Project: CReAting a Dynamic archive of responsibLe AI Ecosystems in the context of Creative AI(CRADLE)

AHRC/UKRI BRAID Programme 2024

Output 4: Structuring a Dynamic Archive of AI ecosystems in the context of the creative industries[1]

Funders — Bridging Responsible Artificial Intelligence Divides (BRAID) – and Arts and Humanities Research Council (AHRC).

[1] To cite this output use: Farina, L., Webb, H., Giannachi, G., Moore, J., Benford, S., Jordan, S., Stahl, B., Perez-Vallejos, E., Vear, C., Drury, M.R.F., Miles, O., Brundell, P., 2024, Structuring a dynamic archive of AI ecosystems in the context of the creative industries, BRAID/ AHRC OPP18206, 4.

Authors Lydia Farina, Helena Webb, Gabriella Giannachi, Steve Benford, John Moore, Bernd Stahl, Elvira Perez Vallejos, Spencer Jordan, Craig Vear, Megan R. F. Drury, Oliver Miles, Pat Brundell

§1. Introduction

§1.1. Overview

In this document we sketch the structure of a dynamic archive based on an analysis of 7 recent research projects focusing on the use of AI and advanced technology applications in the creative industries.[2] To ensure that a dynamic archive is structured in a way which aligns with RAI UK priorities, it is necessary to provide accurate models of the AI ecosystem the archive is an expression of. To this aim, in section 2 we give a brief discussion of each research project analysed in this scoping project focusing on the AI ecosystems modelled by each individual case study. This enables us to make more visible the different stakeholders relevant in each ecosystem. In turn, this will enable us to approach responsibility questions arising within each AI ecosystem which should be reflected in the structure of each cast study archive.

We use a Diagrammatic Ecology (DE) approach as analysed in the workshop dedicated to Cat Royale (see Workshop report 2). Diagrammatic ecologies (DE) are visual representations of the situated contexts of the research projects; they consist of a series of nested concentric circles. The general structure for each DE moves outwards from *the project* located in the core, to the *public context*, and ends with the *governance context*. As we show in the following section, applying this model to our case studies allows for individual differences embedded in each project. The particularities of each project's DE are discussed in the individual sections below. Though DEs are exceptionally useful for mapping the stakeholder relations of each project they also have limitations since they cannot be used to visually represent the temporality of these projects and their iterability in a clear way. As we argue in section 3, these two features need to be reflected in the structure of a dynamic archive and as such, the DE module needs to be supplemented so that it includes some of the overarching priorities of the structure. These are considered necessary but not sufficient conditions so that the structure aligns with the requirements of RAI UK for responsible AI applications.[3] We consider these features to be necessarily reflected in the structure of the dynamic archive so that it ensures responsible use of AI in this context.

[2] In this document we interpret AI applications in a broad sense as applications that attempt to do the sort of things human and non-human animals can do (Frankish and Ramsey 2014)

[3] Borrowing from RAI UK, we take these applications to be responsible when they are both trustworthy in principle, and trusted in practice by individuals, society and government. In so doing, we are committing to the following guiding principles as stated by RAI UK: cooperation, multidisciplinary, commitment to excellence, accountability and anticipation of consequences, fairness, transparency and commitment to public benefit. These will guide everything the ecosystem does, including its internal management, and the research, engagement, and skills programmes. For more detail see https://rai.ac.uk/guiding-principles/

In section 4, we suggest a hybrid structure which can be used as the dynamic archive framework by combining two different approaches used in the Human-Computer Interaction (HCI) literature. This hybrid structure is a work in progress, resulting from bottom-up evidence collected in the 7 research projects we analysed during this scoping project. The aim is that this hybrid is used to address some of the limitations of the DE approach.

§1.2. Preliminaries: Defining a Dynamic Archive

We interpret an archive as ‘a "heterogeneous set" including "virtually anything" under the same heading: "discourses, institutions, buildings, laws, police measures, philosophical propositions, and so on, precisely because the apparatus is the network that is established between these elements..." (Giorgio Agamben, from Foucault, 2009).[1]

As discussed during the project workshops, a dynamic archive should:

- continue to grow over time
- reconstitute itself in accordance with keywords set by whoever wishes to consult it
- reconstitute itself by making visible different time-based versions of itself
- be used curatorially and creatively as a live (generative) archive.

In addition, a dynamic archive can be “referred to as an ‘apparatus’” because it can be operationalised as much as a structure and tool for construction as a set of documented content’. [1]

As mentioned in the previous section, to create a structure for an archive one needs to know the boundaries and stakeholders of the AI ecosystem the archive relates to. We define AI ecosystems as the “digital counterparts of biological ecosystems, exploiting the self-organising properties of biological ecosystems, which are considered to be robust, self-organising and scalable architectures that can automatically solve complex, dynamic problems” [5]. The exercise of modelling an ecosystem is intimately related to structuring the archive relating to that ecosystem. Partly this is because the archive needs to be accessible to all stakeholders whilst also having the ability to reflect differing values between these stakeholders. In addition, the structure of a dynamic archive must also allow for an element of temporality so that the archive can be adapted and reconstituted to reflect both changing values and changing stakeholders. We give more detail on these necessary features in section 3.

§2. Case studies' overviews

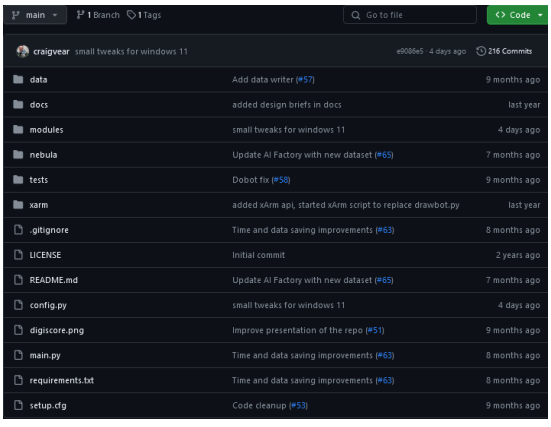
In this section, we review the key themes and features of some of the AI ecosystems arising from our case studies and examine the relationship between models of AI ecosystems and their archives.

§2.1. Jess + and Cat Royale

§2.1.1 Textual Overview

-Jess+

-The *Jess+* project brought together musicians, sensory probes, and an AI-enabled robotic arm, as a novel means of "…build[ing] an embodied-AI system that facilitates co-creation for an improvising ensemble of disabled and non-disabled musicians…"(Jess+). The project was part of the wider European Research Council (ERC) 'digiscore' project, investigating how AI and robotics can transform musicianship and aid people with disabilities. It received further funding from Nottingham University (Faculty of Arts) and Trustworthy Autonomous Systems (TAS) hub. It aims to encourage a responsible approach to performance, composition, and interaction, by means of joint authorship or co-creation between humans and AI applications. It also aims to bring about the breaking down of barriers between composer and performer, and it employs a user-centred, inclusive, and flattened hierarchy approach to physical mobility. *Jess+* affords knowledge of the participants' perceptions of being 'in the loop' in terms of creation of the digital score and constitutes a shared space for creativity.



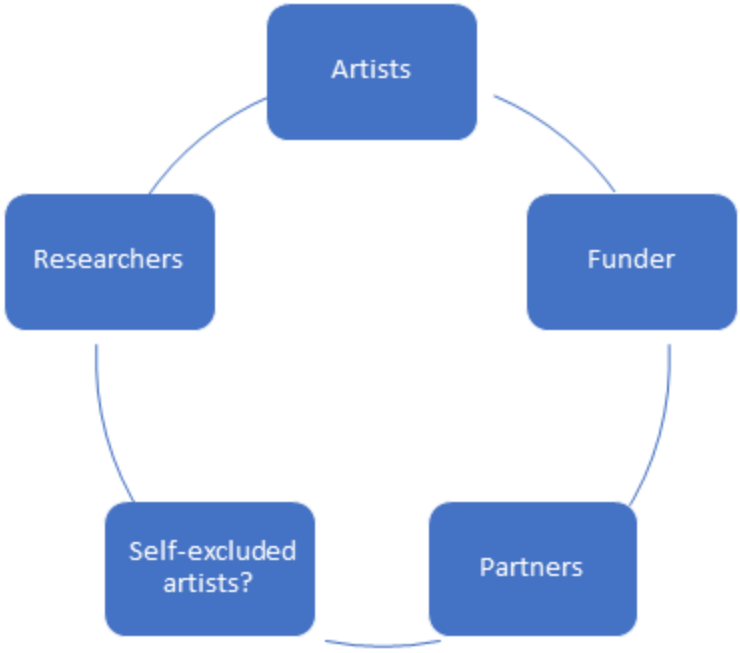
Artefacts of Performance Artefacts of Experience

As an ecosystem, *Jess+* can be described as an attempt to foster an emerging set of principles aiming to:

- Ground a hierarchy of trust.
- Build a sense of togetherness.
- Build a forefront of creativity and blended inclusivity.
- Invite discussion on the creative agency of the robotic arm.

Thinking about what an ecosystem of *Jess+* might look like, during the first workshop of the scoping project, project lead Craig Vear together with Solomiya Moroz and Adrian Hazard (researchers in *Jess+)* reflected that data elicited both gives insight into the practice led experience itself, as well as into the experiential output of the participants. Crucially therefore, the artefacts of *Jess+* as a performance (the instruments, sensory probes, robotic arm) are not the points of focus for this scoping project. Rather, the project should focus on the phenomenological experience, which is captured in a variety of archived sources (e.g., a GitHub repository). We reflected that this becomes nuanced when considering the role of certain components of the original *Jess+* performance, for example the use of a preexisting Jazz dataset. This invoked the “no decision about me, without me” principle.

Sketch of ecosystem stakeholders in *Jess +*:



-Cat Royale

Cat Royale is the product of a collaboration between the University of Nottingham's Mixed Reality Lab (MRL) which is "grounded in the disciplinary field of Human-Computer Interaction" (HCI) and the Brighton-based artist group, Blast Theory (BT), which focuses primarily on "creating groundbreaking new forms of performance and interactive art" that "explores interactivity and the social and political aspects of technology".

The key question underpinning *Cat Royale* was '*can AI make us (humans) happier?*.' By using cats as proxies for human animals, the project wanted to raise questions on whether we can increase happiness using AI applications. Happiness was interpreted as increased engagement with the AI applications. Thus, artists set about building an enclosure designed for the maximalisation of feline comfort and luxury where three cats would spend six hours a day for twelve days. During this time, the cats would interact with a robotic arm that could perform a series of actions determined via an AI system connected to a computer vision system. The arm/AI system would monitor the behaviour of cats (measured against a 'Participation and Play' scale) and respond by learning which activities led to the greatest observable happiness.

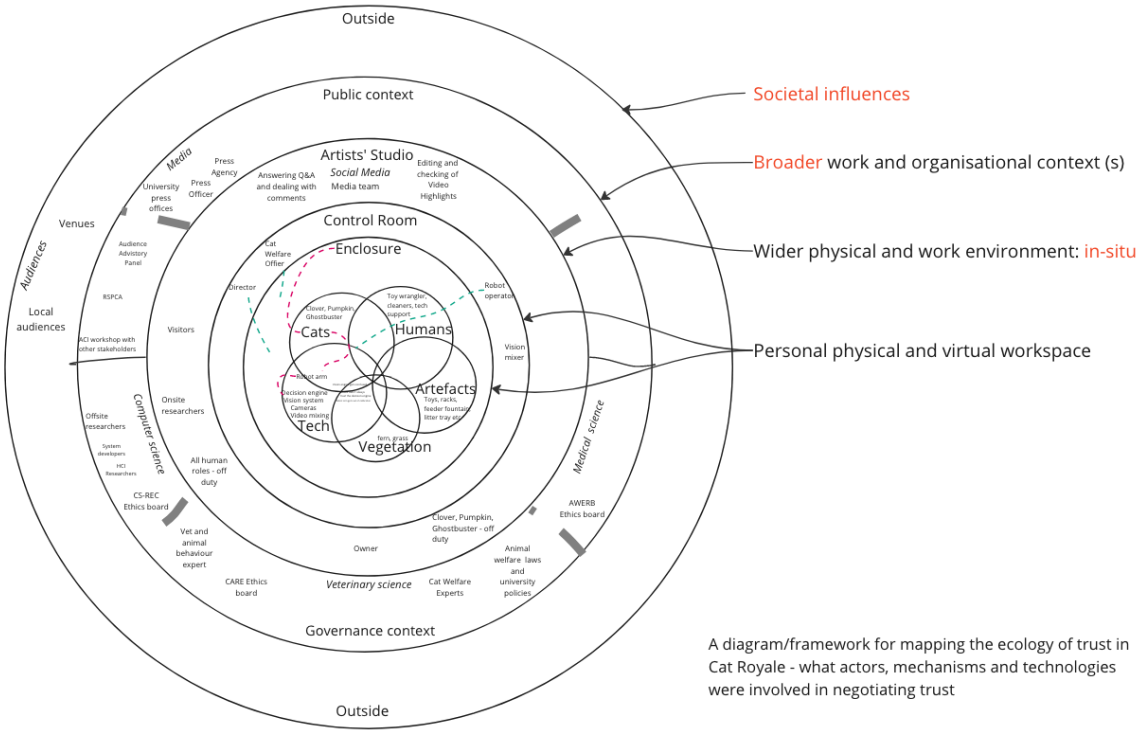
To gain further understanding, we watched three YouTube video excerpts from Blast Theory's promotional work: We include screenshots from videos below.

Intro day 1: *Cat Royale* is introduced as a utopia for cats. In this utopia, a robot arm is central to interactions in the system. The overarching goal is to 'increase and maintain the cat's happiness. The cats, – Ghostbuster, Pumpkin, and Clover –, are unique actors with their own stories and characteristics. Artefacts included thirty-six toys and games for the cats to interact with. Questions revealed a concern for Responsible Research and Innovation (RRI) such as 'is it best for the cats if the AI learns they love snacks, and acts accordingly by offering more,' or 'should the human operators intervene?'

Highlights: The highlight video from day ten drew attention to the number of appearances of artefacts as meaningful touchpoints for narrating interaction. This evoked questions such as 'are the cats aware of what they are doing?' And 'are their interactions significant beyond the immediate and obvious?'

End video: The final video summarised the filming of *Cat Royale* and included final measures, such as the provision of a game 'rating' by the AI. This provoked questions such as 'has the AI or the cat led the interaction in terms of provision of favourite toy?'. In total, 7500 video clips were tagged from eight iPhones positioned around the room. The robot operator controlled the arm, and automation was continually tweaked and altered. Anecdotally, further contextual data was captured here. For instance, Clover and Ghostbuster were observed to 'queue to get in' the enclosure. In summary, *Cat Royale* included many different perspectives, such as from vets and animal behavioural psychologists. In terms of the artist's perspective, the analogy was that the cats in *Cat Royale* represent the human user. The enduring thought was, 'what are the costs of instant gratification'?

Cat Royale maps this AI ecosystem as an example of ecologies of trust; a framework consisting of concentric circles (see Figure 2 below) in which the original performance is depicted as central, moving out through layers of ecological zones and diverse stakeholders. *Cat Royale* depended on a large number of human and non-human animals, and artificial agents operating successfully. By mapping the entities and environments in this manner it is possible to identify main stakeholders and relevant 'trust trajectories' (i.e., the circuits of trust firing within the ecosystem between stakeholders).



A diagram/framework for mapping the ecology of trust in Cat Royale - what actors, mechanisms and technologies were involved in negotiating trust

The resource became central to discussing what needs to be in place so that interactions are maintained. This cartographic approach to responsibility enables the identification of key stakeholders or foci of responsibility at different levels or layers, working from the nucleus outwards.

Figure 2: Ecology of Trust in Cat Royale

Explanation of the different levels:

The Enclosure

The inmost sphere – The Enclosure – is the environment of the primary stakeholders (i.e., the cats, humans, artefacts, vegetation, and technologies). Specifically, other than Ghostbuster, Clover, and Pumpkin, the robot arm was pre-programmed to move and manipulate toys, with eight iPhones capturing interaction inside of this layer. The Enclosure was the space that could be affected by the actions taken from the next layer, the Control Room, which housed the director, welfare officer, robot operator and vision mixer.

The Control Room

In The Control Room, human operators have a responsibility to ensure the effective management and safety of those within The Enclosure. This included the robot arm operator, the cat behaviour specialist, vets, computer scientists, and director. The systems in place

allowed for the monitoring and overriding of the AI decision engine and the operation of the robot arm (e.g., to prevent over-feeding, underfeeding, feline stress) which were sometimes guided by the decisions and recommendations of the cat behaviour specialist (measuring feline scores against the play participation scale).

Tasks included the completion of a cat stress score (participation in play scale), yielding a measure of 'happiness' which could be fed into a decision engine resulting in recommendations. There were two modes for the algorithm; random and exploitation (using things learned by the algorithm). Critical questions include 'who should handle and hold executive decisions over recommendations?', and 'what objectively constitutes too stressed/too relaxed'?

The first showing of *Cat Royale* was at the Brisbane festival (Blast Theory, n.d.) which was the closest to the live performance (24 hours delayed). We reflected on this latency, both from the perspective of the audience, and from the perspective of any stakeholder who might wish to analyse the original performance.

Artist's Studio

In terms of the Artist's Studio, various stakeholders were present including visitors, the onsite researchers, the owner, the cats, and all off-duty human roles. This space is important because many different agendas were operating in tandem to innovate in a responsible manner. Practically, a meeting took place in this space once a day to discuss welfare, daily agendas and what – if anything – needs to change. This element was not within public view and represented a 'boundary within boundaries'.

The Studio was used to reunite the three cats with their owner. It also allowed space for ethnographers and other researchers to interact with the project team. Social media teams also present in the Studio managed the public face of the project.

Public Context

The public context represents the first domain located completely 'outside' of the *Cat Royale* enclosure. Stakeholders encompassed third-party actors such as social media and overseas audiences. This space crucially includes the gallery and the theatre and audiences of current exhibition.

In terms of the governance context, there were three components: Discussion in computer science, in the veterinary school, and the cross-university animal committee. This interaction led to a positive, constructive process. However, and with regards to creating a dynamic archive, there is a question of where this process sits. Clara Mancini's (2011; 2017) work regarding animal computer interaction (ACI) and the distinction between *contingent*

and mediated consent is of relevance here [1]. Contingent and mediated consent is a concept denoting a model for garnering consent in the context of human-animal-computer interaction research studies. The *mediated* dimension of this concept refers to the way in which consent is given on behalf of an animal (e.g., a companion animal, animals in zoos, or animals in shelters) by their relevant human companion and/or their legal guardian (Mancini 2017, 230). The *contingent* dimension refers to the conceptualisation of the act of giving consent not as a singular event, but a sustained process whereby it is possible for the relevant human mediator to withdraw consent on the behalf of the animal participant at any time (Mancini 2017, 230).

§2.2. When the Future Comes

§2.2.1. Textual Overview

When the Future Comes is a collection of research-based artistic projects, including *Performing the Future, The Future Machine,* and *The Cabinet of Curious Places*. For the purposes of CRADLE's research, *The Future Machine* has been selected for close analysis given its role as "an artistic technology probe" (Jacobs et al., 2023,12). In other words, The Future Machine can be understood as an inchoate or embryonic prototype of a fully-fledged future machine that is dependent on an Artificial Intelligence model (for more on technology probes, see Mattelmäki 2006).

Publicly funded by a series of academic and cultural institutions and consortia[4], *The Future Machine* is best described as a longitudinal study of five geographic locales and the changing perceptions of and attitudes toward the climate crisis (When the Future Comes, n.d.; Jacobs et al., 2023, 1). Situated within the field of human-computer interaction (HCI), the artistic research project aims to break from the standard dichotomy evidenced in the literature between artworks that eschew emotive or phenomenological affects in favour of empirical data, and those that eschew empirical data in favour of emotive and phenomenological affects (Jacobs et al., 2023, 4). This decisive break from what can be standard approaches to artistic explorations of the climate crisis allows *The Future Machine* to emerge as a unique project, both in terms of its rationale and its timescale, since the project is projected to be operational until the year 2050. The '*artistic space*' of the project invites visitors to think and feel beyond "the short-term time frames [humans] normally engage with in [their] lives" with a positive reflection rather than resulting in doom-saying

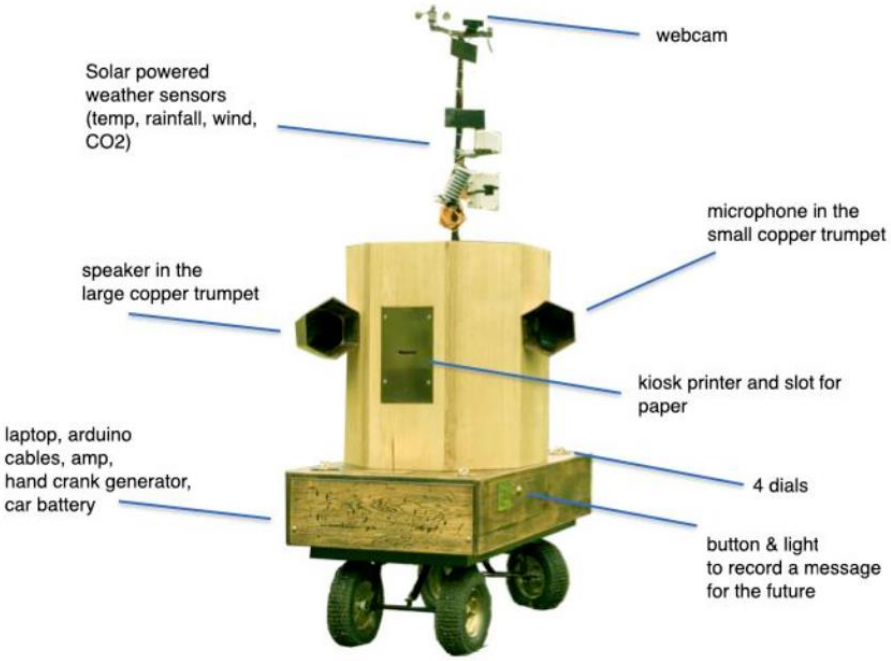
[4] "Arts Council England, Furtherfield, Horizon Digital Economy University of Nottingham (ESPRC/RCUK), the Engineering and Physical Sciences Research council (EPSRC) through the Trustworthy Autonomous Systems Hub (EP/V00784X/1)" (Jacobs et al., 2023, 14).

(Jacobs et al., 2023, 4). *The Future Machine*, therefore, is designed to be an arbiter of hope amidst discourses of existential threat.

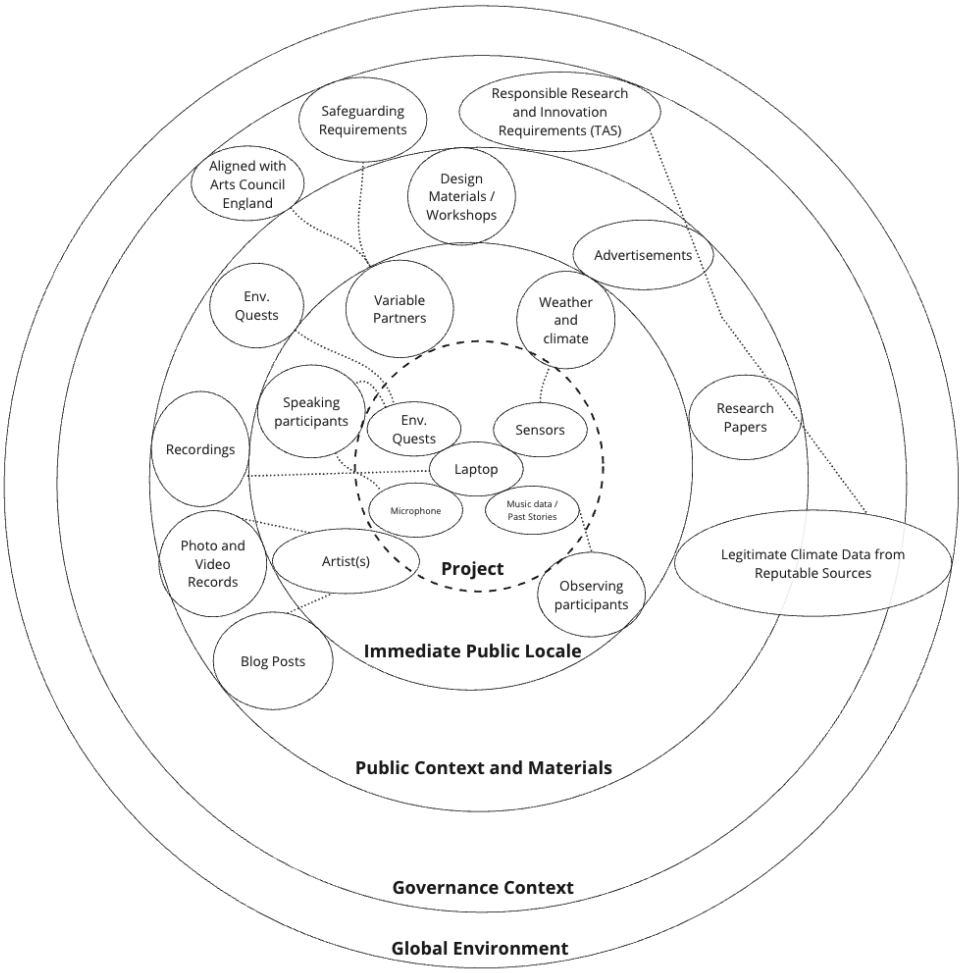
The Future Machine itself, as a material object, is "a large sculptural artwork" that is interactive (Jacobs et al., 2023, 5). Participants are invited to accompany *The Future Machine* on its annual travels to five locations around the United Kingdom[5] as key seasonal thresholds are reached (e.g., the blossoming of spring, the solstices, and the falling of leaves) (Jacobs et al., 2023, 9). The project encourages audiences to interact with and appreciate the rhythms of the natural world and to reflect on and record, by speaking to *The Future Machine*, their hopes and fears, and those of others, for the future of the Earth. It is for this reason that *The Future Machine* is frequently described by the creative team as a *witness*; it is an archive of past stories, and vehicle for the projection of the present into the future.

More than a vocal recording and playback device on wheels, however, *The Future Machine* is also equipped with a webcam that records the horizon and skylines of its environment, solar powered sensors that record the temperature, rainfall, wind, and carbon dioxide levels that collect localised environmental data (Jacobs et al., 2023, 5, 10). This data is both stored for longitudinal analysis as the data accumulates, but also for the production of music generated by the digital system housed within the octagonal ash, oak, and brass exterior from pre-selected layered tracks (Jacobs et al., 2023., 10).

[5] These are: Christ Church Gardens, Nottingham; the River Leven, Cumbria; Rotherfield Peppard, Oxfordshire; Cannington and Kilve, Somerset; and Finsbury Park, London.


webcam
Solar powered
weather sensors
(temp, rainfall, wind,
CO2)
microphone in the
small copper trumpet
speaker in the
large copper trumpet
kiosk printer and slot for
paper
laptop, arduino
cables, amp,
hand crank generator,
car battery
4 dials
button & light
to record a message
for the future

§2.1.2. AI ecosystem on the basis of a Diagrammatic Ecology (Static Archive/Stakeholders)

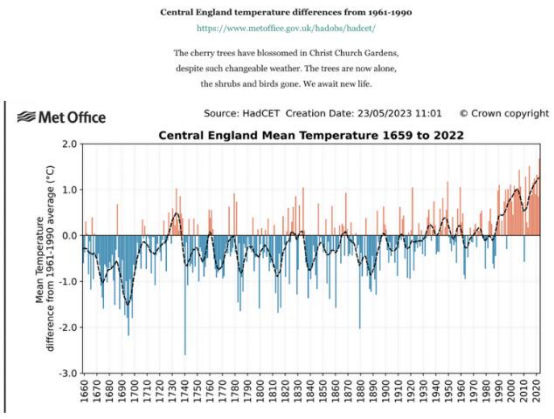


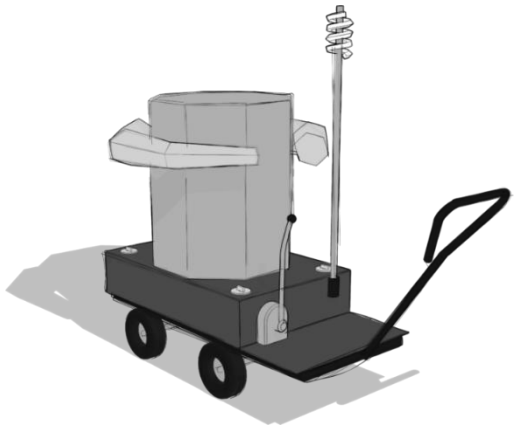
The static archive, or diagrammatic ecology, for The Future Machine (above) is divided into five strata. The innermost nucleus maps the components of The Future Machine which is constituted by a Windows laptop connected to a car battery, powered by a trickle lever, and outputs instructions to the kiosk printer, microphone, and amplifier. The sensors feed environment data into the laptop where the data is stored. These processes, however, are always unfolding within an immediate public locale, hence the decision to mark the boundary of the two innermost strata with a hyphenated line, highlighting the permeability or transience of this distinction.

The second stratum concerns the *immediate locale*, which itself is variable referring to either Christ Church Gardens, the River Leven, Rotherfield Peppard, Cannington and Kilve, or Finsbury Park depending on the time of year that the static archive is attempting to map.

With each immediate locale, there are a series of different stakeholders that may be relevant to each iteration of the project. For example, in Christ Church Gardens, Mellers Primary School partners with the When the Future Comes Collective[6] to facilitate a 'light box' making ritual. In Finsbury Park, Furtherfield Commons hosts *The Future Machine* and other local collectives, such as The Drumming School to facilitate musical performances. These geographically variable partnerships help to produce, at each location, a new ritual for when the future comes. The immediate locale also includes the participants, both those who come to watch *The Future Machine* and engage with the rituals, and those who speak directly to *The Future Machine* and thus to future audiences (and their future selves). This immediate locale is indissociable from the project itself, since *The Future Machine* is an experientially embedded project.

The third stratum, labelled the *public context and materials,* refers to the dimensions of the project that are less bound by the grounded and embodied character of *The Future Machine* and more stably persist through time. For example, the Guardians of *The Future Machine* are those individuals that have signed up to take responsibility for the machine itself and aid in its transport from place-to-place and reliably occurs more than once a year; these Guardians are encouraged to document their travel with the machine and to ritualize the process. Other forms of documentation (e.g., photographs, videos, early design sketches, and the climate data used to situate the project) are also persisting records of the project and have been used to disseminate information about *The Future Machine* both to the public and to potential and actual funding bodies.





[6] The When the Future Comes Collective is formed by the collaboration of six artists (Rachel Jacobs, Frank Abott, Juliet Robson, Wallace Heim, Caroline Locke, and Esi Eshun), two musicians (Alexandrea Yemaoua Dayo and David Kemp), and two computer specialists (Robin Shackford and Dominic Price).

The fourth stratum, termed the *governance context,* concerns the project's responsibilities to its funding bodies, for example, the Arts Council England and UKRI. Examples of such responsibilities are the Responsible Research and Innovation principles set out by the UKRI's Trustworthy Autonomous Systems Hub, as well as the responsibilities that the project team hold with each of its partners, specifically when working with primary schools in Cumbria, Somerset, and Nottingham. The specificities of the obligations put upon the When the Future Comes Collective and their commitment to fulfilling them are discussed in more detail in section 3.5.

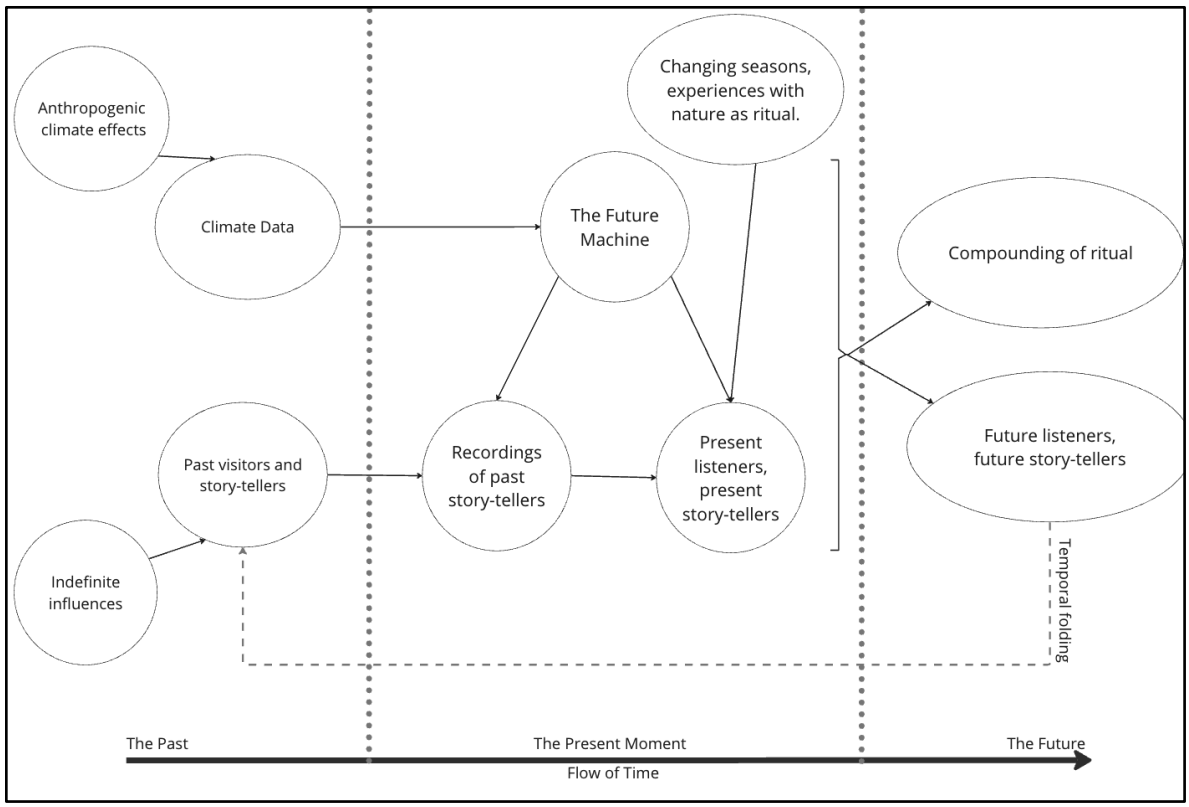
Finally, each of these four strata are encapsulated within the temporally bound and dynamic context of the global environment. The very impetus for the project in its entirety concerns the rapid march of the global community towards irreversible climate thresholds, which is directly connected to *The Future Machine* via the stories (located in the public context and locale strata) written by lead artist Rachel Jacobs and climate scientist Professor John King that are influenced by the global events precipitated by the collective influence of human activity on the climate.

To summarize, the generation of a diagrammatic ecology for The Future Machine yields an interesting way of conceptualizing the project and mapping the potential materials for inclusion within a dynamic archive. These materials include, the global climate data drawn from the global context, the documentation of the design processes and the photos, videos, and textual records of the project's multiple iterations, the recorded testimonies from participants, the localized climate data recorded by the machine itself, and the stories generated for the visitors to *The Future Machine*.

The creativity of the project concerns the weaving of time, place, and narratives (personal, collective, environmental). This weaving of narrative through technology produces *The Future Machine* not as the focus of attention but as the conduit for human connection to other humans and to our more-than-human others. It's not an input/output project; the artwork is inherently networked or rhizomatic. In other words, there is no central locus of significance. This much is evidenced by the artists' and researchers' own admissions that as the project continued, they discovered that Future Machine is peripheral:

> *"Rather, we are researching the future itself: what happens to the locations that temporarily host Future Machine each year, what happens to the artists and their visions, what happens as stakeholder groups come and go, what happens as public understanding of climate change shifts; [...]; what happens to the physicality of where we and Future Machine interact; and what happens when everything shifts again, and again, and again."*
>
> (Jacobs et al., 2023, 12)

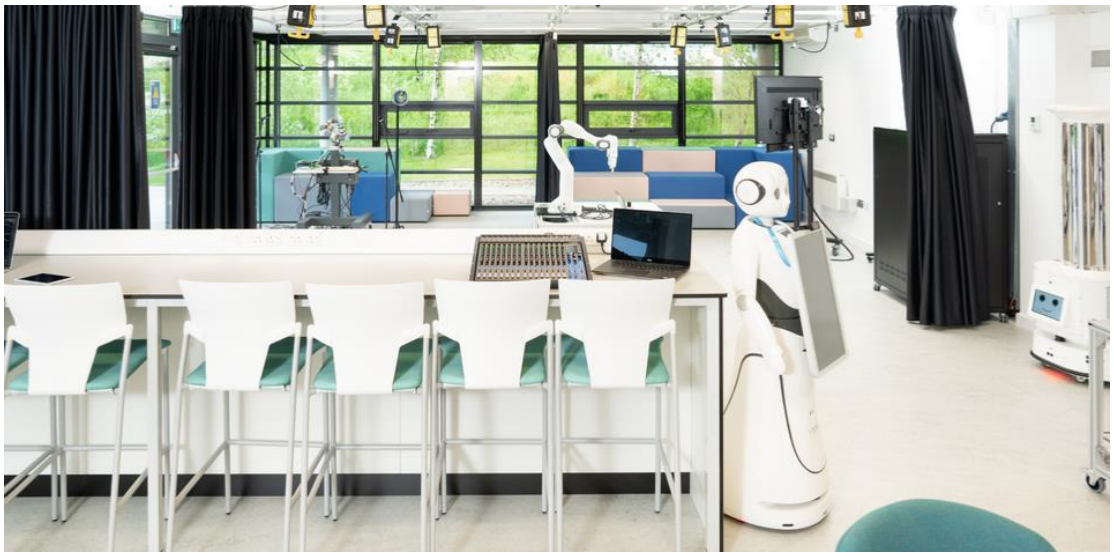
The meaning of the work, then, is given by the flow instantiated within the whole assemblage, including those innumerable and indefinite atoms of experience that influence the stories told, the reactions invoked, and so on. *The Future Machine* as project (rather than technology) gathers all these together and connects them via a circuit/trajectory. The authenticity, the honesty, of those stories told to the machine is central to the When the Future Comes project, which is facilitated by the intentional design of the techno-artefact as a *friendly* technology. In an attempt to represent the dynamism of the project, we include below a model of an *affective cartography* where embodiment and affect trajectories are essential for conceptualising the project (Carasco Segovia 2024). The Affective cartography approach foregrounds this dynamism as a departure from the diagrammatic ecology approach presented above. A dynamic archive will need to be able to reflect this embodied, temporal nature rather than freezing time and interaction out of the picture.



§2.3. Embodied Trust in Dance

§2.3.1. Textual Overview

Embodied Trust in Dance is a Trustworthy Autonomous Systems (TAS) Hub project. The work explores the concept of embodied trust through interactions between robots and professional disabled dancers. The following summary is based on information documented on the TAS hub website, on formative work and publications, and on a meeting held on 25/04/24 between the scoping project team and the *Embodied Trust in Dance* team, where project participants discussed the work, its findings, implications, and the processes involved.



REF: https://cobotmakerspace.org/about-the-space/

The purpose of exploring how professional dancers with various disabilities interact with robots broadly aligns with an overarching objective of TAS research; to 'improve people's physical and mental wellbeing' (Embodied Trust in TAS). In this case, interaction entailed '…examin[ing] the machine/body interface and reimagin[ing] bodily contact with robots as …creative, expressive and trustworthy rather than harmful and a problem to be avoided' (Embodied Trust in TAS). In practice, this meant exploring the interactions between professional dancers with physical disabilities, and a selection of robots (including, but not exclusively dancing). Four practical workshops took place, investigating a range of interactive research scenarios with assistive technologies, prosthetics, and through the application of somatic (body and movement) based methodologies such as 'contact improvisation and soma design' (Embodied Trust in TAS). A range of research questions were posed by the team, which included:

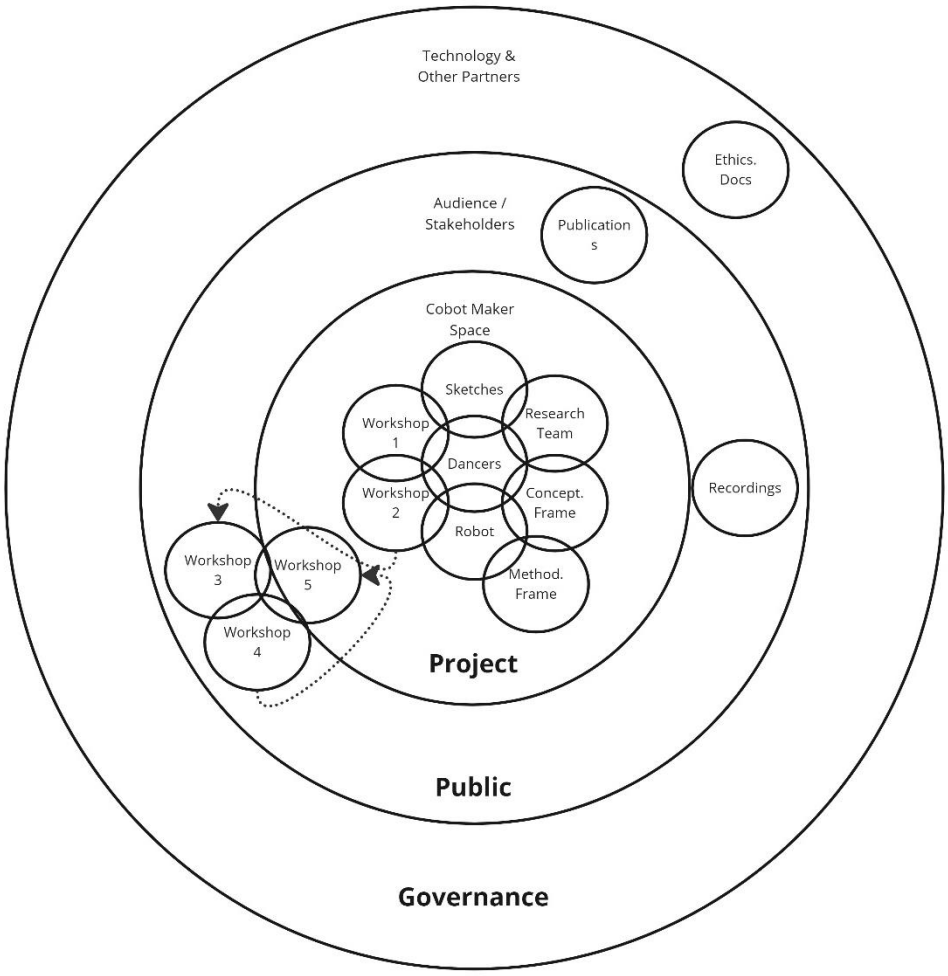
- '… what happens when [a dancer's] "partner" is a robot?

- ‘Is there a shared responsibility in moving together?’
- ‘What is the ‘language’ of the robots’ touch, or the sensation of physical contact with a human body?’
- ‘How can robots be more responsive, sensitive, and alert to contact with ‘live’ bodies?’
- ‘What might we learn about human-robot interaction that can be translatable to other contexts?
- ‘In programming robots to always avoid collisions, how might the dancer influence thinking about the creative potential of ‘collision’?

The dancers worked through various scenarios of interaction between dancers and robots including: one robot physically manipulated by a dancer whilst a second dancer responded to live/mirrored robot movements, or both dancers interacting with independent robots. Many interactions were explored including robot-dancer, robot-controller, and dancer-dancer, some live, and some with recorded robot movements. Moreover, interactions with other stakeholders in the room were considered. Interactions with others in space included camera operator and research/technical staff who controlled emergency safety switches and were tasked with stopping the robots if at any point health and safety was at risk. See Figure above where dancers are observed by researchers with “red buttons”.

§2.3.2. The ‘Embodied Trust’ Ecosystem as a Diagrammatic Ecology (Static Archive /Stakeholders)

If Embodied trust in TAS were to be curated as a static archive (DE), we may envisage the dancers as occupying central exploratory roles in the project space (the cobot maker space); workshops one and two, were organized with the dancers as the central focus. In workshops three and four the dancers were still pivotal, but activities were increasingly initiated and sometimes directed by the research team to explore specific themes. These workshops were more systematically recorded (in a more public context) with replay and analysis in mind. Workshop five continues the trajectory toward ever more public space (audience and other stakeholders), where recordings are replayed, analysed by all stakeholders and observations are reflexively and collaboratively produced. From an academic perspective, the project generated publications and publicity materials. Finally, we may consider ethics documentation, located in the governance space including technology and other partners.



The above diagram sets out these domains as concentric circles, and the artefacts within them as discreet nodes. Even as a static representation, the dynamic nature of the project is evident in the location of the workshops, which began in a well-defined project space but became comparatively more public as they progressed to include a broader group of stakeholders. This is exemplified by the arrow, showing workshops three and four occupying an audience-facing space alongside publicized recordings and publications.

§2.4. Trustworthy Accessible Robots for Inclusive Cultural Experiences (TARICS)

§2.4.1. Textual Overview

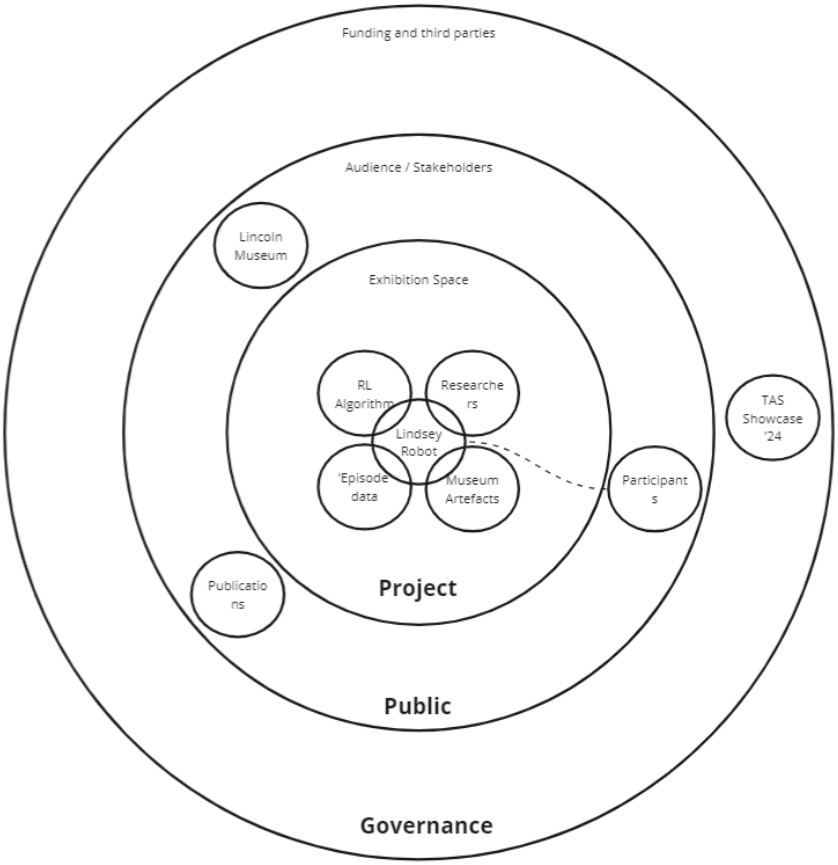
Another TAS Hub project, *TARICS* explores interactions between members of the public 'with learning disabilities and/or autism' and a social robot, with the objective of improving museum accessibility and 'interactive cultural experience [during] museum visits' (TARICS).

REF *LINDSEY' - TARICS tour guide robot, https://tas.ac.uk/research-projects-2022-23/tarics/*

In this project a tour guide robot - 'LINDSEY' – (an autonomous museum guide robot) was deployed to Lincoln Museum to test learning and interaction techniques in this real-world environment. More specifically Lindsey used several methods including: novel combination of 'state and action specification' (a list of sequential actions and action successors serving as rules for the robot); the 'engagement model' (measuring user engagement via continuous interactions to reinforce learning); and 'behaviour adaptation' (using an algorithmic operationalisation of 'the "optimism in the face of uncertainty" principle') (Duchetto and Handheide 2022).

§2.4.2. The TARICS Ecosystem as a Diagrammatic Ecology (Static Archive /Stakeholders)

Available archive materials include the algorithmic design data from the Lindsey robot, contextual information regarding its museum deployment, design information including ethical procedures, and interactional data and analysis including audiovisual, statistics, algorithmic modelling, and interpretation. Additionally, the TAS showcase, which took place in March 2024, London, offers supplementary information of early academic-funder engagement in the form of impact and conference brochures, pictures, presentations, and news media.



As with the TAS dance project, *TARICS* lends itself to static representation through the concentric circle- diagrammatic ecology model. The participatory nature of the data collection effectively means the project-public space is difficult to delineate.

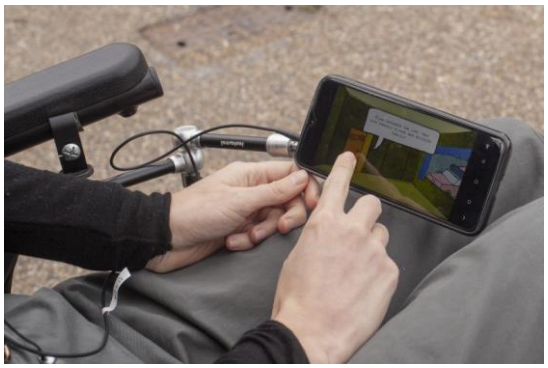
§2.5. Rider Spoke

§2.5.1. Textual Overview

Rider Spoke was originally developed by Blast Theory (artist group based in Brighton) in collaboration with the University of Nottingham's Mixed Reality Lab, Sony Net Services, and the Fraunhofer Institute as part of the European Union's funded project, 'The Integrated Project on Pervasive Gaming (IPerG)' which ran between September 2005 to February 2008

(Blast Theory, n.d.; CORDIS, 2008). The project has seen continued interest and staying power, evidenced by the repeated international iterations of the artwork.[7]

Rider Spoke, in the simplest sense, consists of multiple mobile phones or tablets that are mounted to the handlebars of bicycles (provided by the sponsor, Trek, or by the participants themselves) that display a user-interface inspired by Mexican votive art, religious imagery related to prayer and offerings, and naval tattoos evoking tales of adventure (Chamberlain et al. 2011, 6). As the lone player cycles away from Blast Theory's (sometimes temporary) HQ, they are asked to reflect on different moments in their lives, their relationships, their hopes and dreams, and so on.



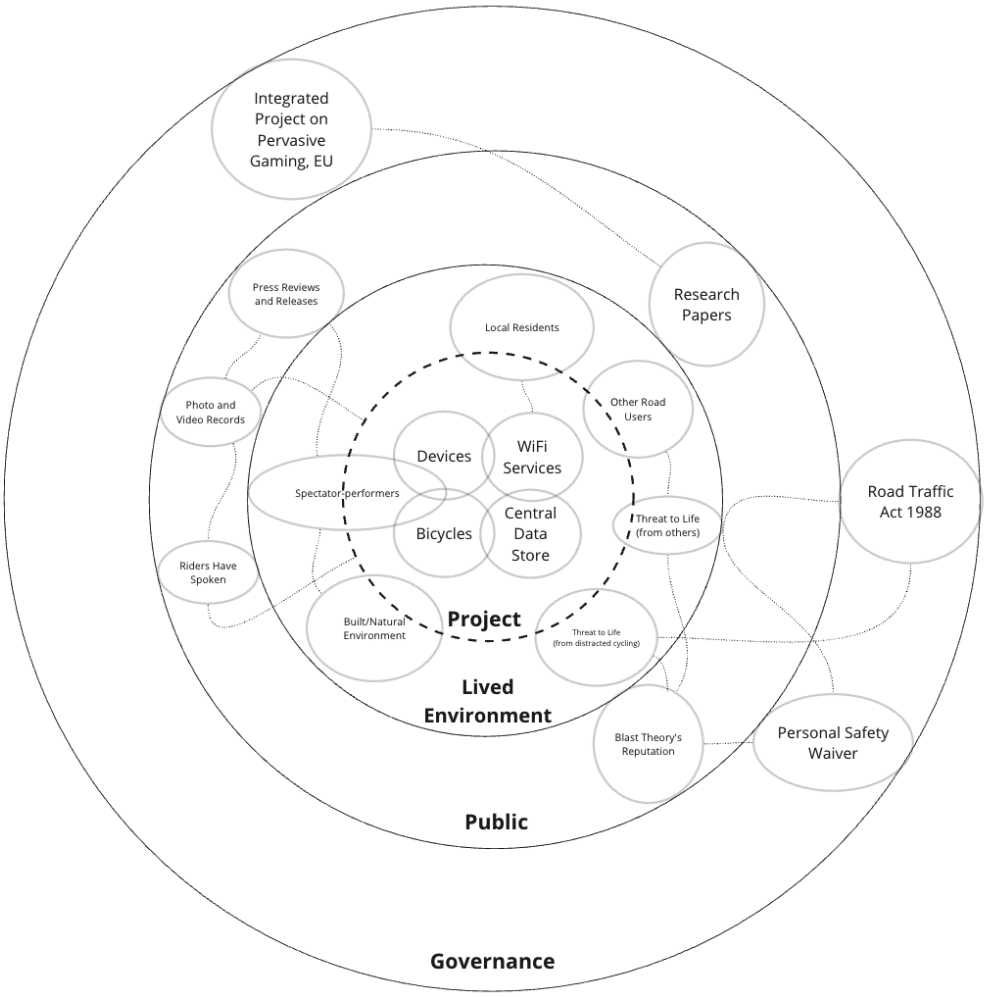
In one sense, the aim of the project is to elicit from these participants personal stories or intimate confessions (Chamberlain et al. 2011, 4, 6; Quigley 2016, 97). In another sense, as Blast Theory's Matt Adams explained in 2007, *Rider Spoke* challenges the traditional bounds of theatre and artwork by integrating networked technologies to, first, produce a digital artwork accessible only through participation, and second, to have this artwork produced not by team members, but by the audience themselves (Lavendar et al., 2007). As the individual cycles around, reflecting on prompts, recording questions, and 'hiding' them virtually, they can also explore the area for recordings left by other players. Thus, following Quigley (Quigley 2016), we refer to the individual players as *spectator-performers*.

Rider Spoke has several unique features. First, it continually toys with the tension between fiction and reality, between authenticity and fantasy. The proliferation of online spaces and contexts has provided opportunities for individuals to fabricate for themselves new identities; identity fragmentation and the possibilities for technology to encourage sincere confession as opposed to promulgating false testimonies presents itself as a core feature of *Rider Spoke* (Lavendar et al., 2007). Second, the project is unique as regards its very

[7] Rider Spoke has been performed in Adelaide, Athens, Brighton, Bristol, Budapest, Cambridge, Copenhagen, Edinburgh, Falmouth, Newbury, Norwich, Kupio, Leeds, Linz, Liverpool, Madrid, Sydney and Terni since 2007. Its most recent iteration being in 2021 at Brighton Festival (Blast Theory, n.d.).

composition; it is based on the operation of a road vehicle whilst engaging in mixed reality or pervasive gaming as the daylight fades, presenting significant opportunities for analyzing the connection between creativity and responsibility.

§2.5.2. Commentary on the Diagrammatic Ecology (Static Archive /Stakeholders)



The diagrammatic ecology we have generated for Blast Theory's *Rider Spoke* shares a key feature with the DE for *The Future Machine* (§2.2.2) in that the project is inherently experiential and embodied, thus we have chosen to represent this by marking the boundary between the core of the project and the space in which the project unfolds (the lived environment of the city) with a hyphenated line. *Rider Spoke*, at its most fundamental level consists of the relations between the mobile phones or internet tablets on which the *Rider Spoke* program is loaded, the bicycles upon which these devices are fixed, and the bespoke Wi-Fi location service that creates location fingerprints, and the central data store to which each device is synced each night during an iteration of the project.

Rider Spoke is nothing, however, without the participation of members of the public (spectator-performers) and their engagement with both the project itself and the built environment around them. Spectator-performers must contend with the liveliness of city-spaces, particularly as concerns other road users (whether other cyclists, car-users, or pedestrians) and be able to balance project engagement with situational awareness (see also the section on responsibility). Similarly, the project depends upon the indirect engagement of local residents insofar as the location technology depends upon recording local Wi-Fi signals (though this engagement is minimal, insofar as *Rider Spoke* does not seek to gain access to these Wi-Fi connections but use them as digital landmarks) (Chamberlain et al. 2011). In this lived environment stratum, we have further key concerns related to the threat to life sustained from injury, and the threat to life sustained from others. We note that participants identified as women noted that the practice of cycling solo around the city as night falls, or being encouraged to find somewhere to hide, caused significant anxiety.

Beyond the two blended strata of the project and its enactment in the lived environment of a city, the public context refers to those materials and resources that more stably persist through time. These include numerous reports and reviews published in the press and online as well as photographs of the project in progress and interviews with members of Blast Theory (e.g., Nick Tandavanitj at Brighton Festival 2021). Without a doubt, one of the most significant artefacts of *Rider Spoke* is the archiving project, Riders Have Spoken, developed by Gabriella Giannachi, Duncan Rowland, Steve Benford, and Dominic Price.

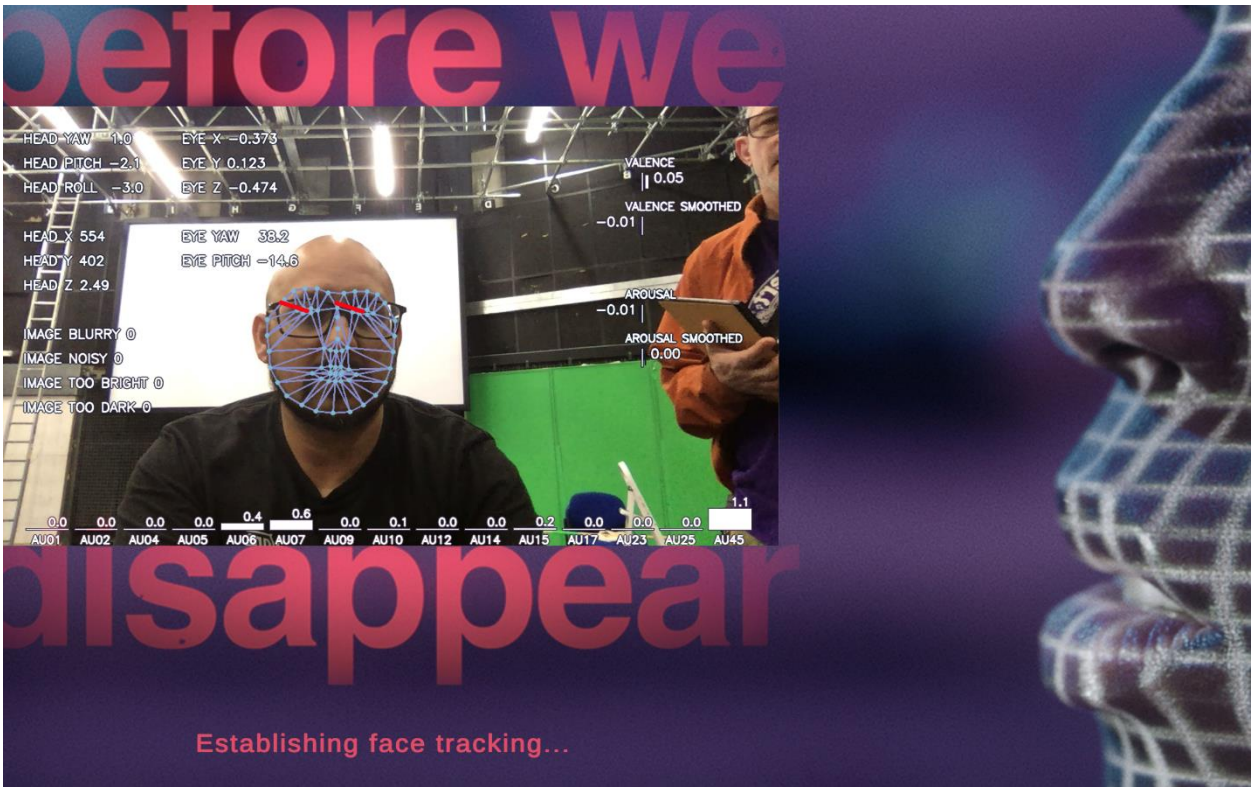
The final stratum for discussion in this subsection is the governance context in which Rider Spoke operates. As a product of the European Union's Integrated Project on Pervasive Gaming (IPerG), the artists that worked to develop *Rider Spoke* did so under the condition that their work would contribute to the rapid development of knowledge and understanding of how wireless mobile technologies could be integrated into the development of emergent leisure activities, with the aim of ensuring "European leadership in the development [...] of future mobile media content" (CORDIS, 2008). The team at Blast Theory developed an intensive Terms of Use and Indemnity Policy document, of which only the Melbourne iteration is publicly available online (Rider Spoke Terms of Use & Indemnity, 2022). The document states that all participants must have received a safety briefing regarding proper and safe use of the bicycle, and that all relevant traffic laws are observed. Although Blast Theory gains participant consent, some users sign off on falsehoods, such as familiarity with the local traffic laws of the area (Benedictus 2007); a phenomenon that Stenros, Montola, and Waern (2009) discusses in her chapter, "The Ethics of Unaware Participation in Public Interventions", as being all too common.

§2.6. ‘Before We Disappear’

§2.6.1. Textual Overview

Before We Disappear (see publicity video) is an interactive film created by Richard Ramchurn and AlbinioMosquito. It uses face recognition technology to adapt its storyline to the audience. The film is set in the year 2042, with the UK having become a tropical region with floods, fires, storms and civil unrest, and challenges the audience to face the issue of climate change https://www.albinomosquito.com/before-we-disappear/.

The filmmakers designed narratives which adapt to viewer’s engagement and emotional response. Designed to work with a standard laptop camera (or similar) audience member’s behavioural responses are assessed while watching the film using BlueSkeye AI’s (our industry partner’s) B-Social system. Facial features are recognized, tracked automatically (see Fig X below) and continuously analysed. These analyses are interpreted to produce metrics including affect, arousal and valence. The measures of valence and arousal are judged from similarity to a training set produced by filmmakers manually marking up what they thought the intended viewer emotion was for specific scenes. Researcher Mani Telamekala wrote an open-source program to collect the marking up data. The film narrative is changed according to how the system interprets the viewer’s changing emotional state. *Before We Disappear* has 3 possible endings, and around fifteen narrative constructs.



Before We Disappear is intended to be released through the Steam distribution platform. The film is embedded with b-social into a single package.

§3. Priorities and Features of the Structure

In what follows we discuss necessary features that need to be reflected in the structure of the dynamic archive on the basis of insights and data analysed from the case studies discussed above. These features are necessary but not sufficient. We hope to complete a full list of necessary features during the demonstrator project. The following sections give a brief description of the following features:

- Temporality
- Adaptability
- Trust
- Accessibility
- Inclusivity
- Transparency
- Consent
- Sustainability
- Replicability of the AI model

Several of the above features are identical to some of the strategic priorities of RAI UK such as transparency and fairness (RAI principles). We describe each feature below whilst discussing supporting evidence from the case studies. A detailed description and analysis of these features will be developed during the demonstrator project.

§3.1. Temporality and Adaptability

- Temporality

The When the Future Comes Collective, and more specifically *The Future Machine* makes use of creative techniques, traditional craftsmanship, and quantitative empirical data to manifest an experience for audiences that inculcates a sense of responsibility for the future of other people, both in their immediate geographic locale and at the global level. The project does not simply engage in the promulgation of myth and ritual for the sake of it, but rather to encourage a sense of cross-temporal connection with the past and the future; the focus of this connection concerns the responsibility for the climate.

What is most fascinating about this project's intense focus on the climate crisis and human responses to it – both fearful and hopeful – concerns the *temporality* of the project. As a longitudinal study projected to finish in the year 2050, the relation that it draws between current speakers and future listeners can be likened to a Mobius strip, where the listeners become those who are listened to, and they, in turn, become the speakers to be listened to. The present is simultaneously connected to the past and the future – this is one of the distinctive dimensions of *The Future Machine* as an inherently dynamic project: it's not about listening to the past; it's about speaking for the future. It's not about objectifying the past as an aesthetic artefact; it's about relating to the future people *qua* people in the world with their own stories and affectivity.

- Adaptability

TARICS uses '...several methodologies in order to have a unified framework that allows the robot to explore and learn online without the need of having separate phases for data collection and learning'. The technology used can learn by interacting with the public and adapt to future users. As described by the *TARICS r*esearchers, a rationale for use of the "optimism in the face of uncertainty" principle is given in relation to handling behaviour adaptation during uncertain user engagements: '...if the model is correct, you have no regrets (exploitation); otherwise, you have effectively learned something new about the world (exploration)' (Learning on the job 2022 p4).

§3.2. Trust

As evidenced in *Jess +* and in *Embodied Trust in Dance*, it was felt that to trust a robotic arm as a co-creator or a dancer should be analogous to trusting a human dance partner, with important similarities and differences. Trust with the robot arm was built through several steps: (i) rigorous health and safety training including description and exploration by all parties of the research environment. (ii) safety features and personal incorporated into the dance and research context so any physical risks were mitigated (iii) freedom of the dancers to stop whenever they wanted to, but also to explore and develop relationships with both the robots and the other dancers over time.

In a similar way, it may be necessary that users are given adequate time and resources to develop this trust with the AI applications they will be using. This is especially the case when embodied advanced technologies such as the ones used in *Cat Royale*, *Jess +* or *TARICs* interact with human and non-human animals.

§3.3. Accessibility and Inclusivity

- Accessibility

Rider Spoke revolves around the operation of what is essentially a road vehicle within a city-space whilst engaging with mixed reality media. Blast Theory, of course, were aware of these difficulties and thus took steps to navigate them (Benford and Chamberlain 2008, 37). First, the participants were to listen to the artwork through a *single* ear bud, which meant that situational awareness within the city was not compromised. Participants were also instructed not to listen to or operate the device whilst cycling (2008, 37-38). However, Giannachi reflected that some users, including herself, would be unfamiliar with the roads of the area, the traffic laws pertaining to the use of bicycles, and the riding of the bicycle itself (Lavendar et al., 2007, 11). This does of course affect the participant experience – "I didn't [...] feel liberated rather in difficulty" (Lavendar et al., 2007, 11).

- Inclusivity

Embodied Trust in Dance contributes to a tangible understanding of practical, RAI as it includes perspectives which are typically not included in the development stage. By addressing expectations and perspectives from dancers, audiences, researchers and technicians, it uses these to reveal emergent affordances – or practical values – relating to the robotic arm and its specific use within the *Embodied Trust in Dance* ecosystem. It creates a more inclusive ecosystem including minority stakeholders as main actors within this ecosystem.

In a similar way, *TARICS is* focused on '...increasing the accessibility of the cultural experience in museums for people with learning disabilities. As mentioned in the project publication the team plans 'to explore ways in which the proposed learning framework can enable the robot to behave in a more inclusive way and take into account the different user's characteristics and preferences' (Learning on the job 2022 p.7)'. The researchers took an 'experimental validation' approach, ultimately enabling conclusions to be drawn that the robot increased the engagement of the targeted members of the public: '...the robot policy learned to maintain the engagement of users for longer, with an increase of 22.8% over the initial static policy in the number of items visited during the tour and a 30% increase in the probability of completing the tour' (Learning on the job 2022).

Matters of inclusivity are also highlighted in *Rider Spoke*: participants reported feeling unsafe with the time of day, noting that "listening on the side of a dark street alone" was not a comfortable experience. Though Benford and Chamberlain (Benford and Chamberlain 2008, 39) claimed that "being mugged" is a "relatively exotic" and "unlikely" risk, the likelihood of harassment and violence against women is a separate issue and ought to be taken into consideration since a majority (58%) of the participants in the first iteration were women. In addition, there is a need to evaluate whether the production of these novel social spaces is acceptable and accessible to broader demographics. For example, are bicycles

essential to the performance piece or would it be possible to include wheelchair users as participants? *Rider Spoke*, as such, suggests that when trustworthy autonomous systems are used to construct mixed reality spaces, it is important to consider their production and operation as being socio-politically located and, therefore, as needing to respond to relevant concerns to demonstrate adaptability, accessibility and inclusivity.

§3.4. Transparency and Consent

- Transparency

Films produced within a research context such as *When we Disappear* are located within the public accountability of universities and their funding bodies. There is also accountability related to filmmaking in universally important areas like climate change. This is an open question – is there a requirement for a film with any sense of climate realism to be based upon research data for instance? There are multiple steps (visible, public or not) between "raw" data such as precipitation measurements, solar activity and so forth and a prediction of how or when the climate might change. In this case, steps between data and predictions are likely to involve computational models. Whether a responsible filmmaker uses and makes transparent the processes of research and inference is an open question.

Moreover, any lack of transparency has consequences for the establishment of trust between the audience and the artist or the artwork. More importantly when the artwork is a collaboration between artists and researchers under the providence of public institutions, transparency becomes more significant if we wish the users to trust the creators and engage with the artwork itself.

Finally, when reporting research findings generated in part from facial recognition and interpretation, the raw data from facial tracking and the interpretive algorithms may require to be publicly available for scrutiny. An understanding of how affect-recognition models are created, of whether there is the risk of inherent biases (eg. Racial, age related) in the training data and how the data may be captured and used are important concerns.

- Consent

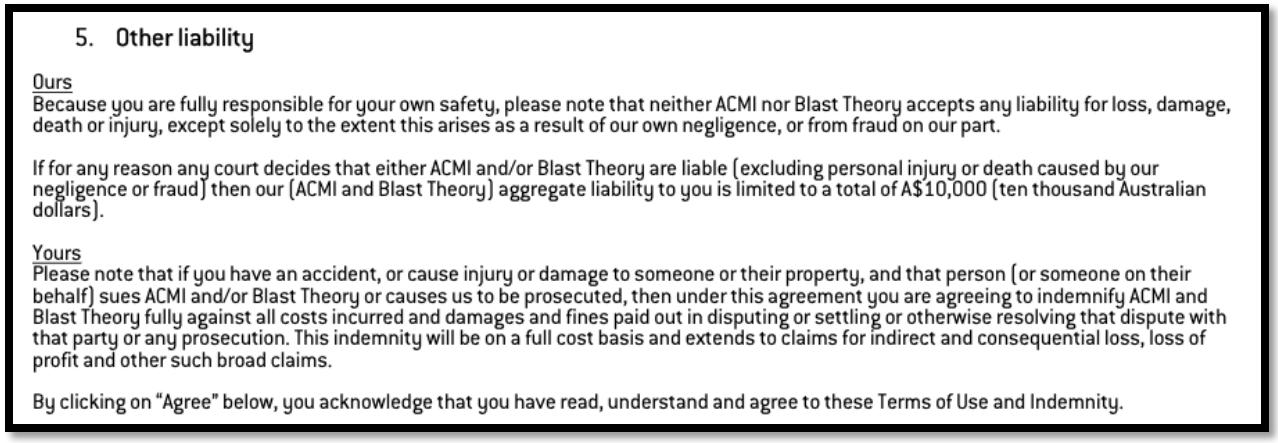
A significant concern regarding *Before we Disappear* regards the use of the AI facial recognition system: b-social. These are concerns that might be applicable to most (if not all) systems designed to collect and use personal data to infer emotional or other cognitive states. Are these considered as data collected from audience members? Is the collection and use transparent to viewers and are they able to watch the film without consenting? Are the processing models revealed to audience members and are they publicly available for

scrutiny? What data is stored, where and for what purposes? There is also consideration for filmmakers in general about their distribution strategies. B-social requires a license to work. Does this mean that viewers must be online to view the film and in online communication with Blueskye AI? If the film is reliant on the correct functioning of multiple related systems, are there responsibilities to ensure hardware and software capabilities are maintained, and for how long?

Rider Spoke also highlights a consent issue, one that is summarised by Benedictus (Benedictus 2007, 24, emphasis ours) in his preview of *Rider Spoke* published in The Guardian. He writes,

> "I have to sign a form, giving away my credit card details, *approving such falsehoods as*: "I know, understand and will comply with the road traffic laws under the Road Traffic Act 1988"".

This form, that eager participants will sign whether or not they are familiar with the traffic laws, also includes a waiver for any death incurred through participation within *Rider Spoke*. The difficulty with such procedures in the context of participatory art and pervasive gaming, however, is that when risk is involved, consent is often – as is well known – a complex matter (Stenros, Montola, and Waern (2009)). Whilst obtaining *implicit consent* is often sufficient for projects that involve minimal or negligible risk to participants, participatory art and pervasive gaming experiences have often "subjected participants to non-negligible risks, to harm, and to humiliations"; though consent is often obtained, "it is not necessarily fully informed" (Stenros, Montola, and Waern (2009). Whilst Blast Theory's Terms of Use and Indemnity document takes great pains to systematically work through the potential dangers of the project and secures consent that the individual participants are knowledgeable of the local traffic laws, evidently some participants are content to feign full awareness (e.g., Benedictus 2007, above).



5. **Other liability**

Ours
Because you are fully responsible for your own safety, please note that neither ACMI nor Blast Theory accepts any liability for loss, damage, death or injury, except solely to the extent this arises as a result of our own negligence, or from fraud on our part.

If for any reason any court decides that either ACMI and/or Blast Theory are liable (excluding personal injury or death caused by our negligence or fraud) then our (ACMI and Blast Theory) aggregate liability to you is limited to a total of A$10,000 (ten thousand Australian dollars).

Yours
Please note that if you have an accident, or cause injury or damage to someone or their property, and that person (or someone on their behalf) sues ACMI and/or Blast Theory or causes us to be prosecuted, then under this agreement you are agreeing to indemnify ACMI and Blast Theory fully against all costs incurred and damages and fines paid out in disputing or settling or otherwise resolving that dispute with that party or any prosecution. This indemnity will be on a full cost basis and extends to claims for indirect and consequential loss, loss of profit and other such broad claims.

By clicking on "Agree" below, you acknowledge that you have read, understand and agree to these Terms of Use and Indemnity.

When engaging with *Rider Spoke*, then it is of critical importance that the participant understands that Blast Theory have been absolved of their responsibility for the wellbeing of the participants. Interestingly, it is possible to read this transfer of responsibility to the participant as taking place within the frame of the experience also. Consider, for example, that it is not the responsibility of the *Rider Spoke* team to ensure that participants are engaging with the experience in the prescribed manner (that is, through honest, authentic, reflective confessionals). Whether the participant tells the truth, whether they record messages that they can feel as though they've imparted part of themselves, is the responsibility of the participant. The participant is not only responsible for their well-being, but also for their engagement with the pervasive gaming experience itself.

Rider Spoke offers two interesting ways of looking at the relationship between creativity and responsibility. On the one hand, Blast Theory generates a pervasive gaming experience that invariably opens up players to a multitude of risks and absolves themselves of responsibility through an intensive and well-considered Terms of Use and Indemnity policy. Simultaneously, this legal transfer of responsibility can be read as effectuating a shift of quasi-artistic responsibility to record testimonies that are in the spirit of the experience; as described in §3.2.4, some *Rider Spoke* participants, when they encounter emotive recordings come face-to-face with this artistic responsibility, and experience a radical shift in their responsibility as a participant.

§3.5. Sustainability and Replicability of the AI model

- Sustainability

When the Future Comes, as has been mentioned, is oriented towards tracking human responses to fluctuations and changes in local environments over the course of 30 years. The project seeks to challenge not just the views of the public and their daily interactions with the world, but also the design of technological infrastructure and the contemporary attitudes towards technologic innovation and economic progress.

The contemporary focus of economic growth as the mark of progress has a significant history and is often taken as equivalent to the development of technology, leading to the term, *innovation economy* (e.g., West 2011). Jacobs et. Al. are explicit about their desire for *The Future Machine* to raise "awareness of the full *environmental* costs of HCI, the internet of things and artificial intelligence" (Jacobs et al., 2023, 4, emphasis ours), meaning that the problem with the rapid rate of technologic expansion is not simply the mounting costs with which these developments are associated, but the often invisibilised environmental damages, including "the resulting CO_2 emissions and water requirements for cooling the

vast systems that power the necessary increases in computing power and efficiency, alongside the damaging extractions required for manufacturing the hardware." (Jacobs et al., 2023, 4).

The When the Future Comes Collective has a clear-sighted focus on environmental sustainability. *The Future Machine* is a manifestation of the desire for "a much more responsible holistic way" of thinking about and producing projects "from conception to design and deployment" (Jacobs et al., 2023, 4). The technology that is hidden away inside the ash and brass exterior is "mostly refurbished or repurposed from previous projects" (Jacobs et al., 2023, 11) and is powered by a repurposed car battery activated by the manual turning of the exterior crank. The battery itself can be charged either via mains supply or else through the machine's solar umbrella. Transportation also proves to be a challenge to the collective's aim for a net-zero device, but nevertheless the collective continues to research innovative *and* ecological ways to reduce the device's footprint (Jacobs et al., 2023, 11-12).

By challenging the expected appearance of technology (e.g., by drawing on timeless materials with mythic significance) and working towards net-zero in the process (e.g., using repurposed wood, brass, and technology) to produce a single object, When the Future Comes sought to challenge "our throwaway, fast culture and transient relations to everyday objects", a decision that contributed to the *presence* of *The Future Machine* itself as "'cathartic', 'celebratory', 'joyful' and [as] 'a friendly technology'". *The Future Machine*'s presence is *trustworthy* precisely because it breaks from the expectations that circulate the public imaginary; trustworthiness can flow from responsible design decisions.

The When the Future Comes Collective is committed to the promulgation of the myth of *The Future Machine* always describing it as a *witness*, a *ritual*, or as a "mysterious and mystical device" (When The Future Comes, n.d.). Powered by a windows laptop and car battery, *The Future Machine* is an extensive longitudinal "artistic technology probe" (Jacobs et al., 2023, 12). The aim of this probing project is to measure and identify changes in emotive responses to climate change.

- ### Replicability of AI model

Insofar as the digital system of *The Future Machine* concerns the cataloguing of a series of recordings, each pre-labelled by the user under a certain myth, mood, or theme (e.g., Lamenting Eden, Bringing on the Apocalypse, Constructing Babel, Celebrating Jubilee) and replaying a relevant recording at a future moment when one of these themes is selected by the user the AI model is replicable in other contexts. A very similar model was employed by Giannachi, Rowland, Kwastek, and Blast Theory during the archival project that followed *Rider Spoke*, *Riders Have Spoken*.

In terms of model replicability, *TARICS* was considered '...a promising step toward behavioural adaptation in long-term scenarios for robotics applications in social settings' [11]. Taken as a whole, the physical and non-physical artefacts associated with the project are therefore highly replicable when repeatedly deployed in the same or similar environments. However, the more the environment changes – (a different set of museum artefacts, a different museum, a different public etc) – the less replicable the AI model is.

Moreover, 'a limitation to the applicability of the proposed framework for more complex problems is that it cannot handle high dimensional state and action spaces' [11]. Thus, and presumably, its application is optimised for spaces with known dimensions.

Rider Spoke incorporates innovative location technology that was designed bespoke. As described previously, Rider Spoke uses technology that periodically scans the available Wi-Fi signals and records the available networks to position each individual participant relative to these overlapping Wi-Fi signals, building up a digital map of the city with no pre-existing model (Lavendar et al., 2007; Chamberlain et al. 2011; Opperman et al. 2011). As Oppermann et al. (2011, 3) observe this is "a lightweight and adaptable approach to location" which permits flexible redeployment of *Rider Spoke* to different and repeat locations without the need to engage in extensive city scans or pre-planning (Lavendar et al., 2007, 9). This is evident given the immense success of *Rider Spoke*'s international touring history, which began with Athens in 2008, the year following its debut in Brighton.

The location technology, which is arguably the fundamental structure of *Rider Spoke*, is therefore easily adaptable to varied cities around the world providing that these cities are well-populated with Wi-Fi signals. It is therefore conceivable that *Rider Spoke* could be deployed in any town-sized environment. Beyond the locative services, the game is run on the tablet which is affixed to the handlebars of the bike, as mentioned previously, it would be possible to open up the *Rider Spoke* performance to other modes of transport, perhaps most plainly mobility aids such as wheelchairs. Indeed, one can imagine an iteration of *Rider Spoke* that is directed towards the production of a mixed reality space that records testimony of disabled people recording their experiences with ableist architecture and design as a means of educating able-bodied people about the barriers that frequently go unnoticed. Such a project could be used not to challenge the phenomenon of identity fragmentation and authenticity, but to challenge the widespread acceptance of ableist architecture and city design (social model of disability). To summarise, *Rider Spoke* was designed with the ease of redeployment in mind. This simple and supple design means that it is a very malleable approach to the production of pervasive gaming and can be redeployed in a variety of contexts.

The AI model used in *Before we Disappear*, b-social, is a standalone system developed by Blueskye AI and as such is usable on its own or during any media presentation, social interaction or indeed anywhere a human face is visible to a digital camera. B-social can also be used to analyse a video recording to produce output metrics. B-social requires a license to function and in that sense the model may be functionally replicable, however unless the model is publicly available, it may not be clear to filmmakers (and other interaction designers etc.) how they would incorporate b-social into their projects.

Before we Disappear as developed with an explicit focus on the responsibility and ethical concerns inherent when using personal data, in particular face processing mechanisms. These concerns being particularly salient when filmmakers or other producers, are creating works designed to invoke emotional reactions, deal with contentious or anxiety inducing concepts and may be consumed by a public unknown to the producers with personal or social characteristics which may make them vulnerable in various ways.

The work was in collaboration with BlueSkeye AI who has incorporated privacy by design into its technology from its inception. Data collection and storage is minimised wherever practical, and all data is processed on people's own devices, without using the cloud. This gives users control over who they share their data with, when and always with end-to-end encryption. The intended release as an interactive app, "incorporating an awareness of potential abuse of the user's data, and safeguarding any personal data on the device used to watch it". More details on Horizon case study.

§4. A hybrid framework for structuring a dynamic archive

In this section, we introduce a speculative design for a framework from discussions during the project workshops. This framework is based on enabling the four foundational assumptions of dynamic archive functionality, mentioned in section 2.

During workshops we discussed two HCI approaches regarding the mapping of ecosystems:

a) in terms of object-interactions (stakeholders, artefacts, etc.,) and
b) in terms of trajectories tracking participants' perspectives.

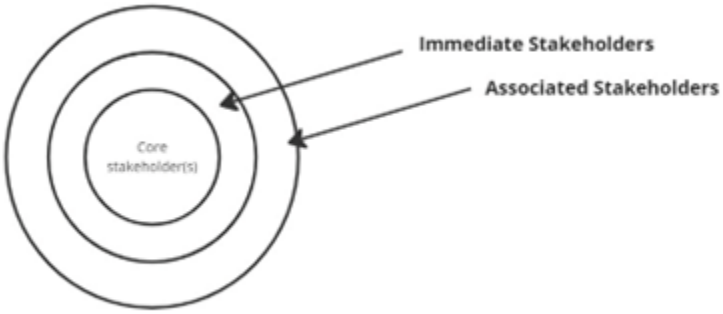
The first approach uses concentric circles and draws heavily on preexisting notions of boundaries and domains of interaction. The second approach uses a linear representation of interaction stages emphasising the passage of time and emergent pathways taken by individual participants or artefacts.

In this section we combine these frameworks, critically examining their affordances when conceptualised as interactive, dynamic archives, to see whether this hybrid model can

accommodate the necessary features mentioned above, and whether it is aligned with fundamental RAI principles more broadly.

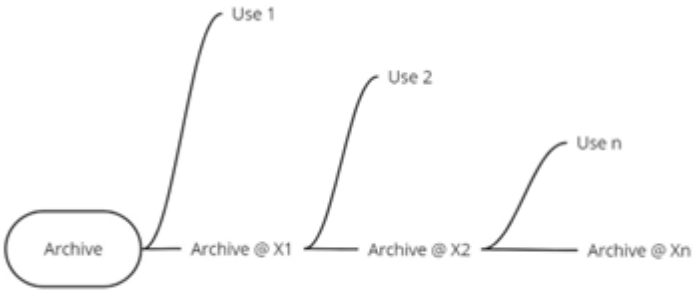
§4.1. From Concentric Circles to Trajectories

The concentric circle model sets out components of a project as a static archive, depicting human and technological actors, as data repositories or interactive touchpoints; crucially these components stay within the boundaries of defined stakeholder groups and contexts. On the other hand, the trajectory model – a formal framework for developing HCI propositions – has been instrumental in rendering the projects dynamic; delineating the various pathways of interaction conceivably useful to future stakeholders.

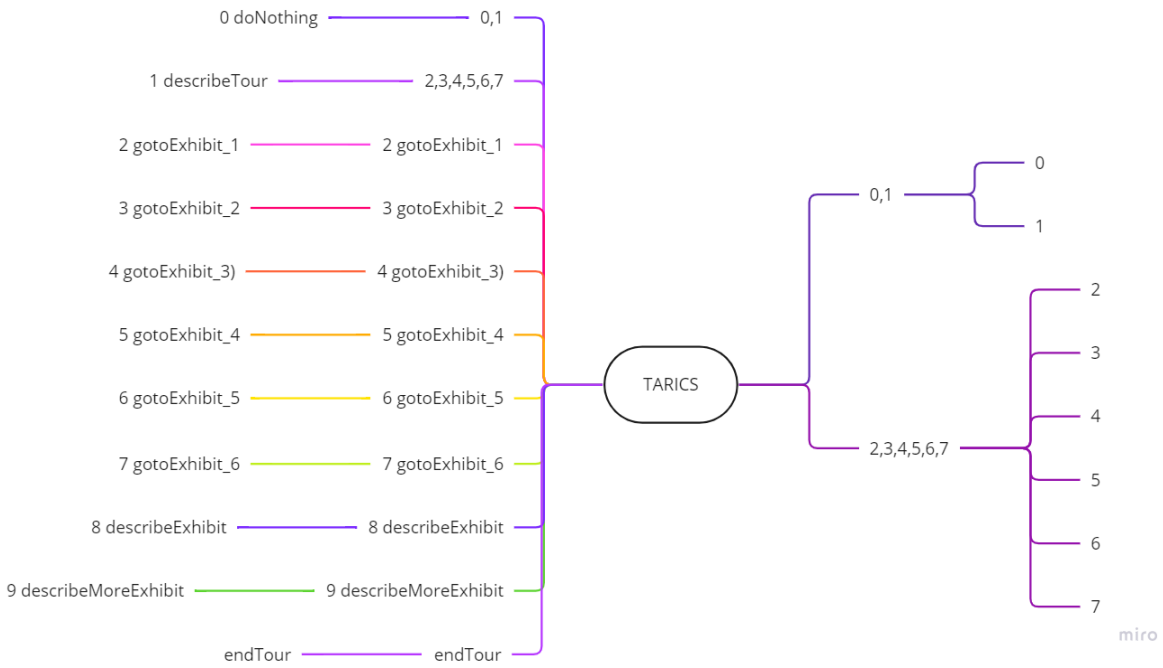


The above concentric circle diagram illustrates AI ecosystems, centred on the primary artefacts, stakeholders, and interaction between them as a core feature. Immediate stakeholders - those present and influencing the interaction(s) such as researchers and technicians, or present and observing the interaction(s) such as an audience or the wider public – typically form a first layer peripheral circle. Associated stakeholders – such as funders, ethics committees, and the societal context in which the project takes place – typically form a second layer peripheral circle [8].

This model exemplifies a static approach to archiving the ecosystem, and while dynamic aspects such as instances of interaction, or transcendent themes such as shared stakeholder values, can be depicted as such, the implication of the model is one of a fixed delineation of boundaries asserted by an assigned expert: An alternative or lay-user is likely to have a different perspective of what is core, and what is peripheral. The boundaries may in many cases be fuzzy, if not arbitrary.

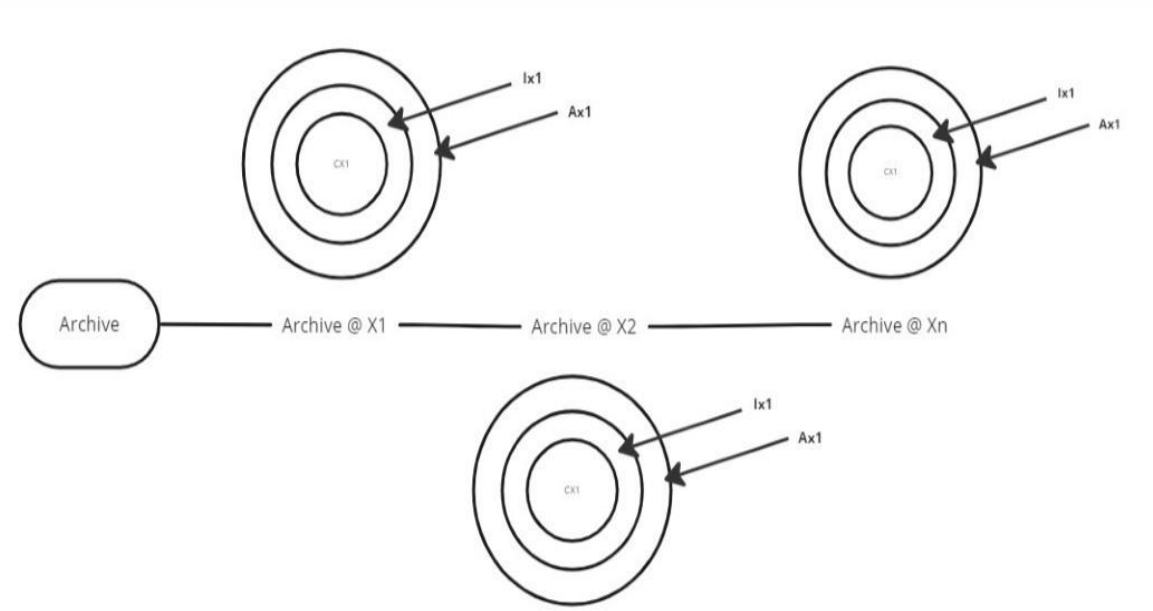


Another way in which ecosystems can be illustrated is via the human computer interaction (HCI) paradigm of the trajectory (see figure above). Compared to the concentric circles model, this more closely resembles the nature of interaction as a series of progressive events as observed by various stakeholders, rooted in an original instance (the archive) and documenting the use and reuse of artefacts at different stage.



An example of this is evident in the *TARICS* project's use of prescribed robot 'actions' and 'action successors' (see figure above): On the left, 10 actions (0-9) are defined and assigned logical successors, covering all possibilities of participant interaction with LINDSEY in the museum, though description and navigation to exhibits, and ending of the tour. Imagining these actions as trajectories, it would be possible to compile instances of the real-world

substantiation of their successors. For instance, to ‘doNothing’ can be archived as instances of non-engagement ‘0’ or progression to ‘describeTour’, where successors are the sum of interactions with LINDSEY(Learning on the job 2022). Moreover, ‘gotoExhibit’ might be substantiated by all instances of user response to the exhibits in situ ‘2-7’ (Learning on the job 2022).



Finally, we discuss the combination of the concentric circle and trajectory models. In the above illustration, the concept of the trajectory is combined with the concentric circle approach, such that multiple trajectories of use cases can be captured for different users at each stage of interaction X_1, X_2... Xn creating different archives focusing on different stakeholders. This hybrid model arose from the need to retain original ecosystem boundaries as defined by project research teams, while affording them ongoing use sensitive to the subjective and currently unknown requirements of future users.

Below we use some of our case studies to speculate what a dynamic archive of these cases would be like and whether the hybrid model developed in this scoping project can better accommodate some of the required features and priorities of a dynamic archive. As this hybrid model is a work in progress, it does not address all features and priorities. Further research is required to ensure the structure can accommodate all features and priorities which can be undertaken during the demonstrator project.

§4.1.1. When the Future Comes

Drawing on the reports from audience members who connected with *The Future Machine* as a friendly technology, a dynamic archive for *The Future Machine* could be constructed on the premise that the archive is a means of conversing with *The Future Machine* itself; that is to say, the dynamic archive could function as an analogue for *The Future Machine*'s apparent inner-life, a fabrication of the first-person phenomenological experience of the mysterious machine outside of time itself. To expand on this, the dynamic archive, could be designed to be interactive and fill the role of an old friend or warm acquaintance, perhaps one met long ago who has access to all data relating to the ongoing project. What would it mean to converse with *The Future Machine* (pseudo-machine) considered as a dynamic archive in this way? What could be possible?

Incorporating an AI model to be able to respond to requests, for example, to hear stories from different years, or to listen to how the weather-made music changed in a single location every five years, could be interactive and can make use of the power of retrospect to synthesise the findings of *The Future Machine*'s longitudinal study into a matter of minutes in different ways. Similarly, timelapses of the photographs taken by *The Future Machine*'s horizon camera in a single place could show the changes that took place across a significant time span that might not be readily perceivable in other ways. It may also be possible to create movable graphs that represent changes to both the local areas studied and the globe in general, perhaps creating an interactive timeline of climate events against the data recorded by the *The Future Machine*.

Alternatively, the same archive could allow a pseudo-machine to narrate the history of the design process, the decisions and rationale, the highs and lows of the artists and the art project itself (e.g., the pitfalls and breakthroughs that came as a consequence of COVID-19) all of which have been recorded in documentation. When *The Future Machine* is over, dismantled and no longer running, the pseudo-machine could function as a mythological relic or creation story – a living remain rather than a distant memory.

Finally, this pseudo-machine functioning as a dynamic archive may be able to provide visitors with their own future quest in much the same way that *The Future Machine* does as a way of preserving what may be taken to be the primary interaction between users and artefact.

§4.1.2. *Embodied Trust in Dance*

Creating a dynamic archive of *Embodied Trust in Dance* based on the concentric circles structure entails placing the artefacts and interacting stakeholders (dancers) front and

centre. Technicians, followed by academic stakeholders, ethics committees, the TAS hub, and the public, would occupy progressively peripheral circles. On the other hand, creating a dynamic archive based on trajectories would emphasise individual stakeholder perspectives e.g. dancers, academics, audience etc. over time, more so than the concentric circle model. Combining these approaches might allow research aims such as development of trust over time to be explored by identifying and highlighting relevant touch-points, interactions, dialogues and actors. This hybrid approach has an inherent sense of chronology; any concept or theme of relevance might be revealed by mapping across either or both time and space.

§4.1.3. Trustworthy Accessible Robots for Inclusive Cultural Experiences (*TARICS*)

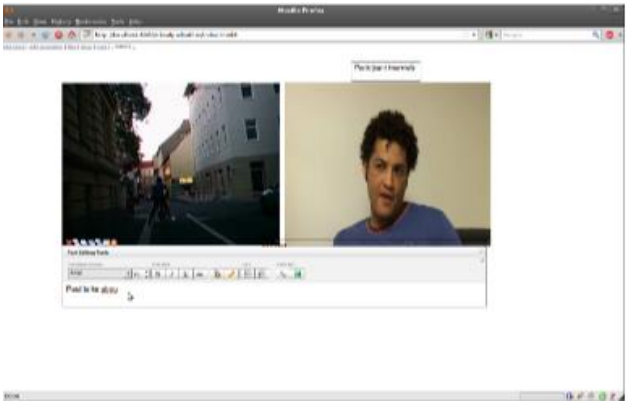
Imagining *TARICS* as a series of concentric circles, or a diagrammatic ecology, we assigned project artefacts such as the LINDSAY robot, its algorithms and episodic data, alongside the researchers and museum artefacts in a central 'exhibition' or project space. Peripheral to this and in the 'audience/stakeholder' or public space, we placed the Lincoln Museum and publications. Participants were also included here, but with the explicit linkage made to the central exhibition space via the LINDSAY robot emphasising their potential movement between spaces. Finally, and in terms of the 'funding/third party' or governance space, we included the (project) ethics data/procedures, as well as the 'TAS showcase' event in 2024, as instances in which RAI were explicitly conveyed.

Imagining *TARICS* as a trajectory or 'affective cartography', we noted that this was already evident in some of the interaction frameworks present in the design of the LINDSAY robot. The hybrid model would allow for more adaptability, temporality and replicability of the AI applications used in this project.

§4.1.4. Rider Spoke

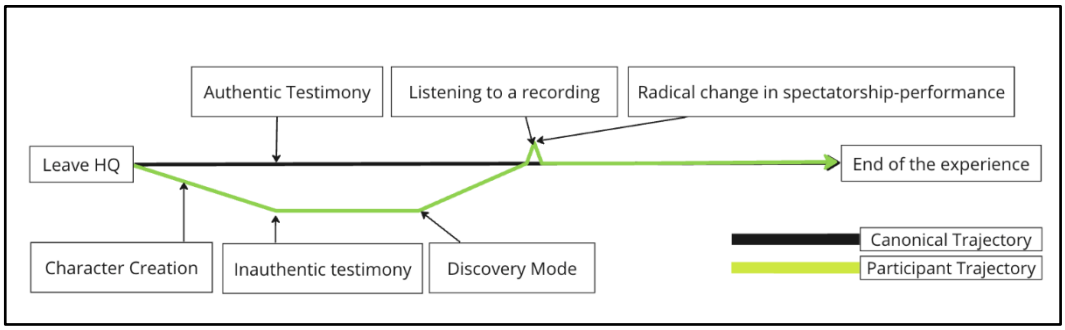
Rider Spoke is accompanied by the project, *Riders Have Spoken*, led by Gabriella Giannachi and funded by Horizon Digital Economy Research. In short, *Riders Have Spoken* concerned "an interactive archive of selected recordings from Rider Spoke" (Riders Have Spoken, n.d.). During the course of *Rider Spoke*, over ten thousand audio recordings were made by players and pinned to quasi-physical locations, accessible through a *Rider Spoke* handset.

The archive produced as the *Riders Have Spoken* Project was presented as a hand drawn city scape, with lights on in some of the windows. Users simply had to click on a lit window in order to hear an individual's response to a prompt, whether it is about their father, their deepest hopes, or their first-time holding hands with someone. This specific design was used to give users that had no experience of *Rider Spoke* a sense of the atmosphere and "specific context in which the recordings had been made" (Riders Have Spoken, n.d.). Taken as a whole, the archive consisted of the user interface (the drawn cityscape) which permitted access to video recordings of a participant *in situ* (captured by a trailing ethnographer) and a post-participation interview (Riders Have Spoken, n.d.). (see figures below).



The archive was inherently iterative insofar as users and visitors were able to annotate the videos recorded of nine participants (Giannachi et al. 2010). Chosen participants were recorded from the front and from the rear and were GPS-tracked to sync visual and audio media to specific locations and were also interviewed following their engagement with Rider Spoke. The archive itself was a CloudPad system, a platform for multi-media mashups that could make handling videos, audio files, and textual annotations very simple.

The aim of the archive is to collate and record historic trajectories formed by the recollections of the participants as they recall their own participant trajectories through the world of *Rider Spoke*. Users or visitors of the archive can access these historic trajectories through the lens provided by Blast Theory of *Rider Spoke* itself; that is to say, it is through the framing of the project itself that one comes to contextualise the individually contextualised experiences of the project as it unfolded. For Giannachi, making *Riders Have Spoken* iterative or dynamic is fundamental insofar as these users are able to draw their own participant trajectories through the work as they engage with it asynchronously. Below we have chosen to include a simple diagram of Quigley's participant trajectory, showing how the encounter with authenticity-despite-anonymity challenged her own inauthenticity-because-of-anonymity and resisted playful antagonism within the frame.



These two projects, *Rider Spoke* and *Riders Have Spoken*, are prime candidate projects where the hybrid model could be applied to ensure that a dynamic archive of these projects satisfies the necessary features sketched in this scoping project and as such aligns with the priorities and commitments of RAI UK and RRI. We aim to explore the viability of this framework during a demonstrator project.

References

[1] G. Giannachi, *Archive Everything: Mapping the Everyday*. The MIT Press, 2016.

[2] S. Blackburn, *The Oxford Dictionary of Philosophy*. OUP Oxford, 2005.

[3] S. Benford and G. Giannachi, *Performing mixed reality*. MIT Press, 2011.

[4] G. Giannachi, H. Lowood, D. Rowland, S. Benford, and D. Price, "Cloudpad - a cloud-based documentation and archiving tool for mixed reality artworks," in *ADHO Stanford*, 2011.

[5] G. Briscoe and P. De Wilde, "Digital Ecosystems: Evolving Service-Orientated Architectures," 2006.

[6] A. Butterfield, G. Ekembe Ngondi, and A. Kerr, *A Dictionary of Computer Science*, 7th ed. Oxford: OUP, 2016.

[7] R. Velt, "Putting Trajectories to Work: Translating a HCI Framework into Design Practice."

[8] S. Benford, G. Giannachi, B. Koleva, and T. Rodden, "From Interaction to Trajectories: Designing Coherent Journeys Through User Experiences," 2009.

[9] R. Owen, J. Stilgoe, P. Macnaghtnen, M. Gorman, E. Fisher, and D. Guston, "A framework for responsible innovation," in *Responsible innovation: managing the responsible emergence of science and innovation in society*, R. Owen, J. Bessant, and M. Heintz, Eds., Wiley Online, 2013, pp. 27–50.

[10] René von Schomberg, "A Vision of Responsible Research and Innovation," in *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*, Wiley Online, 2013, pp. 51–74.

[11] F. Del Duchetto and M. Hanheide, "Learning on the Job: Long-Term Behavioural Adaptation in Human-Robot Interactions," *IEEE Robot Autom Lett*, vol. 7, no. 3, pp. 6934–6941, Jul. 2022, doi: 10.1109/LRA.2022.3178807.