Journal of Responsible Technology 13 (2023) 100057

Contents lists available at ScienceDirect

Journal of Responsible Technology

journal homepage: www.sciencedirect.com/journal/journal-of-responsible-technology

Towards a master narrative for trust in autonomous systems: Trust as a distributed concern

Joseph Lindley [a,*], David Philip Green [a], Glenn McGarry [b], Franziska Pilling [a], Paul Coulton [a], Andy Crabtree [b]

[a] Lancaster University, United Kingdom
[b] University of Nottingham

A B S T R A C T

In this paper we present findings of research into Trust, specifically within the context of Autonomous Systems. The research is based upon an exploratory workshop attended by domain experts from academia and industry. The aim of the work is to synthesise interdisciplinary and high-level understandings of pertinent issues into a singular and cohesive Master Narrative relating to Trust an Autonomous Systems. The inquiry constructs a Master Narrative that casts Trust as a notion that is necessarily constructed by complex relationships, disciplinary lenses, and multiple concurrent stakeholders. We term this 'Trust as a Distributed Concern'. The paper describes the research and analysis which underpins the concept of Trust as a Distributed Concern and discusses how the concept may be operationalised in research and innovation contexts.

Introduction

Autonomous Systems (AS) are being rapidly adopted, driven by ubiquitous computing, the growth in the digital economy, advances in robotics, and rapid developments in AI. Such systems are those that can take actions with little or no human supervision: automatic checkouts, elevators, traffic lights, and autonomous vehicles. The adoption of AS introduce many societal challenges including ways to manage the relationships between humans and machines, and how to articulate, define, and perceive our Trust relationships (The IEEE global initative on ethics of autonomous and intelligent systems). Different research disciplines necessarily articulate aspects of trust differently, for example by delineating aspects of the design, engineering and operation of a given AS (Charisi et al., 2017). Relating this work directly to Responsible Research and Innovation (RRI) principles [e.g., (Fraaije & Flipse, 2020)] is the assertion that Trust is frequently a "prerequisite to partake in RRI at all" (Asveld et al., 2015) and the role of Trust in RRI is of *particular* importance in the context of AS and AI technologies [cf. (Stahl, 2021, Winfield and Jirotka, 2017)]. The paper draws upon a wide variety of exemplars of designing AS with Trust in mind.

Trust is a cross-cutting challenge, and touches upon tangential research interests including explainability, accountability and transparency; verification, validation, and reliability; governance and regulation. Alongside a wide range of regulations, governance structures and guidelines are regularly published in this domain (Scotti, 2020). This paper describes a research process, which aspires to synthesise the multitude of perspectives which exist on the topic and define a 'Master Narrative' relating to Trust in the context of AS. A Master Narrative is defined as, "*culturally shared stories that guide thoughts, beliefs, values, and behaviours*" [(McLean & Syed, 2015):323]. The premise of the work is not to 'reinvent the wheels' of perspectives on Trust, but rather to explore the potential to unify those perspectives in a productive way. We note that this research is at an early-stage and future inquiry based on this work will seek to triangulate and interrogate the findings.

The paper proceeds in three sections. First, we describe the research approach, explaining how we collected and analysed the data on which the research is based. Next, we discuss the findings, supported by excerpts from the underlying data. Finally, the paper concludes with a discussion section that introduces the rationale for why 'Trust as a Distributed Concern' has value as a Master Narrative for leveraging Trust as an element in Responsible Research and Innovation within the context of AS.

Research approach

This work is based upon a workshop which was designed, facilitated and conducted to explore the themes discussed in the paper. The workshop was staged and captured online using a combination of telepresence (Zoom) and interactive whiteboard application (Miro). The Miro board we used was custom designed to facilitate and support the

* Corresponding author.
E-mail address: j.lindley@lancaster.ac.uk (J. Lindley).

https://doi.org/10.1016/j.jrt.2022.100057

Available online 14 December 2022

two workshop activities (see Fig. 1). The workshop had ethical approval from the organisers' institutional ethics committee.

Data was captured in the form of participant feedback directly in the interactive whiteboard, through recorded discussions between participants that were transcribed for analysis, and an artist (see Fig. 4) captured the event in the form of 'sketch notes', thereby providing a visual summary of proceedings. In total, 22 experts participated in the workshop; 16 were University-based, 1 from the media industry, and 5 from industries with direct interests in AS. Participants responded to an invitation distributed among a large national research hub that is focused on Trust and AS. Although our approach and findings are in concert with the Special Issue's focus on RRI, we did not explicitly ask that participants describe their prior knowledge of RRI. The justification and rationale for our participants experiences being integral to the workshop is grounded upon the implicit role of Trust in conceptions of RRI (cf. Lotte Asveld 2015) combined with the participants expertise in domains directly relevant to or adjacent to Trust and AS. Table 1

The workshop was structured around two core activities. The first activity generated a range of example use-cases, which could be used as the basis for thought experiments relating to Trust (see example in Fig. 2). Generated use cases included recommender systems, autonomous vehicles, building management systems and emergency communication infrastructure.

The second activity sought to map the spectrum of themes relating to Trust across all the use cases and identify commonalities between them before summarising these into overarching ideas that would serve as the basis for Master Narratives. The second activity had three distinct phases. The first (see Fig. 3) centred on identifying themes which spanned the identified use cases, the second mapped mechanisms influencing Trust, and the third specifically sought to propose potential 'Master Narratives'.

The notes generated across the workshop activities and discussions which took place among organisers and participants throughout constitute the data upon which this paper is based. These data, and the workshop itself, were constructed around the ethnomethodological tradition (Crabtree et al., 2012), an approach which is agnostic to pre-existing social theory, and hence is well-suited for our aim of cohering multiple, sometimes non-complementary, expert perspectives.

Findings

In this section, we describe key themes which emerged from the qualitative data gathered at the workshop. We reiterate at this point that this research represents the early stages of an ongoing project, subsequent inquiry will seek to triangulate and test these findings via engagement with non-expert audiences. The findings presented in this paper focus on only *some* aspects of the data and discussions that took place at the workshop. The full scope of our discussions is beyond that which can be represented in this short paper format, to contextualise this paper in terms of those broader discussions please refer to the separate workshop report (McGarry, Lindley, & Mason, 2022). Consonantly whilst only a handful of the workshop participants are quoted directly in this paper, most workshop attendees *did* significantly contribute, and the findings presented here arose from those conversations throughout the various stages of the workshop as well as our own engagement in the process as researchers.

Table 1
Participant list showing occupations and sectors (we note that only some participants are referred to in this paper, please refer to the full workshop report for a more detailed exploration of the data).

ID	Occupation	Sector
TM1	Lecturer	University
TM2	CTO	Industry
TM3	Professor	University
TM4	Researcher	University
TM5	Lecturer	University
TM6	UX Writer	Media
TM7	Policy Research Fellow	University
TM8	PhD student	University
TM9	Research Associate	University
TM10	Research Fellow	University
TM11	Senior Backend Engineer	Industry
TM12	Senior AI Engineer	Industry
TM13	Head of Product	Industry
TM14	PhD Student	University
TM15	Research Fellow	University
TM16	Professor of Marketing	University
TM17	PhD Researcher	University
TM18	Associate Professor	University
TM19	*Withdrawn from research*	–
TM20	UX Designer	Industry
TM21	Research Associate	University
TM22	Research Associate	University

Trust is relative

A recurring theme in the workshop was the idea that trust is, "…*very much circumstantial, and - based on the application, and who you're affecting - your trust is going to differ*" [TM21]. In this section, we explore how vital aspects of any conceptualisation of Trust are shaped according to the context of use and the stakeholders involved.

The example of a recommendation system was a popular talking point in the workshop. Here, it is used to highlight how Trust is assembled from many aspects of a given system, and how a broad 'ecosystem' of stakeholders and interested parties extends beyond the 'service provider to end user' relationship into various other relationships, including people-to-organisation, organisation-to-organisation, organisation-to-system, and system-to-system.

> [TM2]: First of all, there's the bit of trust (relating to the) recommendation system itself … Then there's the client platform, so the platforms that integrate (our system) recommendations, like online marketplaces, you might think of e-commerce platforms like eBay, things like that. There are also other kinds of platforms like lending platforms, gig economy platforms. Then there are the end users who actually use these platforms. There are potentially two types of end users: you have the e-commerce side; and then you have the buyers. But even then, some of them also have marketplaces themselves like Amazon, so you also have sellers. Basically, service providers, service consumers, product sellers, product buyers.

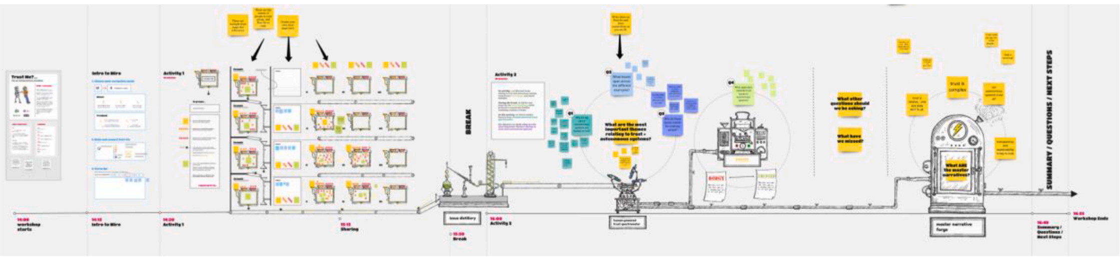

Fig. 1. High level overview of the custom-designed Miro board. See detail view of activity 1 in Fig. 2 and detail view of activity 2 in Fig. 3.

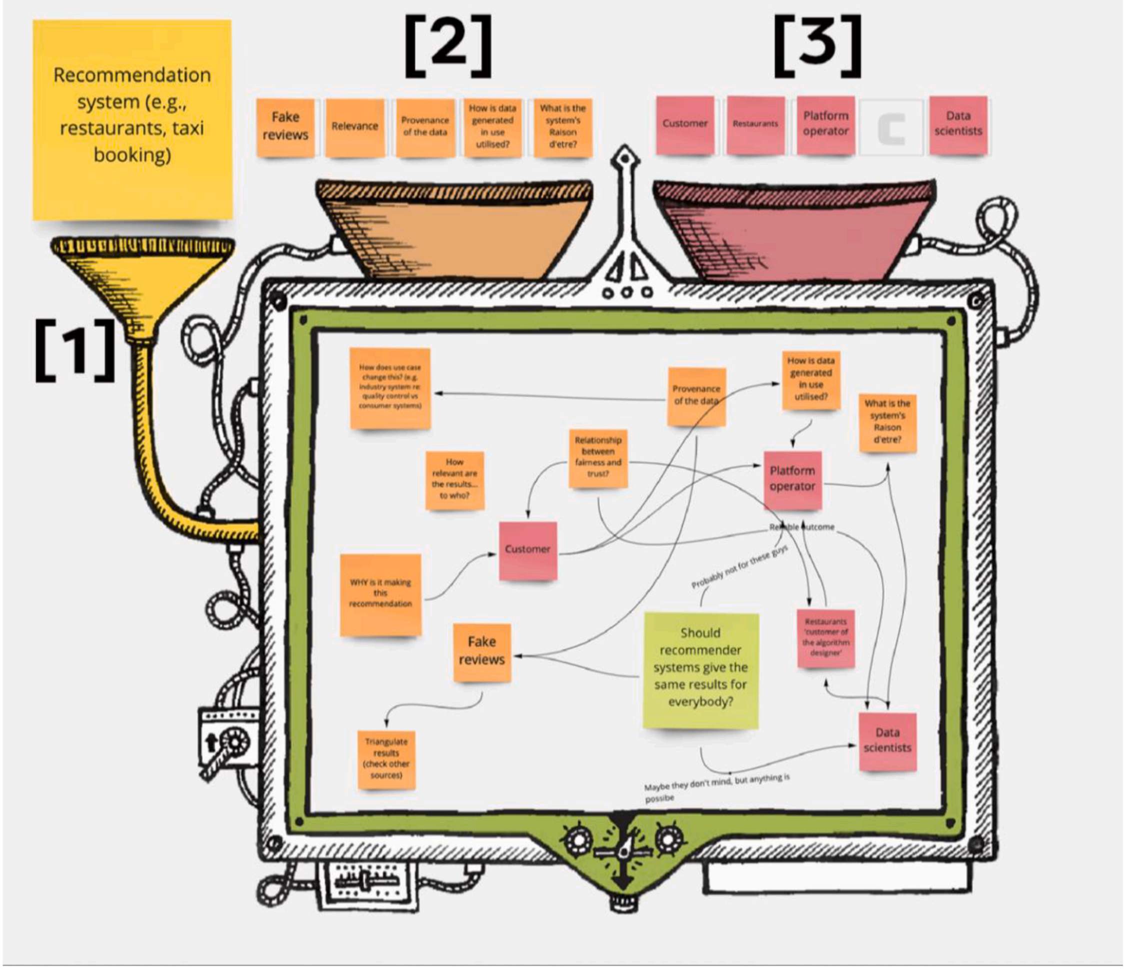
Recommendation system (e.g., restaurants, taxi booking)
[1]
[2]
Fake reviews
Relevance
Provenance of the data
How is data generated in use utilised?
What is the system's Raison d'etre?
[3]
Customer
Restaurants
Platform operator
Data scientists
How does use case change this? (e.g. industry system re: quality control vs consumer systems)
Provenance of the data
How is data generated in use utilised?
What is the system's Raison d'etre?
Relationship between fairness and trust?
How relevant are the results... to who?
Platform operator
Customer
Reliable outcome
Probably not for these guys
WHY is it making this recommendation
Fake reviews
Should recommender systems give the same results for everybody?
Restaurants 'customer of the algorithm designer'
Data scientists
Triangulate results (check other sources)
Maybe they don't mind, but anything is possibe

Fig. 2. Detail view of one group's Miro notes from the first activity. Participants were asked to create Trust Maps consisting of: (McGarry, Lindley, & Mason, 2022) a use case (e.g., a recommendation system), (Asveld et al., 2015) Trust concerns (e.g., 'fake reviews'), and (Charisi et al., 2017) stakeholders (e.g., 'Customer'). The relationships between these factors were mapped collaboratively using connecting arrows.

In the example above, the expert's back-end recommendation system is syndicated across several e-commerce platforms to serve a complex network of goods and services providers, consumers, and organisations. The parts of these interconnected systems rely up on Trustful relationships between them.

> [TM2]: We are building a system that enables people to see what their trusted friends and persons physically recommend or give feedback on stuff.
>
> …
>
> [TM2]: Depending on the level of detail that the recommendation has, if it shows you that other users prefer a certain doctor or a certain pub or whatever, then that might have privacy implications. There are various cryptographic techniques available to mitigate that, but then there is also a trade-off with respect to efficiency of the system and so on. So how can the system be trusted in this respect?
>
> …
>
> [TM2]: So basically, you would need to trust the system to read your situation correctly, to identify persons that you trust correctly for the particular service, and context and situation.

In the case of the recommender service, the nature of Trust relates largely to privacy, data, and meaningful interpretations of context. A particular recommendation may involve the exposure of sensitive personal information, for example the circumstances surrounding a clinic's recommendation, or time and location patterns that might be inferred from a taxi service recommendation. The apparent technical challenge for this use case lies in the trade-off between these privacy issues and the efficiency of the autonomous (in this case AI-driven) part of the system to analyse and deliver personalised peer-to-peer recommendations autonomously. In more general terms, trade-offs such as these crucially turn upon the *risks* involved in relation to the system's context of use and its design constraints.

Trust is the "other side of the coin" to risk

The likelihood an AS could cause harm to a human is a key element of any construction of Trust. Hence, 'risk' was described as "*the other side of the coin*" to Trust because "*where you require trust, there is some risk that something [bad] will happen*" [TM2]. Hence a significant component of Trust relates to the ability to understand, qualify, and quantify levels of risk. In turn, risk is relative to a given system's context of use. For example, in systems which act in the physical world notions of risk relate to whether malfunction or failure could cause physical harm to people in and around the operating environment. In this sense, Trust is a reflection of *safety* risk.

> [TM5]: The safety would be the first (issue) […] There are a couple of things. One is people generally, or even the regulators, or when you talk to the CAA (Civil Aviation Authority), they're worried about the aircraft falling from the sky for whatever reason, or crashing into a building, crashing into something it's not supposed to crash into, landing somewhere it's not supposed to land, these sort of things. So:

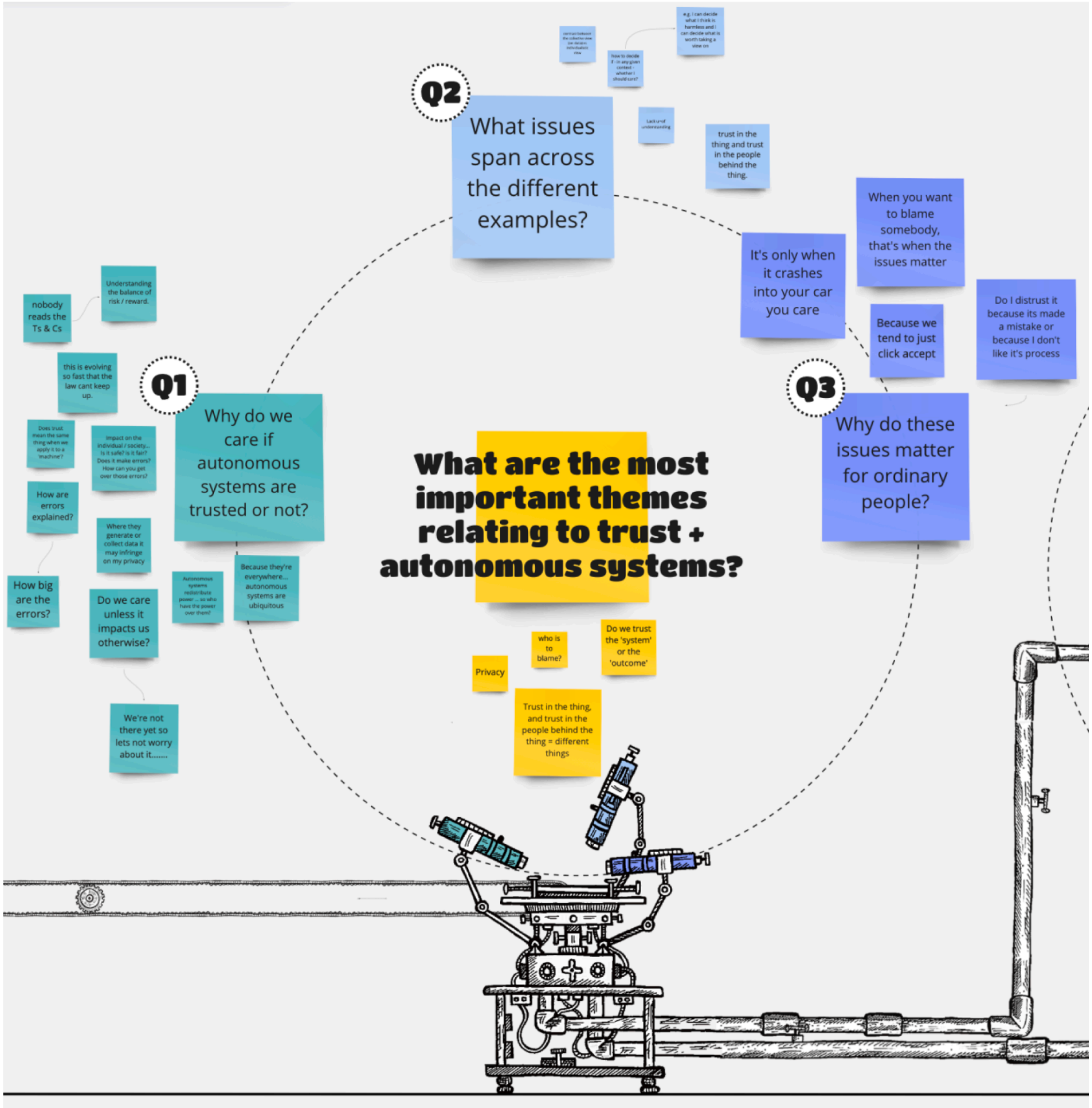
What are the most important themes relating to trust + autonomous systems?
Q1
Why do we care if autonomous systems are trusted or not?
nobody reads the Ts & Cs
Understanding the balance of risk / reward.
this is evolving so fast that the law cant keep up.
Does trust mean the same thing when we apply it to a 'machine'?
Impact on the individual / society... Is it safe? is it fair? Does it make errors? How can you get over those errors?
How are errors explained?
Where they generate or collect data it may infringe on my privacy
How big are the errors?
Do we care unless it impacts us otherwise?
Autonomous systems redistribute power ... so who have the power over them?
Because they're everywhere... autonomous systems are ubiquitous
We're not there yet so lets not worry about it.......
Q2
What issues span across the different examples?
contrast between the collective view (on data) vs individualistic view
how to decide if - in any given context - whether i should care?
e.g. I can decide what I think is harmless and I can decide what is worth taking a view on
Lack of understanding
trust in the thing and trust in the people behind the thing.
Q3
Why do these issues matter for ordinary people?
It's only when it crashes into your car you care
When you want to blame somebody, that's when the issues matter
Because we tend to just click accept
Do I distrust it because its made a mistake or because I don't like it's process
Privacy
who is to blame?
Do we trust the 'system' or the 'outcome'
Trust in the thing, and trust in the people behind the thing = different things

Fig. 3. Detail view of the first part of the workshop's second activity where participants were asked to begin to identify common themes across the previously developed Trust Maps (see Fig. 2).

> 1) trusting the ability of the aircraft to perform what it's supposed to perform; and 2) trusting that if something goes wrong it's not going to cause a lot of damage or have a very high or significant impact on people and people's lives.

Cyber-only systems (e.g., recommendation systems) present different kinds of hazard. For example, a personal/private data breach is unlikely to pose an immediate risk to physical safety, but it could have negative psychological, social, or cultural impacts. These risks are less clearly-defined but – like physical safety risks – they have given rise to an 'arms race' between usage and regulation. "*Stuff is evolving so fast that regulation and law and stuff like that can't keep up with it*" [TM9]. This, in turn, raises questions about accountability, which links to a multitude of concerns, including levels of autonomy, the role of the system and the role of human actors, the system's context of use, the inherent/residual risks involved, the governance of novel highly automated systems. Based on the spectrum of expertise in the workshop, this balance of responsibilities was conceptualised as being distributed across a system, its designers, and its operators.

Yet *how* this balance is achieved is not consistent across use-cases. For example, information that may be necessary to build up an accurate picture of Trust is obscured by a systems interface (intentionally or not) and/or overlooked by users. In this sense, it is useful to conceptualise trust as a balance of risk and reward. A classic example, which came up in various guises, was the phenomenon of '*skipping the terms and conditions*'.

> [TM6]: It's very much I think a risk/reward thing, because on most of these things no-one's going to read the terms and conditions, no-one's going to read how trustworthy it is or what (data) it's collecting. […] If there's a delivery drone that can get you your stuff in an hour but takes all your data, as opposed to a person who takes 24 hours, a lot of people are just going to go, 'I want the thing in an hour so whatever, yes please, tick all the boxes, I don't care how autonomous it is'. It's only when it crashes into your precious car or picks up your child by accident and drops it in the middle of the road … but it's only when something like that happens that people think, 'Oh, hang on a second, who's to blame?'.

Trust starts with people

A common taxonomy of autonomous systems defines six levels of

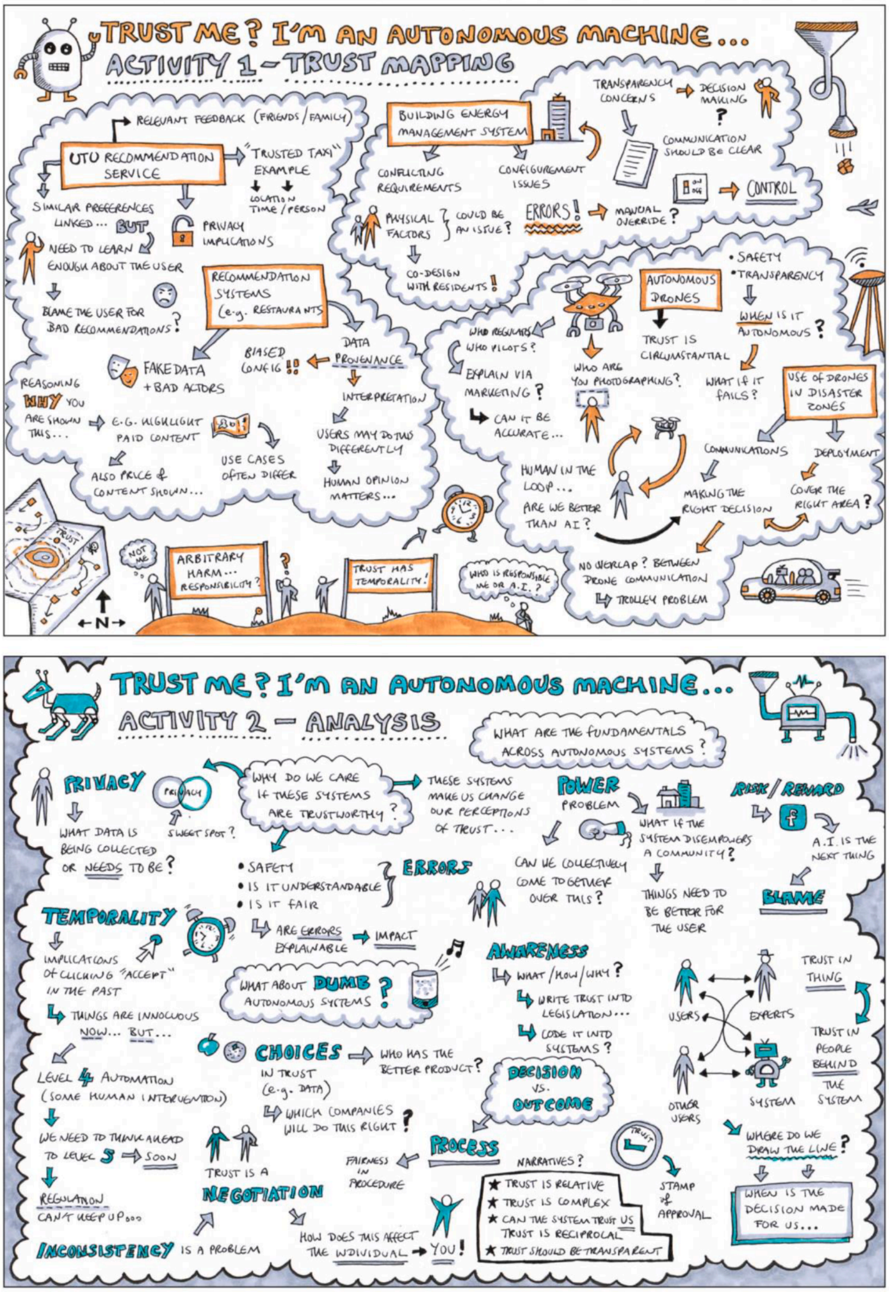
TRUST ME? I'M AN AUTONOMOUS MACHINE...
ACTIVITY 1 – TRUST MAPPING
RELEVANT FEEDBACK (FRIENDS/FAMILY)
UTU RECOMMENDATION SERVICE
"TRUSTED TAXI" EXAMPLE
LOCATION TIME/PERSON
SIMILAR PREFERENCES LINKED... BUT
PRIVACY IMPLICATIONS
NEED TO LEARN ENOUGH ABOUT THE USER
BLAME THE USER FOR BAD RECOMMENDATIONS?
RECOMMENDATION SYSTEMS (e.g. RESTAURANTS
FAKE DATA + BAD ACTORS
BIASED CONFIG !!
DATA PROVENANCE
INTERPRETATION
USERS MAY DO THIS DIFFERENTLY
HUMAN OPINION MATTERS...
REASONING WHY YOU ARE SHOWN THIS...
E.G. HIGHLIGHT PAID CONTENT
ALSO PRICE & CONTENT SHOWN...
USE CASES OFTEN DIFFER
TRUST
←N→
NOT ME
ARBITRARY HARM... RESPONSIBILITY?
TRUST HAS TEMPORALITY!
WHO IS RESPONSIBLE ME OR A.I.?
BUILDING ENERGY MANAGEMENT SYSTEM
CONFLICTING REQUIREMENTS
CONFIGUREMENT ISSUES
PHYSICAL FACTORS
COULD BE AN ISSUE?
CO-DESIGN WITH RESIDENTS!
ERRORS!
MANUAL OVERRIDE?
TRANSPARENCY CONCERNS
DECISION MAKING ?
COMMUNICATION SHOULD BE CLEAR
ON OFF
CONTROL
AUTONOMOUS DRONES
• SAFETY
• TRANSPARENCY
WHEN IS IT AUTONOMOUS?
TRUST IS CIRCUMSTANTIAL
WHO REGULATES WHO PILOTS?
EXPLAIN VIA MARKETING?
CAN IT BE ACCURATE...
WHO ARE YOU PHOTOGRAPHING?
WHAT IF IT FAILS?
USE OF DRONES IN DISASTER ZONES
COMMUNICATIONS
DEPLOYMENT
HUMAN IN THE LOOP...
ARE WE BETTER THAN AI?
MAKING THE RIGHT DECISION
COVER THE RIGHT AREA?
NO OVERLAP? BETWEEN DRONE COMMUNICATION
TROLLEY PROBLEM
TRUST ME? I'M AN AUTONOMOUS MACHINE...
ACTIVITY 2 – ANALYSIS
WHAT ARE THE FUNDAMENTALS ACROSS AUTONOMOUS SYSTEMS?
PRIVACY
PRIVACY
WHAT DATA IS BEING COLLECTED OR NEEDS TO BE?
SWEET SPOT?
WHY DO WE CARE IF THESE SYSTEMS ARE TRUSTWORTHY?
THESE SYSTEMS MAKE US CHANGE OUR PERCEPTIONS OF TRUST...
POWER PROBLEM
WHAT IF THE SYSTEM DISEMPOWERS A COMMUNITY?
THINGS NEED TO BE BETTER FOR THE USER
CAN WE COLLECTIVELY COME TOGETHER OVER THIS?
RISK/REWARD
A.I. IS THE NEXT THING
BLAME
• SAFETY
• IS IT UNDERSTANDABLE
• IS IT FAIR
ERRORS
ARE ERRORS EXPLAINABLE → IMPACT
TEMPORALITY
IMPLICATIONS OF CLICKING "ACCEPT" IN THE PAST
THINGS ARE INNOCUOUS NOW... BUT...
WHAT ABOUT DUMB AUTONOMOUS SYSTEMS?
AWARENESS
WHAT/HOW/WHY?
WRITE TRUST INTO LEGISLATION...
CODE IT INTO SYSTEMS?
USERS
EXPERTS
OTHER USERS
SYSTEM
TRUST IN THING
TRUST IN PEOPLE BEHIND THE SYSTEM
LEVEL 4 AUTOMATION (SOME HUMAN INTERVENTION)
WE NEED TO THINK AHEAD TO LEVEL 5 → SOON
REGULATION CAN'T KEEP UP...
CHOICES → WHO HAS THE BETTER PRODUCT?
IN TRUST (e.g. DATA)
WHICH COMPANIES WILL DO THIS RIGHT?
DECISION VS. OUTCOME
TRUST IS A NEGOTIATION
FAIRNESS IN PROCEDURE
PROCESS
NARRATIVES?
INCONSISTENCY IS A PROBLEM
HOW DOES THIS AFFECT THE INDIVIDUAL → YOU!
★ TRUST IS RELATIVE
★ TRUST IS COMPLEX
★ CAN THE SYSTEM TRUST US TRUST IS RECIPROCAL
★ TRUST SHOULD BE TRANSPARENT
TRUST
STAMP of APPROVAL
WHERE DO WE DRAW THE LINE?
WHEN IS THE DECISION MADE FOR US...

Fig. 4. During both workshop activities an artist summarised discussions using a "sketch note" technique. In addition to the themes which underpin this paper, the sketch notes highlight the diversity of the discussions that took place at the workshop.

automation (0-5). It is only at level 5 that humans cease to have agency in the actual operation of the system (i.e., are no longer needed), hence for most AS there is a 'human-in-the-loop'. The workshop participants recognised this and accordingly common was the idea that any conceptualisation of Trust in an AS needs to account for human behaviours.

> [TM10]: I think the human operator of the autonomous system (is a stakeholder in the system) as well - it depends on the level of autonomy - if we have like a human in the loop, so there is some kind of accountability for a human controller as well.

While, on a positive note, it was felt that human input can often *improve* the functionality of autonomous systems, the complementary notion of human fallibility was also raised.

> [TM6]: I think it's really interesting that human in the loop thing, because it assumes that a human knows best, and we all know humans haven't always made the best decisions all the time. So yeah, I just find it interesting that we think that humans can save computers and AI.

> [TM14]: I was just talking about the human in the loop, that there are different approaches when the system has doubts about the human if the human is suggesting a wrong decision. So, there should be communication saying 'do you think that is the best thing to do or not?'; 'this is why I think it is not the right thing to do'; or, 'okay I accept your recommendation'. It's interesting.

Related to the question of how humans might (or might not) play a functional role in the operation of an AS is the corresponding risk of humans acting in bad faith. This applies to system designers and developers who might make AS do undesirable things, nefarious users, or hackers.

]

> [TM2]: Definitely hacking is an issue. Then there's also the issue of giving fake reviews, like you're recommending your friend's car service even though you know your friend is actually not such a great mechanic. So, in our system we basically allow only to give reviews when the reviewer actually has used the service, probably. Okay, probably is a big word, but if the system believes it.

This example relates to the idea of 'transitive trust', which was explained as "*a bit 'trust-by-proxy', saying okay, because you trust this person for this particular service and this context to give you a recommendation, that means that if that* person *recommends the service and trusts the service basically then you will also trust the service*" [TM2]. When Trust relies on specific data then it is related to a much more general type of problem—data quality.

> [TM2]: The system focuses or hones-in one particular aspect of that user because that's the data they got, but that might not be very well represented.
>
> …
>
> [TM2]: Like you have groups of people who trust each other, and the system learns that, and then you only ever see recommendations from those people, you never look outside. So that makes it hard for newcomers or for maybe minorities in some cases to get a foothold in certain services or industries.

Discussion

The workshop this paper is based upon identified a huge variety of insights and perspectives, however the themes identified above have been selected here to support and describe the rationale for a 'Master Narrative'—*Trust as a Distributed Concern*. In this final section, we reflect on how the workshop findings give rise to this narrative, and how Trust as a Distributed Concern may be operationalised when applying RRI in a AS context.

Across our findings, each theme exhibits breadth and complexity. Often, they are constructed from more than one disciplinary perspective, and equally often, they are dependent on (multiple) contextual qualifiers, i.e., "ifs" and "buts". Moreover, as most AS are part of network-enabled computer systems, any agency within the system must immediately be considered as relative to the other actants in the system (McGarry, Lindley, & Mason, 2022). Such systems thinking connects to a similarly broad and diverse range of literature and viewpoints from, for example, Science and Technology Studies (Franklin, 2017, Haraway, 2016), Human-Computer Interaction (Hauser et al., 2018), and Design (Coulton & Lindley, 2019). Whilst contemporary AS frequently have 'humans-in-the-loop' (Enarsson et al., 2022), even this well-supported approach is not straightforward given that humans are variously unpredictable, inconsistent, and unreliable. The layers of complexity that are evident when considering Trust are necessary to address the issue holistically, but they are hard to manage in practice. It is in response to this complexity that we suggest a 'Master Narrative' such as the one we propose may be a useful heuristic for embedding principles of responsibility into research and innovation processes from the outset.

Based on this research, the Master Narrative we propose is that Trust should not be considered as something which is binary (i.e., present, or not) but as a relative concept (i.e., something which exists to a greater or lesser degree) or even one which may be measured on a spectrum. Most aspects of Trust are relative. For example, any measurement or perception of Trust is liable to change over time and/or be attenuated by contextual circumstances. In this sense we might imagine Trust as a *series* of inter-related spectra—a concept that might be best described in terms of a 'gamut'. Finally, our research clearly describes how different actants' views on the same system offer concurrent but incongruent conclusions. Hence, the gamut which might represent Trust is actually distributed across many different possible perspectives. This is the tentative formulation of Trust as a Distributed Concern that our workshop findings point towards.

We note that our finding resonates with many examples of prior research that finds similarly that matters of Trust can rarely be described in absolute or binary terms. Whether considering how a social theory of Trust should be constructed (Misztal, 1992), deconstructing technical terminologies used to articulate aspects of Trust (Lipton, 2018), or exploring how technology mediates contextual trust (Semaan & Mark, 2011), scholars consistently identify the complexity and nuance that surfaced in our workshop. Our proposed Master Narrative, or 'model', is intended to serve as an abstraction or proxy for these complexities. It is an organising principle to connect the wealth of abstract positions to the minutiae of specific research questions or innovation challenges. Future work based on this research will prototype ways that the Mater Narrative may be operationalised in practice, in the meantime we conclude with some potential routes forward.

While a gamut of possibilities distributed across multiple interested parties represents an unmanageable amount of variation, it is precisely that flexibility which makes this Master Narrative powerful. Moreover, it is relatively trivial to identify the relevant spectra for any given AS, context, or set of stakeholders—this was a process that our experts consistently went through (and navigated with ease) during the workshop. Future publications will detail prototype toolkits that—although beyond the scope of this publication—have been produced with the intention of helping practitioners identify and define the spectra of Trust relevant to a given situation. However, to operationalise the Master Narrative, we can advise three practical steps.

1) *Identify relevant spectra.* Any given AS may have several of these that are relevant, and those which are relevant may change depending on context, e.g., how 'regulated' or 'risky' a particular AS may be for the user.
2) *Establish measures or quantities for each spectrum.* These may be numerical (e.g., 'out of 100') or relative (e.g., 'above average', 'below average', 'average').
3) *Contrast different measures according to different contexts.* With spectrums and measures identified, the final step is to contrast how varying contexts will vary reasonable points on the spectrum (e.g., an autonomous vehicle with a full load of passengers constitutes a distinct context to an autonomous vehicle *without* passengers).

Further research is necessary to evaluate how these steps integrate with existing RRI approaches.

The obvious limitation of this approach is that we must accept for any given AS it is impossible to give a conclusive answer to the question *Is this system Trustworthy?* Instead, when working with the notion that Trust is a Distributed Concern, we would have to answer *it depends*. However, this Master Narrative does provide a reasonable means to structure the information upon which the perception of Trust is contingent and provides the framework to answer the follow up question; *it depends on what?*

In this paper, we have reported on research which aimed to capture a group of experts' first-hand experience of practicing Responsible Research and Innovation. Through the analysis of our qualitative data, we have critically reflected on how complexity poses challenges to conceptions of Trust. We propose the Master Narrative *Trust as a Distributed Concern* as a potential means to overcome those challenges.

Declaration of Competing Interest

This work was funded by the EPSRC TAS Hub which has an association to the special issue. No other conflicts of interest.

Acknowledgements

This work was supported by supported by UK Research and

Innovation (reference MR/T019220/1) and the Engineering and Physical Sciences Research Council (reference EP/V00784X/1).

References

Asveld, Lotte, Ganzevles, Jurgen, & Osseweijer, Patricia (2015). Trustworthiness and responsible research and innovation: The case of the bio-economy. *Journal of Agricultural and Environmental Ethics, 28*(3), 571–588. https://doi.org/10.1007/s10806-015-9542-2

McGarry, Glenn, Lindley, Joseph, Green, David Philip, Mason, Zach. 2022. Workshop Report: Trust me? I`m an autonomous system. https://www.research.lancs.ac.uk/portal/services/downloadRegister/356980872/Trust_me_Report_1.0.pdf.

Vicky Charisi, Louise Dennis, Michael Fisher, Robert Lieck, Andreas Matthias, Marija Slavkovik, Janina Sombetzki, Alan F. T. Winfield, and Roman Yampolskiy. 2017. Towards Moral Autonomous Systems.

Coulton, P., & Lindley, J. G. (2019). More-than human centred design: Considering other things. *Design Journal, 22*, Article 4. https://doi.org/10.1080/14606925.2019.1614320

Crabtree, Andrew, Rouncefield, Mark, & Tolmie, Peter (2012). *Doing Design Ethnography*. London: Springer-Verlag.

Enarsson, Therese, Enqvist, Lena, & Naarttijärvi, Markus (2022). Approaching the human in the loop – legal perspectives on hybrid human/algorithmic decision-making in three contexts. *Information & Communications Technology Law, 31*(1), 123–153. https://doi.org/10.1080/13600834.2021.1958860

Fraaije, Aafke, & Flipse, Steven M. (2020). Synthesizing an implementation framework for responsible research and innovation. *Journal of Responsible Innovation, 7*(1), 113–137. https://doi.org/10.1080/23299460.2019.1676685

Franklin, Adrian (2017). The more-than-human city. *The Sociological Review, 65*(2), 202–217. https://doi.org/10.1111/1467-954X.12396

Haraway, D. J. (2016). *Staying with the Trouble: Making Kin in the Chthulucene*. Duke University Press.

Sabrina Hauser, Doenja Oogjes, Ron Wakkary, and Peter Paul Verbeek. 2018. An annotated portfolio on doing postphenomenology through research products. In *DIS 2018 - Proceedings of the 2018 Designing Interactive Systems Conference*, 459–472. https://doi.org/10.1145/3196709.3196745.

Lipton, Zachary C. (2018). The Mythos of Model Interpretability. *Queue, 16*(3), 31–57. https://doi.org/10.1145/3236386.3241340

McLean, Kate C., & Syed, Moin (2015). Personal, master, and alternative narratives: An integrative framework for understanding identity development in context. *Human Development, 58*(6), 318–349. https://doi.org/10.1159/000445817

Misztal, Barbara A (1992). The notion of trust in social theory. *Policy and Society, 5*(1), 6–15. https://doi.org/10.1080/10349952.1992.11876774

Scotti, Veronica (2020). Artificial intelligence. *IEEE Instrumentation and Measurement Magazine, 23*(3), 27–31. https://doi.org/10.2139/ssrn.3518482

Semaan, Bryan, & Mark, Gloria (2011). Creating a context of trust with ICTs. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work - CSCW '11*, Article 255. https://doi.org/10.1145/1958824.1958863

Bernd Carsten Stahl. 2021. Concepts of Ethics and Their Application to AI. . 19–33. https://doi.org/10.1007/978-3-030-69978-9_3.

Alan F. T. Winfield and Marina Jirotka. 2017. The Case for an Ethical Black Box. . 262–273. https://doi.org/10.1007/978-3-319-64107-2_21.

The IEEE global initative on ethics of autonomous and intelligent systems. *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems (version 2)*..