# Charting the Ecosystem of Trust in Cat Royale

## Or what it takes to trust a robot to play with cats

Steve Benford[1 [0000−0001−8041−2520]], Pepita Barnard[1 [0000−0003−4518−1207]], Sarah Sharples[2 [0000−0003−0288−915X]], Helena Webb[1 [0000−0002−4303−7773]], Clara Mancini[3 [0000−0003−1555−077X]] Ayse Kucukyilmaz[1 [0000−0003−3202−6750]], Simon Castle Green[1 [0000−0003−0681−2555]], Eike Schneiders[4 [0000−0002−8372−1684]], Victor Ngo[1 [0009−0003−0805−5292]], Alan Chamberlain[1 [0000−0002−2122−8077]], Joel Fischer[1 [0000−0001−8878−2454]], Guido Salimbeni[1 [0000−0001−6666−3518]], Nick Tandavanitj[5 [0000−0003−2830−4850]], Matt Adams[5 [0000−0002−8561−7119]], and Ju Row Farr[5 [0000−0003−2392−0152]]

[1] School of Computer Science, The University of Nottingham, Nottingham, UK
{steve.benford, pepita.barnard, helena.webb, ayse.kucukyilmaz,simon.castle-green, victor.ngo, alan.chamberlain, joel.fischer, guido.salimbeni}@nottingham.ac.uk
[2] Faculty of Engineering, The University of Nottingham, Nottingham, UK
sarah.sharples@nottingham.ac.uk
[3] The Open University, Milton Keynes, UK
clara.mancini@open.ac.uk
[4] Department of Electronics and Computer Science, The University of Southampton, Southampton, UK
eike.schneiders@soton.ac.uk
[5] Blast Theory, Brighton, UK
{nick, matt, ju}@blasttheory.co.uk

**Abstract.** We present a detailed and unusual case study of ensuring trustworthiness in social robots in practice—the artwork Cat Royale in which a robot played with a family of three cats. We reveal how delivering Cat Royale involved tackling diverse aspects of trustworthiness beyond the immediate interaction between cats and robot, from the control room, to animal welfare expertise, extensive ethical review, and carefully planned public engagement. We contribute a framework for describing ecosystems of trust comprising five layers: personal workspace, orchestration, wider workspace, organisational culture, and the public sphere.

**Keywords:** Art · Robots · Trustworthiness · Ecosystems · Cat Royale.

## 1 Introduction

Trustworthiness is an existential concern for social robotics. Extensive previous research has identified diverse factors in trustworthiness, often through laboratory experiments that establish relationships between particular combinations. In this paper, we take a different tack, introducing and reflecting on an artwork

called *Cat Royale* in which a robot played more than 500 games with a family of three cats over the course of 12 days. We adopt a responsible AI ecosystems perspective to comprehensively map trustworthiness factors as they emerged in context: from designing and operating the robot, to gaining ethical approval and managing the media, and much between. In so doing, we make two contributions. First, we respond to calls to move social robotics research 'beyond the lab' in ways that can safely explore the ethical boundaries of trustworthiness through real experience [26]. Second, we introduce a five-layer framework that supports a responsible AI ecosystems perspective on identifying and tackling the many and diverse issues involved in ensuring the trustworthiness of social robots.

## 2 Related work

Trust is a highly diffuse and nuanced concept but is often broadly understood as the willingness of one agent to be vulnerable to the actions of another [25], which may be based on a belief that the other is able to achieve a particular goal. An associated and equally important concept is trustworthiness: the characteristics of the other (the trustee) that enable the agent (the truster) to be willingly vulnerable in this way. Trust lies at the heart of successful interactions with technologies such as robots. As trusters, humans and other agents make themselves vulnerable in their interactions with robots and it is essential to identify the characteristics that robots can have in order to warrant their status as trustees.

Human trust in robots has been explored in industrial, healthcare, domestic, tourism, leisure and entertainment domains, with evaluation metrics developed for intertwining human, robot, and environmental trustworthiness factors [15, 8]. Common themes emerging from this literature are that over-trusting robots can lead to the misuse of robots, while under-trusting them can cause disuse [17]. When human agents experience loss of trust in a robot, this can lead to adaptations in their behaviour and ultimately a reduced likelihood that the robot will deliver its objectives [11].

Previous research has identified many factors influencing the trustworthiness of social robots. Following Khavas et al among others, we consider these under the categories of human-related, robot-related and environmental factors [17, 21, 32, 14].

Human factors include expertise, psychomotor skill and prior experience. These may be reflected in the way in which decisions are made and the strategies selected for interaction. These human characteristics interact with the information present within the environment, including the cues and affordances influenced by robot design, as well as wider systems factors including organisational culture, regulation and societal norms.

Robot factors include safety, transparency, explainability, performance, capability, control, adaptability, failure and recovery, communication modes, apparent listening behaviour [2], morphology (including anthropomorphology), self-adaptation to users' personalities, preferences, habits [18], personality and size [15,

8], mimicking cross-cultural gestures [28], facial expressions and small talk [22]. Environmental factors include ethics, authority, and power dynamics [16, 23, 19].

Temporal factors are also relevant to trust. Evidence suggests that trust in social robots is not instantaneous but develops through phases of appropriation, incorporation, and eventually conversion. These phases unfold over repeated interactions and prolonged exposure [9, 31].

A robot's performance and its perceived willingness to engage in high-risk behaviours [30] are also key factors in trust development. A recent study by Stower et al. [30] found high-risk robot behaviour can result in high trust ratings, *but only if* that risk is matched by high performance. This existing work provides useful insights; however, trust and trustworthiness are also highly context-dependent [8] meaning that attention needs to be given to the particulars of how a robot is deployed in a given setting.

Salem et al. highlighted the ethical challenges of studying trust in robots, including the difficulty of obtaining ethical approval for risky experimental scenarios, including those involving deception [26]. Given participants may inherently feel safe in a lab-based study, they call for research to move beyond the lab [26].

A final relevant area of literature is the small but growing field of research that explores animal-robot interaction. While most work focuses on human interactions with robots, the increasing presence of social robots in homes, at work, and other settings makes it inevitable that robots will come into contact with animals. Romano et al. [24] discuss this emerging field and describe the ways in which social group behaviours add further dimensions of complexity for studies with multiple agents, whether animal or robotic. Abdai et al. highlight the importance of measuring multiple interactive behaviours when assessing an animal's responses to robot animacy [1].

In sum, existing work highlights many nuances of trust and trustworthiness in social robotics. In a given scenario, a wide range of technical and social human, robot and environmental factors are relevant and intertwine in complex ways. This motivates adopting a broad and holistic perspective on the trustworhtiness of social robots. Existing work also highlights challenges of researching trustworthiness in the lab, motivating approaches that take robots into the wild.

## 3 Methodology

We followed the established method of *Performance-led Research in the Wild* in which researchers enable artists to realize artworks technically while studying their rationale, process, and audience responses in order to generalise design knowledge [6]. This practice- and artist-led approach falls under the umbrella of Research Through Design in which research knowledge emerges from open-ended and exploratory design practice [36, 12]. Collaborative research with artists offers many benefits, two key reasons for engaging artists with robotics research are: (i) they can challenge and provoke the field to reconsider important societal questions from new perspectives, and (ii) they can improvise socio-technical responses to these, subsequently, inspiring further technical research [4]. Artworks

can be vehicles through which ethical and cultural values are expressed, transformed, or subverted, offering opportunities for public critique as well as tools of social and moral imagination to illuminate better paths to social good [33]. By nature of their public deployment, artworks also typically require a wide range of issues to be fully addressed related to trust and liability, including ethical and other institutional approval processes, and anticipation of public reactions and media responses. The artwork we analyse in this paper—Blast Theory's Cat Royale—is a prime example of this approach, serving as an artistic provocation in response to the question: *"How might we trust robots to care for our loved ones?"* Employing an artwork to address this question enabled us to construct and play out imaginary scenarios. Specifically, Cat Royale invoked a different set of associations for robots: those of luxury, convenience and 'utopias'. The artwork actively engaged the public with these issues and sought to prompt a wider reflection on the assumptions we bring to our use of the technology. At the same time, the public exhibition of the artwork raised various considerations of trust and liability.

## 4 Theoretical orientation

We adopt an ecosystems approach to understanding trustworthiness in which intelligent systems are considered as "socio-technical assemblages of heterogenous components including individual humans, technological artefacts, and social structures" [29]. A *responsible* AI ecosystem is then one that recognises the complexity of how responsibility, in our case, for trustworthiness, becomes distributed across a diverse network of actors, including developers, users, owners, regulators, legislators, the system itself, among others, as we reveal below. This motivates us to chart the responsible AI ecosystem of Cat Royale. To provide a suitable structure for so doing we turn to Wilson and Sharples's layered 'onion model' of human factors in interactive systems design, which places people, artefacts and technologies at the centre; surrounded by tasks and goals; and then the personal physical and virtual workspace; and finally, the wider physical and virtual work environment on the outside [35]. We find this to be an encompassing and yet tractable framework for mapping the actors and responsibilities that comprise an ecosystem of trust.

## 5 Introducing Cat Royale

Blast Theory was commissioned by the UK's Trustworthy Autonomous Systems Hub (TAS Hub) to create an artwork to engage the public with the question of trust in robots. Their response was Cat Royale, a so-called 'utopia' for cats, a luxurious environment intended to cater to their every need. Following advice of animal-welfare experts, the artists created a bespoke enclosure with ample feeding stations, sleeping dens, walkways, a water fountain, scratching post, cat grass and litter trays, to house a family of three cats (Clover, Pumpkin, and Ghostbuster) for six hours a day over 12 days. In the centre, a robot arm tried to


**Fig. 1.** A montage of images from Cat Royale. *Top left:* Clover. *Top middle:* Pumpkin. *Top right:* Ghostbuster. *Bottom left:* The three cats playing with a toy offered by the robot. *Bottom right:* The robot waves a toy from behind the four magnetic toy racks.

increase their happiness by picking up toys from nearby magnetic racks every ten minutes and wielding them in a series of pre-programmed movements designed to attract their attention and engage them in play [27] (see Figure 1 bottom row). More than 500 games were filmed during the 12 days, with the footage edited into social media highlights[2] and an eight-hour long movie, touring as a video installation to galleries worldwide.

Previous papers have reporting findings from Cat Royale: how the artists designed a multi-species world to accommodate cats and robots and how the robot operator helped improvise new responses [27]; a case study of how collisions with robots can be reconceived as being complex and extended tangles [3]; and an account of the eighteen-month long journey through ethical approval that revealed tensions in multi-disciplinary and multi-species research [7]. The distinct contribution of this paper is to systematically chart the extensive socio-technical ecosystem required to ensure the trustworthiness of Cat Royale.

## 6 Charting the ecosystem of trust in Cat Royale

We now map the ecosystem of trustworthiness in Cat Royale, revealing the many agents that were involved in establishing the trustworthiness of the robot to play with the cats along with the various roles they played and issues that had to

[2] Cat Royale (`https://youtu.be/sl6nr8B5jqQ`)

be considered. Taking inspiration from the model of [35], we present this as a series of 'onion layers' shown in Figure 2, with the **enclosure** that housed the cats and robot placed at the centre, gradually expanding out through the **control room**, surrounding **artist's studio** where the enclosure was located, the **development and advisory team** who enabled the work, and finally the various **audiences** for the work. At this point, we use names and terms that are specific to Cat Royale. Later on in Section 7 we generalise them to give a more broadly applicable framework for considering trust in robots.

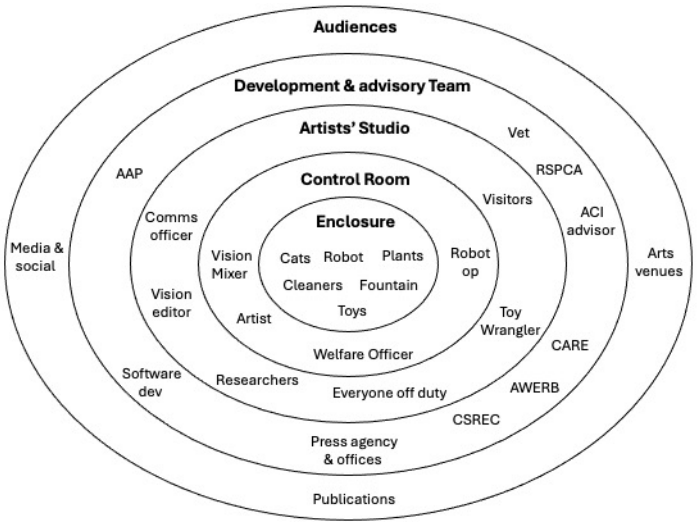


**Fig. 2.** Mapping the social ecology of trust of Cat Royale.

### 6.1 Enclosure

The **enclosure** provides the immediate context in which the *robot* engages the *cats*. Key issues here were whether the robot could be trusted not to harm the cats but also whether the cats could be trusted not to damage the robot. The robot arm, a Kinova Gen3 lite, was chosen to minimise any risk of harm to the cats due to its light payload and minimal range. The entire enclosure, including the robot and its manipulated toys, were designed specifically to be *trustworthy* to the cats, encouraging their voluntary engagement with both the robot and the play activities, for example through the provision of high perches and walkways from which they could observe the robot before choosing to approach it. Its striking visual design was chosen to be appealing to human viewers, hopefully to the cats too (following discussion with animal behaviour experts), and so that the cats would visually stand out from the background with a view to filming and potential uses of computer vision [27]. Also present, and relevant to the

matter of trust, were various *toys* that the robot wielded and other artefacts such as the*drinking fountain*. The toys (which might be considered end effectors for the robot) needed to be safe for the cats to paw and bite (not toxic, sharp or raising the risk of choking). A previous study of Cat Royale showed how these additional artefacts played a pivotal role in safety as they often became loose and got tangled with each other and the robot; for example when Clover wrestled a toy away from the robot, dragged it through the enclosure, tangling with and pulling over the water fountain [27]. Also present were other non-human species in the form of various *plants*. Human *cleaners* would periodically enter the environment to tidy up and clean it so that it was sanitised and safe for the cats. Finally, the enclosure contained eight cameras that were carefully placed to afford wide coverage of the cats' different activities.

### 6.2 Control room

The **control room**, located immediately outside the enclosure and connected via a one-way mirror was an important site for various activities concerned with trustworthiness. A key role here was the *robot operator* who would continuously monitor the robot, prepare it for each game, trigger the pre-recorded series of movements and hold down a deadman's switch while each movement unfolded. Occasionally they had to take manual control and improvise new actions, for example when toys became tangled around the robot or when a cat engaged the robot in a tug-of-war, requiring them to judge the safest moment to release the toy to minimise the risk of it recoiling back and possibly colliding with the cat[6]. The *artist* was also continually present, assuming responsibility for strategic (rather than immediately operational) decisions concerning the robot. They scored each game using the Participation in Play (PIP) scale, an instrument developed by feline behaviourists to score the level of engagement cats exhibits during play [10], feeding the results into the 'decision engine' that learned to recommend games for each cat, which they could choose to accept or reject. Also present was a *cat welfare officer*, a trained professional in cat welfare whose role was to monitor well-being and raise any concerns. Their role was especially important to ensuring trustworthiness, not least because the cats could not directly tell us how they felt about the robot, and so we needed to rely on an expert trained in cat behaviour to judge whether the cats appeared to trust the robot as a form of what is referred to as 'mediated consent' in the animal-computer interaction literature [20]. Finally, a *vision mixer* was present to continuously monitor the outputs of the eight cameras and generate a video edit in real time that best captured the action within the enclosure.

### 6.3 Artists' studio

To prevent the cats from escaping into the wider hazardous world outside, the enclosure was constructed behind locked doors in the *artists' studio*, a complex of

[6] See https://youtu.be/sl6nr8B5jqQ from 3:01–03:27

several rooms, including resting, storage, office, and kitchen facilities. This also ensured only authorised *visitors* could gain entry, an important consideration for the security of both robot and cats, especially given concerns that there could be a negative response to the notion of the project leading to unwanted attention at the studio. An important function of the studio was to provide a stress-free space where *everyone off duty*, cats included, could relax away from the robots. A *toy wrangler*, dedicated to maintenance, cleaning and preparation of the more than 50 attachments that the robot wielded, also worked in the studio along with a *vision editor* who did further editing of video material, for example to generate daily highlights, working alongside a *communications officer* who managed press, media and social media. The studio also provided the location for a daily review meeting where the team met to discuss the previous day's events and make adjustments for the next day. Finally, the studio provided a base for *researchers* to interview people, write up their notes and collate video and other data.

### 6.4 Development and advisory team

Significant to the trustworthiness of Cat Royale was the involvement developers and advisors. Two *software developers* were responsible for the robot control interface and decision engine. A specialist in veterinary behavioural medicine (*Vet Advisor*), an expert on animal-computer interaction (*ACI Advisor*), and the UK's Royal Society for the Prevention of Cruelty to Animals (*RSPCA*) all provided guidance concerning design of the enclosure, games and procedures for monitoring and ensuring cat welfare. A *press agency* was engaged to help devise and execute a media communications strategy, including a crisis management plan in case of having to manage an incident such as (the unlikely event of) a cat being injured or the outbreak of a social media storm. The artists recruited an *Audience Advisory Panel* comprising people with backgrounds across the arts and technology who met regularly to review progress and provide advice. A vital aspect of trustworthiness was ethical approval for the project. Given the ethical sensitivities and United Kingdom's tight regulations surrounding the involvement of animals in research, this unsurprisingly proved to be a complex and lengthy process (see [7] for a full account), with the project passing through three different ethics committees at the University of Nottingham, the Animal Welfare and Ethical Review Body (*AWERB*) which regulates animal research at the University and ensures legal compliance, the Committee for Animal Research and Ethics (*CARE*) in our Veterinary School who gave further advice on ensuring cat welfare, and the Computer Science Research Ethics Committee (*CSREC*) who advised on the technical and data privacy aspects of the project. The extensive and iterative input of these three bodies was essential to ensuring that the robot could be trusted to safely and legally engage the cats and helping us address wider issues about communications and reputational risk.

### 6.5 Audiences

Finally, we note various external **audiences** for Cat Royale who were invited to consider the question of whether one can trust a robot to play with cats, and by extension to care for humans. The movie has toured to multiple *arts venues* (galleries and festivals) worldwide and continues to do so. There has been extensive *media and social media* coverage which has engaged the wider public. The research team has targeted *publications* at venues spanning human-computer interaction, human-robot interaction, and animal-computer interaction.

## 7 Charting ecosystems of trust in social robots

Cat Royale is, of course, an unusual example of social robotics, being both an artwork and focused on cats as the beneficiaries. So what general lessons can we take from it? We now revisit our five onion layers, considering ways in which they might generalise to other applications (see Figure 3). Our aim is to provide developers, researchers and potentially policy makers with a framework to explore the wide variety of factors involved in ensuring the trustworthiness of robots. Note that we often talk about 'spaces'—this is in a largely conceptual sense of being 'design spaces'. Where physical space is important (as it is in several layers), we further note that this could be distributed (e.g., the case of telepresence robots that connect people in different places).

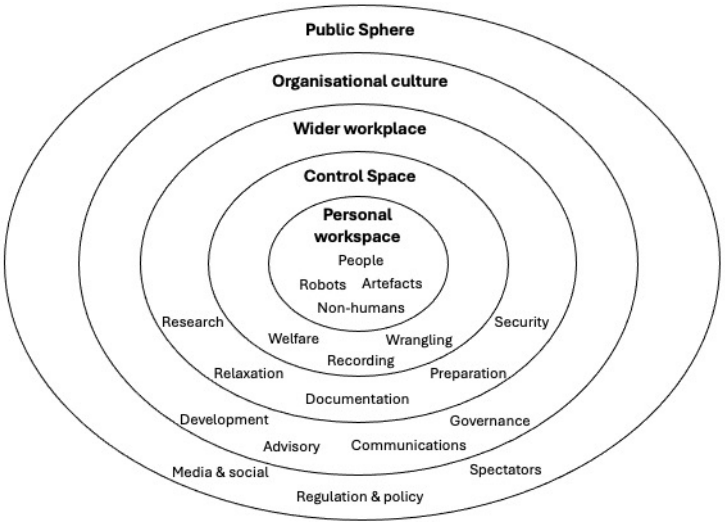


**Fig. 3.** A generalised social ecology of trust for robots

### 7.1 Personal workspace

At the heart of the **personal workspace** are the *people* and *robots* that interact. As discussed earlier, see Section 2, social robotics is replete with knowledge about the many factors that may contribute to trustworthiness between them. However, our experience with Cat Royale highlights additional factors. First are additional *artefacts* that need to be designed to be trustworthy too; they can, for example, also be a cause of mess, for example debris after a collision [3] which can increase the risk of further problems. Cat Royale highlights the presence and impact on *non-human species* and the additional challenges this raises for trustworthiness. This may be deliberate, for example robots that tend to animals in the home, farming or the wild, or accidental, as animals will likely encounter robots in these contexts anyway (see internet videos of cats riding robot vacuum cleaners). Recognising the importance of designing for non-human species reflects a trend towards 'beyond human' thinking in HCI, which aims to decentre humans from design in order to then accommodate wider environmental concerns [13, 34].

### 7.2 Orchestration

The **Orchestration** layer enables the ongoing monitoring and steering of an experience, often from behind the scenes, though sometimes visibly 'front of house'. Orchestration has previously been studied as a feature of interactive artworks [5], but equivalents can be found in other contexts. One aspect of orchestration that is familiar to robotics is *wrangling*, which refers to the control of a robot's ongoing operation, perhaps manually triggering actions, stepping in to improvise, and managing safety such as operating a deadman's switch or emergency stop button, and which may be carried out from a separate (even remote) control space or when co-present with the robot and its users. We have seen how trustworthiness relies on actively managing welfare, especially when participants are vulnerable and unable to consent (e.g., for some care robots), lack knowledge of robots (e.g., delivery robots on the streets), and includes more general workplace safety, well-being and even employment rights. Finally, generalising the role of Cat Royale's vision mixer, the orchestration layer may include mechanisms for *recording* interactions for research, publicity or compliance purposes.

### 7.3 Wider workplace

These play further important roles with respect to trustworthiness. *Security* is important to protect against theft or damage, both accidental due to untrained visitors being present, and deliberate sabotage involving malicious acts. This wider workplace may also house those involved in the *documentation* of robot interactions, including the editing of recordings, but also the generation of additional materials such as field notes, for research, performance analysis and feedback, to capture datasets for training AI models, or for public media. We saw how it was important to support *relaxation* away from robots where people can unwind, socialise and recover without having to constantly attend to the

demands of interacting with robots or being documented, with clear thresholds into the control and personal workspace layers that signal that they now need to attend to these extra demands. Finally, the wider workspace may house workshops and storage to support preparation and maintenance of the robots and ancillary technologies.

### 7.4 Organisational culture

Cat Royale illustrated various ways in which organisational culture influences the trustworthiness of robots. The first is through the *development* process spanning hardware (procurement or development), software, artefacts, surrounding environment and user experience. Also important is introducing external *advisory* roles in the form of expert consultants (e.g., the Vet and RSPCA in Cat Royale) and panels representing potential users and/or the public). Internal *governance* includes local ethical review processes and may need to cover vulnerable users (e.g., animals in our case). Finally, a coherent and carefully planned communications strategy is essential for explaining the project to build trust and dealing with problems if something goes wrong, potentially involving professional advice and connecting press offices where multiple organisations are involved.

### 7.5 Public sphere

The *public sphere* contains important external influences on the trustworthiness of social robots. *Regulation and policy* may span general technology regulation, data protection, security and safety, but also regulation and policy in specific application domains, for example the extensive regulation of healthcare and transportation (or even animal protection as we saw in Cat Royale). Ultimately, however, it is perhaps the public who will be the long-term arbiter of trustworthiness. This includes *spectators* who directly witness robot encounters, including deliberately targeted audiences at demonstrations, educational events and cultural venues, but also ‘ unwitting bystanders’ who happen to encounter robots as they go about their daily lives (e.g., encountering delivery robots or autonomous vehicles on the streets or service robots in hospitals and museums). Critically important is managing the *media and social media* where public perceptions of trust are forged.

## 8 Conclusions

As an artwork in the wild, Cat Royale reveals the diverse factors involved in ensuring social robots are trustworthy. Those concerned with direct interaction with robots (e.g., safety, reliability, and social behaviours) form just one part of a much larger ecosystem of trust. Also important are additional layers concerned with the orchestration of experiences, embedding them into the workplace, connecting them to organisational culture, and ultimately carefully presenting them

within the public sphere—with each of these layers comprising multiple perspectives. In short, ensuring trustworthiness is complex. We do not offer Cat Royale as a solution. Nor is our framework intended as a simple 'cook book' of guidelines that can be readily translated to other settings. Rather, our aim has been to reveal, through a concrete case study and an ecosystems perspective, the breath and complexity of trustworthiness in relation to social robots as an inspiration to further research and the value of artistic methods. For those wanting to apply the ecosystem's perspective themselves, this concrete case study can illustrate different factors that should be considered and Figure 3 offers a starting point to map a social ecology of trust for a specific robotic application. Finally, we highlight the importance of considering non-humans in all of this—animals will encounter robots, by design or accident, and we must ensure that this is trustworthy too.

**Acknowledgments.** This work was supported by the Engineering and Physical Sciences Research Council through the Turing AI World Leading Researcher Fellowship in Somabotics: Creatively Embodying Artificial Intelligence [grant number APP22478], UKRI Trustworthy Autonomous Systems Hub [EP/V00784X/1], Responsible AI UK [EP/Y009800/1], EPSRC Centre for Doctoral Training in Horizon [EP/S023305/1].

**Disclosure of Interests.** The authors have no competing interests to declare.

**Data access statement** The data that support the findings of this study may be available on request from the corresponding author.

## References

1. Abdai, J., Uccheddu, S., Gácsi, M., Miklósi, Á.: Exploring the advantages of using artificial agents to investigate animacy perception in cats and dogs. Bioinspiration and Biomimetics **17**(6) (2022). https://doi.org/10.1088/1748-3190/ac93d9
2. Anzabi, N., Umemuro, H.: Effect of different listening behaviors of social robots on perceived trust in human-robot interactions. International Journal of Social Robotics **15**(6), 931–951 (2023)
3. Benford, S., Garrett, R., Li, C., Tennent, P., Núñez Pacheco, C., Kucukyilmaz, A., Tsaknaki, V., Höök, K., Caleb-Solly, P., Marshall, J., Schneiders, E., Popova, K., Afana, J.: Tangles: Unpacking extended collision experiences with soma trajectories (Mar 2025). https://doi.org/10.1145/3723875, just Accepted
4. Benford, S., Garrett, R., Schneiders, E., Tennent, P., Chamberlain, A., Avila, J., Brundell, P., Castle-Green, S.: How artists improvise and provoke robotics. In: International Conference on Social Robotics. pp. 66–77. Springer (2024)
5. Benford, S., Giannachi, G.: Performing mixed reality. MIT press (2011)
6. Benford, S., Greenhalgh, C., Crabtree, A., Flintham, M., Walker, B., Marshall, J., Koleva, B., Rennick Egglestone, S., Giannachi, G., Adams, M., et al.: Performance-led research in the wild. ACM Transactions on Computer-Human Interaction (TOCHI) **20**(3), 1–22 (2013)
7. Benford, S.D., Mancini, C., Chamberlain, A., Schneiders, E., Castle-Green, S.D., Fischer, J.E., Kucukyilmaz, A., Salimbeni, G., Ngo, V.Z.H., Barnard, P., et al.:

Charting ethical tensions in multispecies technology research through beneficiary-epistemology space. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. pp. 1–15 (2024)
8. Cameron, D., Aitken, J.M., Collins, E.C., Boorman, L., Chua, A., Fernando, S., McAree, O., Martinez-Hernandez, U., Law, J.: Framing Factors: The Importance of Context and the Individual in Understanding Trust in Human-Robot Interaction. In: Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS) (2015)
9. De Graaf, M.M., Allouch, S.B., Klamer, T.: Sharing a life with harvey: Exploring the acceptance of and relationship-building with a social robot. Computers in human behavior **43**, 1–14 (2015)
10. Ellis, J.J.: Beyond "doing better": Ordinal rating scales to monitor behavioural indicators of well-being in cats. Animals **12**(21), 2897 (2022)
11. Esterwood, Jr, R.: Three strikes and you are out!: The impacts of multiple human–robot trust violations and repairs on robot trustworthiness. Computers in Human Behavior **142** (2023)
12. Gaver, W.: What should we expect from research through design? In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 937–946 (2012)
13. Giaccardi, E., Redström, J.: Technology and more-than-human design. Design Issues **36**(4), 33–44 (2020)
14. Gompei, T., Umemuro, H.: Factors and development of cognitive and affective trust on social robots. In: Social Robotics: 10th International Conference, ICSR 2018, Qingdao, China, November 28-30, 2018, Proceedings 10. pp. 45–54. Springer (2018)
15. Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y., De Visser, E.J., Parasuraman, R.: A meta-analysis of factors affecting trust in human-robot interaction. Human Factors **53**(5), 517–527 (2011). https://doi.org/10.1177/0018720811417254
16. Hou, Y.T.Y., Cheon, E.J., Jung, M.F.: Power in Human-Robot Interaction. In: ACM/IEEE International Conference on Human-Robot Interaction. pp. 269–282 (2024). https://doi.org/10.1145/3610977.3634949
17. Khavas, Z.R., Ahmadzadeh, S.R., Robinette, P.: Modeling trust in human-robot interaction: A survey. In: International conference on social robotics. pp. 529–541. Springer (2020)
18. Kok, B.C., Soh, H.: Trust in robots: Challenges and opportunities. Current Robotics Reports **1**(4), 297–309 (2020)
19. Lin, P., Abney, K., Bekey, G.: Robot ethics: Mapping the issues for a mechanized world. Artificial Intelligence **175**(5-6), 942–949 (2011). https://doi.org/10.1016/j.artint.2010.11.026, http://dx.doi.org/10.1016/j.artint.2010.11.026
20. Mancini, C., Nannoni, E.: Relevance, impartiality, welfare and consent: principles of an animal-centered research ethics. Frontiers in Animal Science **3**, 800186 (2022)
21. Manteuffel, C., Dirksen, N., Hartwig, T.: From extra to actor: Facilitating automated conditioning in animal-robot interaction. Computers and Electronics in Agriculture **191**(June), 106496 (2021). https://doi.org/10.1016/j.compag.2021.106496, https://doi.org/10.1016/j.compag.2021.106496
22. Paradeda, R.B., Hashemian, M., Rodrigues, R.A., Paiva, A.: How facial expressions and small talk may influence trust in a robot. In: Social Robotics: 8th International Conference, ICSR 2016, Kansas City, MO, USA, November 1-3, 2016 Proceedings 8. pp. 169–178. Springer (2016)

23. Rodriguez-Guerra, D., Sorrosal, G., Cabanes, I., Calleja, C.: Human-Robot Interaction Review: Challenges and Solutions for Modern Industrial Environments. IEEE Access **9**, 108557–108578 (2021). https://doi.org/10.1109/ACCESS.2021.3099287
24. Romano, D., Porfiri, M., Zahadat, P., Schmickl, T.: Animal-robot interaction—an emerging field at the intersection of biology and robotics. Bioinspiration and Biomimetics **19**(2) (2024). https://doi.org/10.1088/1748-3190/ad2086
25. Rousseau, Sitkin, B..C.: Not so different after all: A cross-discipline view of trust. Academy of Management Review **23**, 393–404 (1998)
26. Salem, M., Lakatos, G., Amirabdollahian, F., Dautenhahn, K.: Towards safe and trustworthy social robots: ethical challenges and practical issues. In: Social Robotics: 7th International Conference, ICSR 2015, Paris, France, October 26-30, 2015, Proceedings 7. pp. 584–593. Springer (2015)
27. Schneiders, E., Benford, S., Chamberlain, A., Mancini, C., Castle-Green, S., Ngo, V., Row Farr, J., Adams, M., Tandavanitj, N., Fischer, J.: Designing multispecies worlds for robots, cats, and humans. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. pp. 1–16 (2024)
28. Shidujaman, M., Mi, H., Jamal, L.: "i trust you more": A behavioral greeting gesture study on social robots for recommendation tasks. In: 2020 International Conference on Image Processing and Robotics (ICIP). pp. 1–5. IEEE (2020)
29. Stahl, B.C.: Embedding responsibility in intelligent systems: from ai ethics to responsible ai ecosystems. Scientific Reports **13**(1), 7586 (2023)
30. Stower, R., Gautier, A., Wozniak, M., Jensfelt, P., Tumova, J., Leite, I.: Take a chance on me: How robot performance and risk behaviour affects trust and risk-taking. In: Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction. p. 391–399. HRI '25, IEEE Press (2025)
31. Sweeney, P.: Trusting social robots. AI and Ethics **3**(2), 419–426 (2023)
32. Tewari, M., Lindgren, H.: Expecting, understanding, relating, and interacting-older, middle-aged and younger adults' perspectives on breakdown situations in human–robot dialogues. Frontiers in Robotics and AI **9**(October), 1–27 (2022). https://doi.org/10.3389/frobt.2022.956709
33. Tollon, F., Vallor, S.: The Responsible AI Ecosystem : A BRAID Landscape Study. Tech. Rep. June, BRAID: Building Responsible AI Divides (2025)
34. Wakkary, R.: Things we could design: For more than human-centered worlds. MIT press (2021)
35. Wilson, J.R., Sharples, S.: Methods in the Understanding of Human Factors. In: Evaluation of Human Work, pp. 1–32. CRC Press, 4 edn. (2015)
36. Zimmerman, J., Stolterman, E., Forlizzi, J.: An analysis and critique of research through design: towards a formalization of a research approach. In: proceedings of the 8th ACM conference on designing interactive systems. pp. 310–319 (2010)