management systems in vehicles which could enable necessary analysis but limit human and third-party access to this.

Art 13 of the AI Act states that HRAIS should be designed and developed to ensure 'their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately'. We return to questions of transparency below in discussing Art 52, but note that this section unpacks what transparency *means* for HRAIS in Art 13 (3). This includes providing information such as identity and contact details of a provider;

'characteristics, capabilities and limitations' of the HRAIS; human oversight measures; expected lifetime of HRAIS including maintenance and care needed (including software updates); changes to its performance; health and safety risks; its 'intended purpose'; performance in relation to intended subjects; specifications for input data and 'any other relevant information in terms of the training, validation and testing data sets used'. This is comprehensive and to further utility and access to this information, Art 13(2) states it should be in an 'appropriate digital format or otherwise that include concise, complete, correct and clear information that is relevant, accessible and comprehensible to users'. The in-cabin environment poses interesting design challenges for delivering on these provisions, which we discuss in more detail below, but this list provides a useful set of legally informed requirements for user experience experts and ergonomists to consider in their practice.

Art 14 HRAIS 'shall be designed and developed in such a way, including with appropriate human–machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use'. The implication for design and system development is significant from this provision, putting human–computer interaction at the forefront of regulation and mandating not just explainable AI (as many debates in law have focused on in relation to Art 22 GDPR, such as Edwards and Veale (2017)) but instead on *effective* human oversight, which may not always require an 'explanation' per se (Crabtree, Urquhart, and Chen 2019). Concerningly, in the UK context, the government is considering changes to UK data protection law post Brexit to remove Art 22 (DCMS 2021). Nevertheless, as oversight was a concern of participants above, it is worth unpacking from the perspective of what Art 14 of the AI Act might require. It views oversight as supporting the ability to act beyond just obtaining an explanation. The oversight system should support monitoring for anomalies or dysfunctionality of the HRAIS; interrupting or intervening in the operation of the HRAIS via a stop button or deciding not to use the HRAIS or ignore its output. In the context of a safety critical setting like a moving vehicle, meaningful oversight poses tough design and human-factors questions, particularly around the timing of interventions and notifications to those overseeing the system.

Before turning to emotional AI in the regulation, we briefly refer to Title 3 Ch 3 as it defines obligations for providers, users, importers, distributors and third parties around HRAIS. This approach recognises the distributed nature of responsibilities and helps to avoid responsibility gaps or some parties ignoring obligations around use of AI. These differ but include the use of CE marks, quality and conformity assessment processes, drafting of technical documentation, in addition to finding mechanisms for management of risk across the HRAIS supply chain. CE marks are traditionally symbols on products that signify conformity with a legal framework, e.g. for electrical safety of consumer goods and showing them to be safe to be available on the market. According to the AI Act, whilst these can be granted, when circumstances arise that call a system into question, the issues should be fixed or risk seeing the CE mark withdrawn. In the automotive sector with a complex, global supply chain, managing responsibilities across it will require coordination. Yet as mentioned by our interviewees, it is quite a secretive industry, making this a challenge with additional organisational and cultural barriers to overcome.

Lower risk AI systems including emotional AI

Emotional AI is explicitly mentioned in the AI Act, not as a HRAIS but as a specifically named class of AI system that gives rise to transparency obligations. This is framed alongside AI systems interacting with natural persons that give rise to concerns around manipulation, deception and impersonation. Art 52(2) and Recital 70 discuss the need for making natural persons aware they are interacting with an AI system, particularly for persons with disabilities who need to be provided with notifications and information in an accessible format. This provision depends if it is obvious, based on that context of use or circumstances that an AI system is in operation. If it is obvious, then it is not necessary. This poses the natural question: when would it be obvious that you are subject to an AI system? With a driver-facing camera, as our participants suggested, this could be used for a variety of applications, including emotion sensing, but it may not be obvious this is the case. If transparency is required, what might this require? Art 13 described above suggests some elements, but Art 52(2) implicates design practices in a major way. We argue this is a wider aspect of the proposed AI Act: whilst it sets in motion principles and a framework for risk-based regulation, giving these principles meaning and implementing them requires a turn to AI system design communities. For example, understanding 'obviousness' would require user testing in vehicles to establish if it is obvious an AI system is in operation, and if not, to co-design mechanisms with users to do this effectively. Such user studies are the purview of interaction designers, human-factors engineering specialists, human–AI interaction specialists and design ethnographers. Furthermore, with transparency, how much granularity of information would be required in car for users and how could this be visualised and the experience designed? How would it be communicated in an environment like a car by using affordances of the technology and signifiers in the environment (Norman 1988), such as from the in-car screen to audio output to heads up display to the manual? The AI Act presents an interesting starting point, but like the technologies themselves, needs to be contextualised to make sense and requires collaboration with technologists to implement it.

Importantly, there has been criticism of the AI Act for not going far enough in regulating emotional AI technologies. Major EU institutions, including both the EDPS and the European Data Protection Board (EDPB), have called for stricter regulation. In their joint EDPS/EDPB Opinion on the AI Act (EDPB-EDPS 2021), they stated 'the use of AI to infer emotions of natural person is highly undesirable and should be prohibited, except for certain well-specified use-cases' (para 21). For the EDPS/EDPB, these use-cases include health and research focused applications, and even then, requiring safeguards and data protection limits to be implemented, particularly around purpose limitation regarding use of data. This indicates the policy landscape around these technologies is shifting and more opportunistic uses will be curtailed. Given the safety critical focus of the use of emotional AI in cars, it could be framed as one of the 'well-specified use-cases', particularly given the governance agenda from other domains of EU law driving the use of driver-monitoring systems we discussed above. What appears clear in any case is that deployments of emotional AI in general, and specifically in cars, are going to be more strictly regulated going forward.

Conclusion: should in-cabin sensing be in cars?

Finally, in cars: are we really safest of all? Whilst it remains to be seen whether interior sensing will lower road deaths, this paper has argued that the use of emotional AI and other biometric profiling raises generates other high-level societal risks. This is a sector that is developing quickly, with prominent start-ups such as Affectiva (now acquired by Smart Eye) seeing stable revenue in the automobile sector. Collectively, these firms offer legacy car makers expertise in tracking human states and emotion expressions. For the car industry itself, development of in-cabin-car sensing is motivated by technological potential, policy initiatives to improve safety and reduce deaths, and logics of personalisation. In the first part of the paper, experts on smart cities, the car industry, emotional AI and data protection policy making, helped us generate key themes. These included safety and the broadly positive view that implementation of these profiling technologies may help create safer roads. Yet, there was a general concern about what happens to data generated through in-cabin profiling and the ends for which it may be used. In addition to data, flow was concerned about the creation of data insights, especially regarding psychological assumptions behind emotion profiling, and the adequacy of training data for profiling algorithms. Whilst issues of bias and effectiveness are relevant to all uses of emotional AI, more unique to the car sector is the issue of trust. The car sector was found to be highly opaque and secretive, due to high levels of internal competition. In turn, this means there are no baseline standards for in-cabin sensing of human behaviour, a point exacerbated by governance actors such as NCAP preferring to let the industry decide its own standards. Lack of common standards and independent scrutiny is for this paper a red flag, given these are safety critical systems. Interviewees also recognised the need for better governance. Whilst all saw issues of ethics-washing, some saw regulation as coming too late, with others being hopeful, seeing scope for positive coexistence between technology (and its developers) and society.

Regulation is a key, because the development of in-cabin profiling has impetus from the EU Vision Zero initiative. Indeed, the ambivalence found at the scoping interview stage was replicated in policy itself as technology was found to be both a solution and a problem. Whilst the EU Vehicle Safety Regulation welcomes safety technologies, other regulations flag human-state measures and emotion profiling as risky. With the proposed AI Act framed in terms of risk, we argue that emotional AI in cars should be deemed a HRAIS on the basis that it is safety critical, and that it has real-time profiling characteristics. Further, especially given the opacity of the car industry, and the lack of collective standards and public validation of systems regarding human-state measures and emotion, the need to avoid responsibility gaps is paramount. The Proposed AI Act coupled with EU Vehicle Safety Regulation, are ushering in new guidelines that provide needed checks and balances on how emotional AI for safety will change human–car interactions through in-car sensing.

Acknowledgements

This work is supported by UK Economic and Social Research Council Grant ES/T00696X/1 and UK Engineering and Physical Sciences Research Council Grant EP/V026607/1.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is supported by the UK Economic and Social Research Council project 'Emotional AI in Cities: Cross Cultural Lessons from UK and Japan on Designing for An Ethical Life' [grant number ES/T00696X/1]; and the UK Engineering and Physical Sciences Research Council project 'UKRI Trustworthy Autonomous Systems Node in Governance and Regulation' [grant number EP/V026607/1].

References

Affectiva. 2018. "In-Cabin Sensing." Accessed June 10, 2021. https://go.affectiva.com/in-cabin-sensing.

Affectiva. 2020a. "Optimizing Road Safety and the Mobility Experience: Euro NCAP and Driver Monitoring Systems." Accessed June 10, 2021. https://go.affectiva.com/road-safety-webinar.

Affectiva. 2020b. "Optimizing Road Safety and the Mobility Experience: Advanced Safety Beyond Driver Monitoring." Accessed June 10, 2020. https://go.affectiva.com/advanced-road-safety-webinar.

Affectiva. 2021. "Optimizing the Mobility Experience: Unlocking Occupant Comfort, Wellbeing, and Entertainment." Accessed June 10, 2021. https://go.affectiva.com/automotive-occupant-experience-webinar.

Aptiv. 2019. "Aptiv Partners with Affectiva to Enable the Next Generation Vehicle Experience." Accessed June 10, 2021. https://ir.aptiv.com/investors/press-releases/press-release-details/2019/aptiv-partners-with-affectiva-to-enable-the-next-generation-vehicle-experience/default.aspx.

Automotive World. 2018. "Affectiva and Nuance to Bring Emotional Intelligence to AI-powered Automotive Assistants." Accessed June 10, 2021. https://www.automotiveworld.com/news-releases/affectiva-and-nuance-to-bring-emotional-intelligence-to-ai-powered-automotive-assistants/.

Barrett, Lisa Feldman, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D Pollak. 2019. "Emotional Expressions Reconsidered: Challenges to Inferring Emotion from Human Facial Movements." *Psychological Science in the Public Interest* 20 (1): 1–68.

Braun, Michael, Jingyi Li, Florian Weber, Bastian Pfleging, Andreas Butz, and Florian Alt. 2020. "What If Your Car Would Care? Exploring Use Cases for Affective Automotive User Interfaces." *Proceedings of Mobile HCI* 20: 1–12. doi:10.1145/3379503.3403530.

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research) 81, 77–91. Accessed June 10, 2021. http://proceedings.mlr.press/v81/buolamwini18a.html.

Crabtree, Andy, Lachlan Urquhart, and Jiahong Chen. 2019. "Right to An Explanation Considered Harmful.". Edinburgh School of Law Research Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3384790.

Cresswell, John. W. 1994. *Research Design: Qualitative and Quantitative Approaches*. Thousand Oaks, CA: Sage Publications.

Department for Digital, Culture, Media and Sport. 2021. *Data: A New Direction*. London: DDCMS.

Department for Transport. 2020. *Vehicle Licensing Statistics: Annual 2019*. London: DfT. Accessed June 10, 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/882196/vehicle-licensing-statistics-2019.pdf.

Edwards, Lilian, and Michael Veale. 2017. "Slave to the Algorithm? Why a 'Right to an Explanation' is Probably Not the Remedy You are Looking For." *Duke Law and Technology Review* 16: 18–84.

Enoch, Marcus. 2018. *Mobility as a Service (MaaS) in the UK: Changes and Its Implications. Future of Mobility Foresight Review*. London: Government Office for Science. Accessed June 10, 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/766759/Mobilityasaservice.pdf.

Euro NCAP. 2018. *Euro NCAP 2025 Roadmap*. Accessed June 10, 2021. https://cdn.euroncap.com/media/30700/euroncap-roadmap-2025-v4.pdf.

European Commission. 2020b. *White Paper: On Artificial Intelligence – A European Approach to Excellence and Trust COM(2020) 65 Final*. Brussels: European Union.

European Commission. 2020. *EU ROAD SAFETY POLICY FRAMEWORK 2021–2030. Next steps towards 'Vision Zero'*. Brussels: European Union. Accessed June 10, 2021. https://op.europa.eu/en/publication-detail/-/publication/d7ee4b58-4bc5-11ea-8aa5-01aa75ed71a1.

European Commission. 2021. *Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*. Brussels: European Union. Accessed June 10, 2021. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence.

European Data Protection Board. 2020. *Guidelines 1/2020 on Processing Personal Data in the Context of Connected Vehicles and Mobility Related Applications*. Brussels: EDPB. Accessed June 10, 2021. https://edpb.europa.eu/sites/default/files/consultation/edpb_guidelines_202001_connectedvehicles.pdf.

European Data Protection Board. 2021. TechDispatch 1/2021 – Facial Emotion Recognition.

European Data Protection Board and European Data Protection Supervisor. 2021. *Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*. Brussels: EDPS/EDPB.

European Parliament and Council. 2019. *Regulation (EU) 2019/2144 of the European Parliament and of the Council of 27 November 2019 on Type-approval Requirements for Motor Vehicles and Their Trailers, and Systems, Components and Separate Technical Units Intended For Such Vehicles, As Regards Their General Safety and the Protection of Vehicle Occupants and Vulnerable Road Users, Amending Regulations […]*. Brussels: European Union. Accessed June 10, 2021. https://data.consilium.europa.eu/doc/document/PE-82-2019-INIT/en/pdf.

George, Damian, Kento Reutimann, and Aurelia Tamò-Larrieux. 2019. "GDPR Bypass by Design? Transient Processing of Data Under the GDPR." *International Data Privacy Law* 9 (4): 285–298.

High-Level Expert Group on Artificial Intelligence. 2018. *Draft Ethics Guidelines for Trustworthy AI*. Brussels: European Union. Accessed June 10, 2021. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=57112, 09/01/19.

Information Commissioner Office. 2021. *Special Category Data*. Wilmslow: ICO. Accessed June 10, 2021. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/special-category-data/.

Institute of Electrical and Electronics Engineers. 2019. "P7014 – Standard for Ethical Considerations in Emulated Empathy in Autonomous and Intelligent Systems." Accessed June 10, 2021. https://standards.ieee.org/project/7014.html.

Kitchin, Rob. 2014. "The Real Time City? Big Data and Smart Urbanism." *GeoJournal* 79 (1): 1–14.

Layder, Derek. 1998. *Sociological Practice: Linking Theory and Social Research*. London: Sage.

Luhmann, Niklas. 1979. *Trust and Power: Two Works by Niklas Luhmann*. Translated by Howard Davis, John Raffan, and Kathryn Rooney. Chichester: John Wiley and Sons.

Lupton, Deborah. 1999. *Risk*. London: Routledge.

McStay, Andrew. 2018. *Emotional AI: The Rise of Empathic Media*. London: Sage.

McStay, Andrew, and Lachlan Urquhart. 2019. "This Time with Feeling?' Assessing EU Data Governance Implications of out of Home Appraisal Based Emotional AI." *First Monday* 24: 10.

Miles, Matthew B., Huberman A. Michael, and Saldana Johnny. 2014. *Qualitative Data Analysis*. London: Sage.

Miller, William L., and Benjamin F. Crabtree. 1992. "Depth Interviewing: The Long Interview Approach." In *Research Methods for Primary Care, Vol. 2. Tools for Primary Care Research*, edited by Moira A. Stewart, Fraser Tudiver, Bass Martin, J. Dunn Earl, V. and Norton, and G. Peter. London: Sage.

Norman, Donald. 1988. *The Psychology of Everyday Things*. New York: Basic Books.
SAE. 2018. "SAE International Releases Updated Visual Chart for Its "Levels of Driving Automation" Standard for Self-Driving Vehicles." Accessed June 10, 2021. https://www.sae.org/news/press-room/2018/12/sae-international-releases-updated-visual-chart-for-its-%E2%80%9Clevels-of-driving-automation%E2%80%9D-standard-for-self-driving-vehicles.
TRIMIS. 2021. "Vision Zero Initiative." Accessed June 10, 2021. https://trimis.ec.europa.eu/?q=project/vision-zero-initiative#tab-outline.
Williams, Robin, and David Edge. 1996. "The Social Shaping of Technology." *Research Policy* 25 (6): 865–899.