TAS for Cats: An Artist-led Exploration of Trustworthy Autonomous Systems for Companion Animals

Eike Schneiders
Alan Chamberlain
Joel E. Fischer
Steve Benford
Simon Castle-Green
Victor Ngo
Ayse Kucukyilmaz
Pepita Barnard
University of Nottingham
Nottingham, United Kingdom
firstname.lastname@nottingham.ac.uk

Ju Row Farr
Matt Adams
Nick Tandavanitj
Blast Theory
Brighton, United Kingdom
firstname@blasttheory.co.uk

Kate Devlin
King's College London
London, United Kingdom
kate.devlin@kcl.ac.uk

Clara Mancini
The Open University
Milton Keynes, United Kingdom
clara.mancini@open.ac.uk

Daniel S. Mills
University of Lincoln
Lincoln, United Kingdom
dmills@lincoln.ac.uk

Figure 1: View of the environment with the Kinova robot at the centre and the end-effector attachment rack in the front.

ABSTRACT

Cat Royale is an artist-led exploration of trustworthy autonomous systems (TAS) created by the TAS Hub's creative ambassadors Blast Theory. A small community of cats inhabits a purpose built 'cat utopia' at the centre of which a robot arm tries to enrich their lives by playing with them. We initially present the design of Cat Royale as an autonomous system, but then reflect on how diverse human (and animal) stakeholders were required to enable its development and live operation when adopting an approach shaped by responsible research and innovation (RRI). In so doing, we unpack how Cat Royale speaks to three core issues of autonomy, trustworthiness and responsibility relevant to the TAS community more broadly.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; • **Applied computing** → **Performing arts**; • **Computer systems organization** → *Robotics*.

KEYWORDS

multi-stakeholder, art as research platform, participatory design

ACM Reference Format:
Eike Schneiders, Alan Chamberlain, Joel E. Fischer, Steve Benford, Simon Castle-Green, Victor Ngo, Ayse Kucukyilmaz, Pepita Barnard, Ju Row Farr, Matt Adams, Nick Tandavanitj, Kate Devlin, Clara Mancini, and Daniel S. Mills. 2018. TAS for Cats: An Artist-led Exploration of Trustworthy Autonomous Systems for Companion Animals. In *TAS '23: First International Symposium on Trustworthy Autonomous Systems, July 11–12, 2023, Edinburgh, UK*. ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

As autonomous systems continue to spread into our daily lives so we need to engage evermore diverse stakeholders through responsible research and innovation (RRI) [6]. One unusual and interesting class of stakeholder is companion animals - the pets who share our homes and who are already encountering autonomous technologies such as automated feeders, cat flaps and toys. How might future autonomous systems such as domestic robots be designed to benefit pets and their owners?

The artists Blast Theory, who are well known for creating works that explore the societal implications of emerging technologies and who are also TAS's Creative Ambassadors, have set out to explore these questions by creating a new artwork called Cat Royale. In their own words,

> "Cat Royale is an upcoming work exploring the impact of AI on animals. Cats will live inside a utopia created by the artists. The cats' every need will be catered for, with spaces for playing and socialising, and spaces that are relaxing and private. They will have walkways, cubbyholes and raised viewing platforms. And at the centre of the room will be a robot arm and a computer vision system offering activities to make the cats happier. The system may throw a ball, dangle a feather or offer a massage."[1]

Cat Royale is a research project as well as an artwork. Following the method of Performance-led Research in the Wild [2], it involves researchers collaborating with artists to help them make and tour an interactive artwork while also studying the artists' design process and the audience's experience as both grapple with emerging technologies and their implications. There are three broad motivations for engaging artists in such a practice-led approach: their creativity often inspires innovative ways of applying new technologies; installations and performances provide a powerful and relatively safe 'lab' for engaging the public; and the creative industries are important economically and socially, and are so worthy of the attention of researchers.

---
[1]https://blasttheory.co.uk/projects/cat-royale

Figure 2: Cats playing with flipping fish toy.

Artists are already embracing AI-based generative tools for language[2] and visual art[3], raising questions about to what extent AI can be creative and its impact on the human artists whose styles it appears to copy. However, Blast Theory are working with robotics to explore what, at first sight, would appear to be a beguilingly simple idea — that these technologies might help deliver a 'cat utopia', including robots playing with cats — but that on further exploration raises a host of complex questions.

(1) Do we trust an autonomous system to make decisions potentially affecting the well-being of living creatures?
(2) Can we understand the rationale behind the decisions of the embodied AI, i.e., are the actions carried out by the AI relatable?
(3) How do we situate an autonomous system within a wider environment, which is simultaneously engaging for spectators, ensures cat well-being, and is suitable for the autonomous system to operate in?

In what follows, we initially introduce Cat Royale as a technical system, but then expand this given our approach shaped by RRI into a *socio-technical system* that involves diverse stakeholders in its design and operation. We then reflect on how, given the RRI approach, the project speaks to the three core TAS concerns of autonomy, trustworthiness and responsibility.

2 CAT ROYALE AS AN AUTONOMOUS SYSTEM

We begin by considering the *idealised* design of the autonomous system that is Cat Royale, by which we mean the collection of technologies that would ultimately be able to autonomously play with a small community of cats. At the heart of the system is a Kinova Gen3 Lite ultra-lightweight robot, a small robot arm with inherent safety-by-design through low speed (25 cm/s) and payload

---
[2]https://chat.openai.com
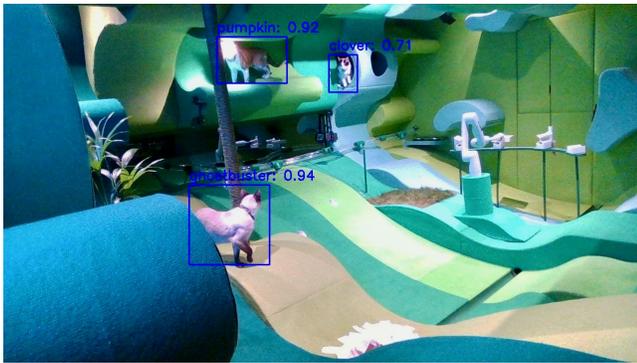[3]https://openai.com/dall-e-2/

Figure 3: Identification of the cats in the environment via the Computer Vision system.

(0.5 kg) limits to work near humans (see Figure 1). The system has various built-in safety features, such as velocity limits, configurable safety thresholds, and an emergency shut-down switch.

In our design, the robot does not directly physically interact with the cats, but deploys a series of toys designed to be safe and engaging (Figure 2). It can, for example, wave various feather and boa toys, drop balls, ring small bells, offer massages, and dispense treats to eat from a tray. These attachments can be stored on four magnetic racks (raised to prevent cats from accessing them) allowing the robot to select them as required. The artists have pre-programmed the robot to conduct specific movements of each toy (typically for a few minutes) in a way that should invite a playful response from the cats.

The cats, robot and racks are housed in a bespoke physical enclosure designed for the safety and comfort of the cats. There are dens to which they can retire, high perches to observe from, raised walkways, a water fountain, and litter trays. All aspects of their care other than play remain the responsibility of humans who observe through one-way mirrors. The cats spend three hours at a time in this environment, being offered a new play task every ten minutes.

Eight iPhones are mounted inside the environment to film the cats from different angles. This material is then edited into a film or live stream, to be shown to public audiences in galleries, with additional captioning to give the artists' interpretation of events. Consequently, cats are not shown live and do not need to be physically present in galleries for the public to engage with the work.

The system involves three key software components. The *robot control system* is used to compose and play back sequences of movements. The *computer vision system* is trained to recognise individual cats, estimate their position in the environment (via bounding boxes around the cats, see Figure 3), and to classify whether they are currently still/sleeping, engaged (attentive), playing, moving grooming, relaxing or feeding. Finally the *decision engine* decides which tasks to offer the different cats at a given time to increase their 'happiness' score. Given the system's raison d'être is to play with cats, it estimates their happiness using an established Participation in Play scale [3].

3 CAT ROYALE AS A SOCIOTECHNICAL SYSTEM

While Cat Royale's idealised version may be an autonomous system, in adopting a responsible approach to the research and innovation (i.e., RRI) needed to deliver Cat Royale, it is in practice necessarily a socio-technical system involving an array of human stakeholders in its design and development. This section will introduce the stakeholders first before articulating the major themes that arise when adopting a responsible approach.

3.1 Human Stakeholders

The core team comprising the Artists and various Computer Scientists (AI, robotics and HCI) established at the outset rapidly expanded to include other stakeholders bringing in new perspectives.

3.1.1 Animal Welfare Specialists. To ensure animal safety and welfare, which was the highest priority throughout the entire project, the research team included researchers specialising in feline behaviour from Veterinary Science and also Animal-Computer Interaction (ACI) as well as professionals from the RSPCA. As expected, the main concern for these stakeholders was animal welfare. They provided valuable input related to topics such as the physical layout of the environment, the robot's role in satisfying basic needs of the cat (e.g., food/water, sleep, or privacy), and direct animal consent. Previous research [5] and input from the animal welfare specialists, greatly informed the design of the environment to ensure that the cats had the opportunity to withdraw from interactions with the robotic arm. Furthermore, these stakeholders shaped the environment by influencing decisions on topics such as environment morphology and colour.

3.1.2 Audience Advisory Panel (AAP). To support the artists in creating an engaging installation from an audience point of view, the design process of the Cat Royale project involved the formation of an audience advisory panel (AAP). This panel was formed of 16 members of the general public, who represented the potential future audience of the Cat Royale installation. These participants were selected from 80 applicants with the goal of getting a diverse group from the UK and the US. The AAP includes cat, art and AI enthusiasts—as well as participants without strong opinions on either—of various ages. The AAP was invited to virtual meetings in approximately 3 week intervals for half a year before the exhibit went live. The AAP's opinions were sought on various topics and questions including, but not limited to: *What is, and how do you even assess, cat happiness? What does an environment need to be considered a 'Cat Utopia'? How does the robot know what to do? Can you trust a robot to take care of cats? How can the audience be informed about the rationale the robot bases its decisions on?* Engagement with the AAP also directly informed the installation, in particular, the production of outward-facing information. This included aspects such as the presentation of the rationale for each robot action and informing on what basis a given decision was made.

3.1.3 Ethical review boards. The design was shaped by three different ethical review boards over the course of nine months, with each board requiring refinements before granting approval. The University of Nottingham's Animal Welfare and Ethical Review

Body (AWERB) governs the use of animals in scientific research (e.g. human medical research), while the Veterinary School's own board focuses more on research intended to benefit animals through veterinary science. Both encouraged the deep involvement of animal experts in the project and the development of detailed protocols to ensure the cat's welfare. The Computer Science committee, on the other hand, focused on the ethics of studying the human participants, including the audience, with a particular focus on the ethical treatment of study data.

3.1.4 Operations team. The involvement of many of these stakeholders continued beyond what would traditionally be seen as design and development into supporting live deployment. In order to operate the autonomous system required the support of human roles including an overall director, robot wrangler, toys and attachment wrangler, live video mixer, and full-time cat welfare officer, who worked in an adjacent control area to ensure the the system was operating safely and appropriately.

3.2 Major themes from an RRI perspective

Here, we consider how Cat Royale relates to three major themes from an RRI perspective - the themes of autonomy, trustworthiness and responsibility – considering the distinct perspectives of various stakeholders on each of these.

3.2.1 Autonomy. Autonomy can be considered from the perspective of several stakeholders. Ideally, the robot would operate with a high degree of autonomy, employing computer vision to observe the cats, reasoning about which play activities are appropriate at any given time, and driving the robot to deliver these. In practice, however, its autonomy is necessarily limited. Its governance protocol requires that, for reasons of safety, humans must approve and monitor actions whilst holding a 'dead man's switch'. However, at the early stages of design, humans are required to integrate – and sometimes partially enact – elements of the system itself, for example: assigning overall happiness scores, improving new play sequences, or acting as the computer vision system while it is being trained on early footage of the cats interacting in the actual environment. The cats are also key stakeholders whose autonomy is clearly impacted. Animal welfare experts stress the importance of maximising their agency and autonomy. In response, the system is designed so that they are free to choose to engage with the robot or to withdraw to dens, perches and other safe spaces within the enclosure. However, for reasons of potential distraction and safety, they are not generally free to pass in and out of the enclosure itself. This loss of autonomy is hopefully mitigated by increasing their agency to make choices when in the environment through the provision of multiple dens, perches, litter trays, walkways and other features. Interestingly, the public audience also enjoys limited autonomy. The cats' experiences are edited into an eight-hour long film that plays out over the course of a day in a gallery space. This is not interactive, so visitors will typically only see the small portion of the material that happens to be playing when they are physically present.

3.2.2 Trustworthiness. The question of trustworthiness also involves consideration of the various stakeholders involved. The most fundamental perspective is whether various humans – especially the artists, owner, cat welfare officer, and audience – trust the system to ensure the safety and welfare of the cats and to enrich their lives. A second is whether we trust that cats will not damage the robot. A third is whether the artists trust the overall set up to deliver an engaging experience for the audience; what, for example, if the cats never played with the robot? They would be safe, but the experience would fail to deliver as either an artwork or a research inquiry. Establishing these various forms of trustworthiness relies on a combination of system design, orchestration and governance. System design factors include the choice of the robot arm (small size and light strength); interaction via soft and flexible attachments and toys rather than directly; careful programming of robotic movements; ensuring the vision system can reliably track cats; and ensuring the decision engine delivers appropriate decisions. Notably, system design extends to designing the surrounding enclosure to provide a safe, stress-free and ideally attractive environment for the cats to engage the robot alongside appropriate food, drink and litter trays while also protecting them from outside dangers. However, the system is far more than a collection of technical components; it is, of course, a socio-technical system that also involves a large team of human orchestrators who constantly monitor it and intervene when necessary. These include an overall director, robot operator, video director, cat welfare operator, and toy wrangler, each of whom follows carefully designed and rehearsed procedures (including a lengthy Cat Welfare Protocol). The technologies must be designed to enable these humans to monitor and intervene, including through software interfaces for triggering and even improvising robot moves and for monitoring the behaviours of the vision system and decision engine and hardware interfaces such as various one-way mirrors looking into the environment complemented by hatchways for accessing the robot and litter trays. Processes supporting trustworthiness also include a five day period of habituation during which the cats are gradually and carefully introduced to the studio, enclosure and robot. Trustworthiness also extends to the governance of the system which in our case involved an extensive ethical review process that passed through three separate ethical boards. As mentioned above, Nottingham's Animal Welfare and Ethical Review Body (AWERB) and School of Veterinary Science addressed the cat welfare aspects of governance, while the Computer Science Research Ethics Committee focused on the involvement of human-study participants.

3.2.3 Responsibility. One important aspect of responsible innovation is the involvement of diverse stakeholders throughout the design process. As discussed, a feature of Cat Royale is the range of different stakeholders, together sharing responsibility, navigating the design space, and sharing their expertise to provide a space and activity for animals that is safe, reliable, trustworthy, and enriches behaviour.

A key question is to what extent the cats themselves can be considered to be participants in a design process? Pushing the field of Animal-Computer Interaction (ACI), Mancini's manifesto on ACI [1, 4] puts a greater emphasis on how we, as a research community, should conduct research involving animals. The goal of ACI should always be aimed at improving the lives of the given species and individual. Furthermore, Mancini [4] highlights the

importance of ethical research to the same extent as we would conduct research within HCI. This includes, amongst others, the need to treat animals as equals to humans, as they are participants in our research who deserve the same consideration, respect, and care. This potentially extends to the need to receive consent from the animals. The consent from animals [5] is specified as having two complementary forms: mediated consent as provided by those who know the animals well, are responsible for their well-being and have the authority to make decisions on their behalf (the cats owner in the case of Cat Royale) and contingent consent, provided by the animals themselves based on their ongoing assessment of research set-ups and expressed by their chosen modality of engagement or withdrawal (as primarily assessed by the Cat Welfare Officer in Cat Royale). Our intention is to engage the cats in hundreds of playtests with our robot to enable us to make sense of their input into the design. The ethical review process might also be viewed as a form of co-design. While it is normal to gain ethical approval early on in the experimental processes, this can be more difficult in more exploratory artist- and design- processes where many design details are unknown until late in the day. In our case, ethical review unfolded iteratively over a period of nine months, successively introducing different perspectives into an ongoing design conversation.

A more general reflection on responsibility concerns the nature of public involvement. Cat Royale is an unusual project designed to stimulate the public to reflect on the societal implications of TAS, especially systems that might potentially care for those we love. It's beguilingly simple and somewhat ambiguous and immediately raises important questions. What would a TAS-driven utopia be like? How can a system know whether cats (or, indeed, humans) are happy? And what is happiness for cats (and, again, humans)? Moreover, it requires significant effort and a large team of humans to design and support a robot to play with cats, revealing the human challenges of designing and delivering AI in practice, which is indicative of the behind-the-scenes human labour often involved in delivering automated results. Framing such projects is a difficult challenge that relies on artistic judgement and experience: how can

we provoke questions about AI while simultaneously being reassuring about the cats' welfare, for example? At the time of writing, we are only able to report on the design and implementation of Cat Royale. We look forward to exploring post-project as to how it has shaped stakeholders' thinking about the responsible design of trustworthy autonomous systems in the future.

4 CONCLUSION

In this short piece we intended to demonstrate how an Animal-Robot system that at first glance may seem simplistic becomes a complex socio-technical system as it is shaped in practice by various stakeholders when adopting a responsible research and innovation (RRI) approach. In so doing, we unpacked the three major themes of autonomy, trustworthiness and responsibility that are pertinent and could be instructive to the TAS community more broadly.

ACKNOWLEDGMENTS

REFERENCES

[1] Devin Anne Nieusma. 2022. Into the Wild: Into the Wild: Opportunities and Challenges for Animal-Computer Interaction in Wildlife Conservation. In *Eight International Conference on Animal-Computer Interaction* (Bloomington, IN, USA) *(ACI'21)*. Association for Computing Machinery, New York, NY, USA, Article 14, 7 pages. https://doi.org/10.1145/3493842.3493904

[2] Steve Benford, Chris Greenhalgh, Andy Crabtree, Martin Flintham, Brendan Walker, Joe Marshall, Boriana Koleva, Stefan Rennick Egglestone, Gabriella Giannachi, Matt Adams, Nick Tandavanitj, and Ju Row Farr. 2013. Performance-Led Research in the Wild. *ACM Trans. Comput.-Hum. Interact.* 20, 3, Article 14 (jul 2013), 22 pages. https://doi.org/10.1145/2491500.2491502

[3] Jacklyn J. Ellis. 2022. Beyond Doing Better: Ordinal Rating Scales to Monitor Behavioural Indicators of Well-Being in Cats. *Animals* 12, 21 (2022). https://doi.org/10.3390/ani12212897

[4] Clara Mancini. 2011. Animal-Computer Interaction: A Manifesto. *Interactions* 18, 4 (jul 2011), 69–73. https://doi.org/10.1145/1978822.1978836

[5] Clara Mancini and Eleonora Nannoni. 2022. Relevance, Impartiality, Welfare and Consent: Principles of an Animal-Centered Research Ethics. *Frontiers in Animal Science* 3 (April 2022). http://oro.open.ac.uk/83128/

[6] Jack Stilgoe, Richard Owen, and Phil Macnaghten. 2013. Developing a framework for responsible innovation. *Research Policy* 42, 9 (2013), 1568–1580. https://doi.org/10.1016/j.respol.2013.05.008