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

AHRC/UKRI BRAID Programme 2024

Output 3: Workshop Report 3: 5 additional case studies[1]

Organizers Lydia Farina, Helena Webb

[1] To cite this report use: Drury, M.R.F., Miles, O., Brundell, P., Farina, L., Webb, H., Giannachi, G., Moore, J., Benford, S., Jordan, S., Stahl, B., Perez Vallejos, E., Vear, C. 2004, Workshop Report 3, BRAID/ AHRC OPP18206, 3.

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

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

§1. Introduction

§1.1. Overview

The following report introduces five additional case studies, selected to inform the creation of a dynamic archive of Responsible AI (RAI) ecosystems. These are: *The Future Machine*; *Embodied Trust in Dance*, T*rustworthy Accessible Robots for Inclusive Cultural Experiences (TARICS*), *Rider Spoke*, and *Before we Disappear*.

As with *Jess+* and *Cat Royale*, these projects share concerns around trustworthiness of AI applications, addressing ethics issues by design, contributing to emancipatory futures and debating AI's place in the world. Continuing the approach of previous reports, we introduce the themes, aims, and iterations of each of these five projects and evaluate their unique implications for creating a dynamic archive prototype. As some of these case studies are live projects, this report should be read as more speculative, compared with those of Workshop 1 (WS1) and Workshop 2 (WS2).

§1.2. Objectives

- To outline five additional creative AI ecosystems informing the CRADLE project
- To describe ecosystems in terms of artefacts and frames (static archives)
- To highlight implications for interactive artefacts and frames (dynamic archives)

§1.3. Participants

Lydia Farina (LF)- Project Lead, Philosophy, UoN (University of Nottingham)

Helena Webb (HW)- Deputy Project Lead, Computer Science, UoN

Gabriella Giannachi (GG)- Co-I (Co-Investigator), English, Exeter

John Moore (JM)- Co-I, The National Archives

Jenn Bunn (JB),-Guest speaker, The National Archives

Steve Benford (SB)- Co-I, Computer Science UoN

Bernd Stahl (BS)- Co-I, Computer Science, UoN

Craig Vear (CV)- Co-I, Music and Computer Science, UoN

Spencer Jordan (SJ)- Co-I, English, UoN

Megan R. F. Drury (MD) – RA – Philosophy, UoN

Oliver Miles (OM) – RA – Computer Science, UoN

Pat Brundell (PB) – RF – Computer Science, UoN

§1.4. Agenda

- **11.00-11.15 Arrivals/ coffee**
- **11.15-11.30 Project update and objectives of workshop (LF)**
- **11.30-11.45 Digital Archives (Jenny Bunn, TNA)**
- **11.45-13.00 Presentation of 5 additional case studies (MD, OM, PB)**
- **13.00-13.45 Lunch**
- **13.45-14.45 Summary findings/ Comparison of case studies (MD, OM, PB)**
- **14.45-15.00 Coffee Break**
- **15.00- 16.15 Towards a prototype design of a dynamic archive (group discussion/ LF/HW)**

  Reminder of key research questions:
  - Who are the main stakeholders in this context?
  - Is the model of AI ecosystem inspired from this context applicable to other contexts?
  - What would a prototype design of a dynamic archive inspired by the 7 case studies look like?
- **16.15-16.30 Summary findings and next steps (group discussion/ LF)**

Defining the Dynamic Archive: A living glossary of terms

Term	Definition
Archive	...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]
Affective Cartography	A type of dynamic archive
Authenticity	The condition of significant, emotionally appropriate living. Contrasted, especially in Heidegger, with inauthenticity: a state in which life, stripped of purpose and responsibility, is depersonalized and dehumanized. [2]
Canonical Trajectory	...an artist's intended journey through the performance or overall narrative' [3]
CloudPad	a cloud-based documentation and archiving tool for mixed reality artworks.[4]
Artificial Intelligence used in the creative industries)	AI applications used in the creative industries context; art, music, dance, etc., https://braiduk.org/creating-a-dynamic-archive-of-responsible-ecosystems-in-the-context-of-creative-ai
Creativity	Data curator has creative role, https://braiduk.org/creating-a-dynamic-archive-of-responsible-ecosystems-in-the-context-of-creative-ai
Dataset (Open AI)	...a term that remains somewhat nebulous. For researchers, open AI may mean collaborative and reproducible science and systems. For technologists, it may centre on free use and distribution of AI models, or perhaps public participation in an algorithm's development, Ding, J., in 'What defines the 'open' in 'open AI'?', The Alan Turing Institute, Blog Post, Accessed 02.07.24
Dynamic Archive	Since the archive has been referred to as 'apparatus', it can be operationalised as much as a structure and tool for construction as a set of documented content. [1]
Ecosystem (Digital)	...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]
Framework	...an outline for the software's structure in the form of objects that not only themselves provide basic functionality but also integrate with each other'.[6]

Historical Trajectory	…a synthesized story of an experience based on one or more participant trajectories'[7]
Participant Trajectory	…an individual participant's actual route through the experience' [8]
Responsible Artificial Intelligence (RAI)	'…developments which 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. These will guide everything the ecosystem does, including its internal management, and the research, engagement, and skills programmes'., https://rai.ac.uk/guiding-principles/
Responsible Innovation	"… taking care of the future through collective stewardship of science and innovation in the present." [9]
Responsible Research and Innovation (RRI)	Attending to '…the bigger picture, including the long-term impacts of their work and its value to society. Invoking ethical acceptability, 'social desirability, and 'sustainability' [10]
Temporal trajectory	…mappings between the actual time of an experience and the "plot" or story time, defined by the narrative participants engage with'[7]

How to use this glossary

The above resource is intended as a helpful resource with working definitions of some key concepts to allow the reader to engage with the report more critically.

§2. Project Overviews

In this section, we review the key themes and features of five projects that exemplify *artificial intelligence* applications used in the context of the creative industries. This section therefore operates as an introductory survey of the case studies selected by the CRADLE team for analysis with a view towards developing a framework for a dynamic archive

In the second workshop (WS2 hyperlink), SB introduced the CRADLE team to a model he called an *ecology of trust*. These diagrammatic ecologies (DE) are visual representations of the situated contexts of the research projects that consist of a series of nested concentric circles. At the core of each DE shown in section 4 is one of the five additional projects. The general structure for each DE moves outwards from: *the project*, to *public context*, and ends with the *governance context,* however given the individual nature of these projects reviewed here there are some differences (e.g., The Future Machine and Rider Spoke). The specificities 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 – hence our decision to follow the discussion of Cat Royale in utilizing this approach – in §4.1, we develop criticisms of the diagrammatic ecologies approach. Specifically, diagrammatic ecologies are limited since they cannot be used to visually represent the temporality of these projects and their iterability in a clear way. As such, these models can provide limited insights on the structure of a dynamic archive.[2]

§2.1. When the Future Comes

§2.1.1. 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).

[2] Each project overview contains a section titled, Commentary on the Diagrammatic Ecology, with each individual DE presented. In section 4 we have also chosen to collate the five DEs together for comparative reference.

Publicly funded by a series of academic and cultural institutions and consortia[3], 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 (Jacobs et al., 2023, 4). The Future Machine, therefore, is designed to be an arbiter of hope amidst discourses of existential threat.

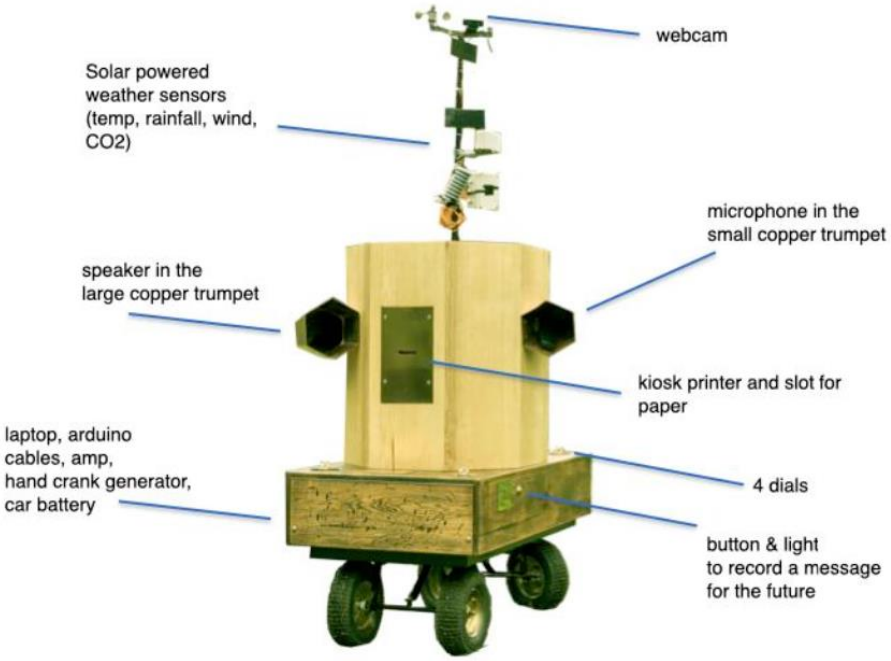
The Future Machine itself, as a material object, is "a large sculptural artwork" that is interactive (Jacobs et al., 2023, 5). Participants are invited to accompany The Future Machine on its annual travels to five locations around the United Kingdom[4] as key seasonal thresholds are reached (e.g., the 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 stored both for longitudinal analysis as the data accumulates, but also to produce music generated by the digital system housed within the octagonal ash, oak, and brass exterior from pre-selected layered tracks (Jacobs et al., 10).

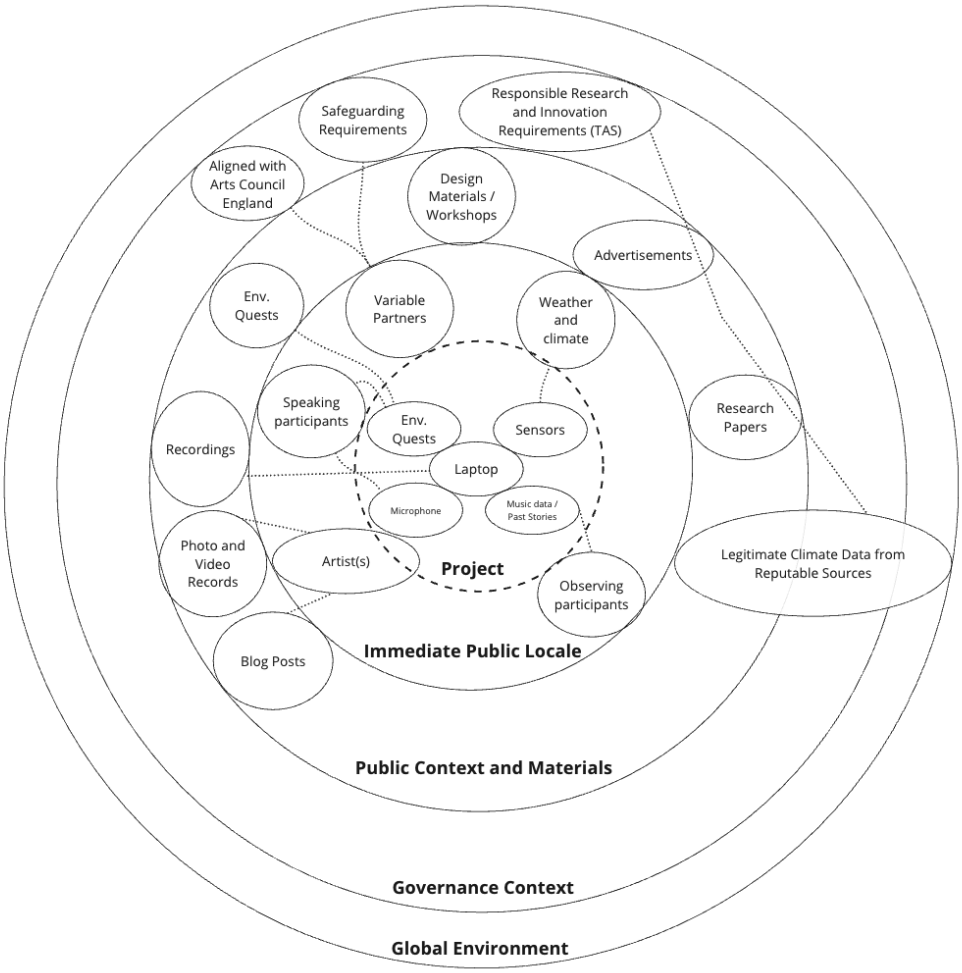
---

[3] "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).

[4] 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. Commentary on the Diagrammatic Ecology (Static Archive/Stakeholders)

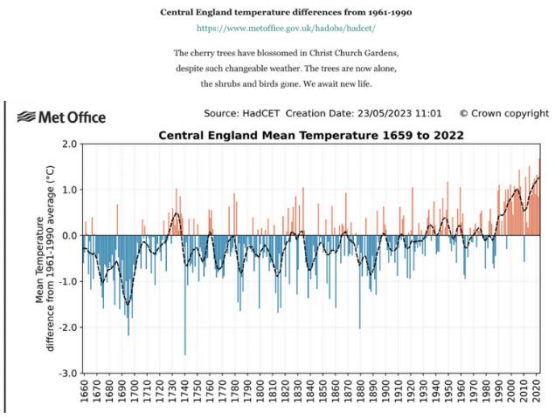


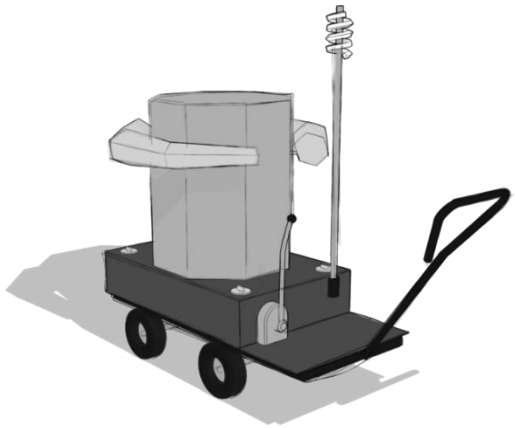
The static archive, or diagrammatic ecology, for The Future Machine (above) is divided into five strata. The innermost nucleus maps the components of The Future Machine which is constituted by a Windows laptop connected to a car battery, powered by a trickle lever, and outputs instructions to the kiosk 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[5] 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 (for more detail see section 4 on authenticity).

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.





[5] 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 (UK Research and Innovation). 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.

§2.2. Embodied Trust in Dance

§2.2.1. Textual Overview

Embodied Trust in Dance is the first of two Trust in Autonomous Systems (TAS) Hub projects included in this report. 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 with the Embodied Trust in Dance team, in which project participants discussed the work, its findings, implications, and the processes involved with the CRADLE team.

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' [tas.ac.uk/research-projects-2023-24/embodied-trust-in-tas-robots-dance-different-bodies/]. 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' [tas.ac.uk/research-projects-2023-24/embodied-trust-in-tas-robots-dance-different-bodies/]. 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' [tas.ac.uk/research-projects-2023-24/embodied-trust-in-tas-robots-dance-different-bodies/]. 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’?

A fifth workshop reviewed and analysed recorded data from previous workshops.

Development and division of activities: Workshops 1-5

Embodied Trust in Dance was not a single event or performance, but a series of workshops which moved from stakeholder, equipment, and data collection familiarization, towards a focused series of recordings based on interactions with the ‘Franka Arm’ robot as the specific device chosen to facilitate robot-dancer scenarios and their eventual analysis.

Workshop	Activities
1	Introductions to stakeholders, to robots, Boston Robotics equipment, familiarisation with sensory and motion capture devices.
2	As above; continued practice and familiarization with stakeholders and technologies, teambuilding.
3	Narrowing to selection of telepresence robots and table mounted Franka arm industrial robots; development of structured scenarios of experiments, documentation on video using multiple cameras including 360-degree motion capture. A greater emphasis on documentation and a focus on robot arms as being the most aesthetically interesting devices to work with.
4	Exclusively dancing with Franka industrial robot arms. Introduction of more scenarios, documentation, and consideration of questions around dance aesthetics and RAI.
5	Data workshop with dancers and researchers. Review of all recorded material in order to choose a selection of robot-dancer interactions for deeper analysis. In depth analysis of 10+ recorded scenarios with Franka Arm robot.

Workshops 3 & 4 Setup

From the perspective of the camera, two dancers – Kat and Welly – were positioned at either end of a rectangular table/secure platform. The table supported a robotic arm towards its right side, facing one of the dancers. The other dancer – positioned to the left of the table – operated the robot.

The dancers worked through various scenarios of interaction between dancers and robots including: one robot physically manipulated by a dancer and 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 X where dancers are observed by researchers with "red buttons".

Physical challenges such as the 'locking out' of the arm – which occurred due to over-manipulation, part of the process of learning to control the robots – provided opportunities for assessing implications for trustworthiness, while simulated limitations – such as the deliberate eye-closing of the dancer, enhanced a requisite sense of trust.

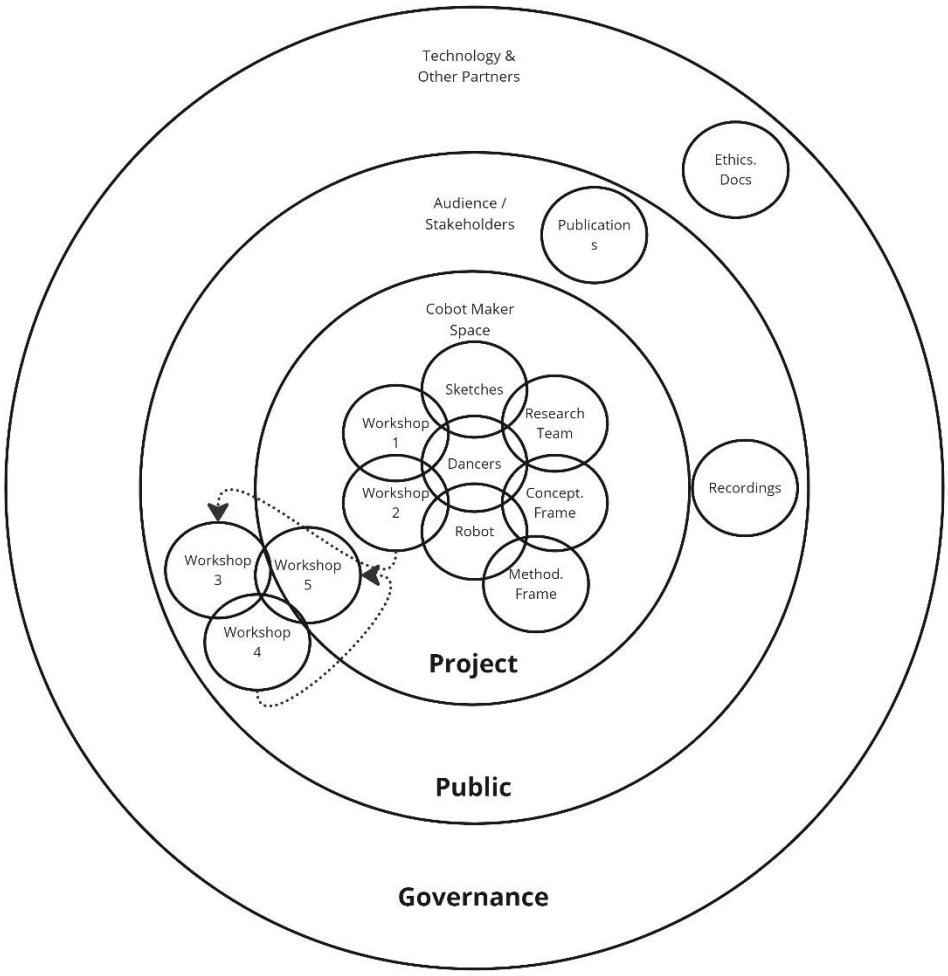
The team discussed the concept of dramaturgy; that everyone played some kind of role and as such, reading elements of trust, control, power, and exchange could be extended beyond the immediate actors. Concurrent themes of vulnerability, autonomy, power, were cited as

crucial in the context of disability. The choice of limb robots was questioned in terms of it being a deliberate one, given the connection with the physical disability of the dancers. Overall, it was suggested that interactions with the robot had the ability to challenge norms, both through human and robot participants.

In total, some 50 experimental scenario recordings with the Franka Arms were captured with audio & video. These documented training sessions, tests, solo dances and duets, and team discussions, divided across 10-20 use case scenarios, with individual videos ranging from 30 seconds to 6 minutes in terms of view time. Many of the experimental scenarios were simultaneously recorded by up to 4 cameras from different viewpoints – including head-mounted cameras, wide angles and 360-degree video. The Embodied Trust through Dance team described a 'triage' process of selecting interesting material from this database, together with a supporting rationale. The team would then vote on these, again, critically assessing why they came to the conclusions that they did. In addition to the archived interactional footage central to the project, documentation of the above processes was captured, along with training, and health and safety data. These resources provide a meta-level description of the work.

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

If TAS dance were to be curated as a static archive, we may envisage the dancers as occupying central exploratory roles in the project space (the cobot maker space); workshops one and two, including the research team, the conceptual and methodological frames, sketches, and of course the robots, 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 particular 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.

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

§2.3.1. Textual Overview

The second 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’ (https://tas.ac.uk/research-projects-2022-23/tarics/).

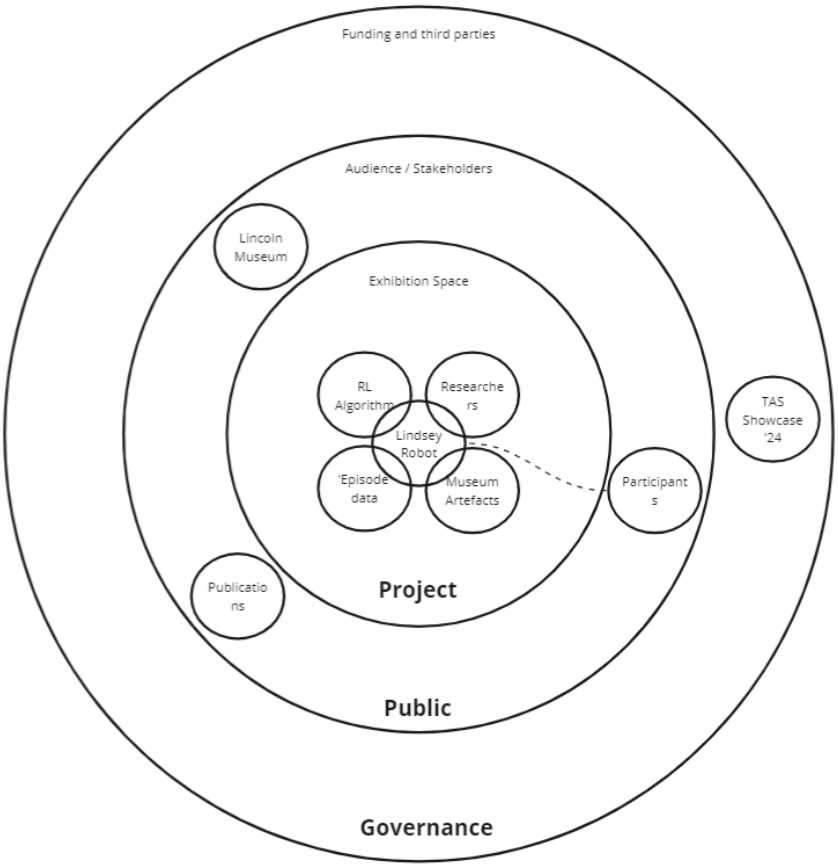
REF *LINDSEY' - TARICS tour guide robot, https://tas.ac.uk/research-projects-2022-23/tarics/*

In this project a tour guide robot - ‘LINDSEY’ – (an autonomous museum guide robot) was deployed to Lincoln Museumto 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 & Hanheide, 2022).

§2.3.2. Commentary on the 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.

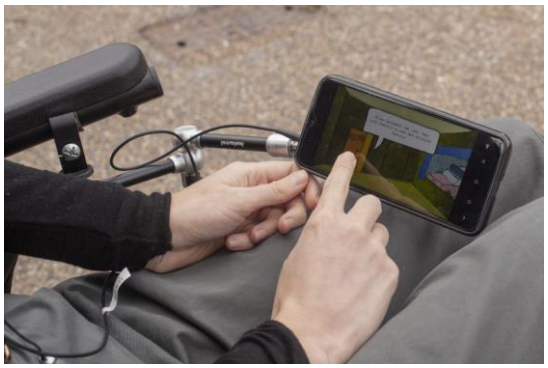
§2.4. Rider Spoke

§2.4.1. 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). Nevertheless, the project has seen continued interest and staying power, evidenced by the repeated international iterations of the artwork.[6]

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.



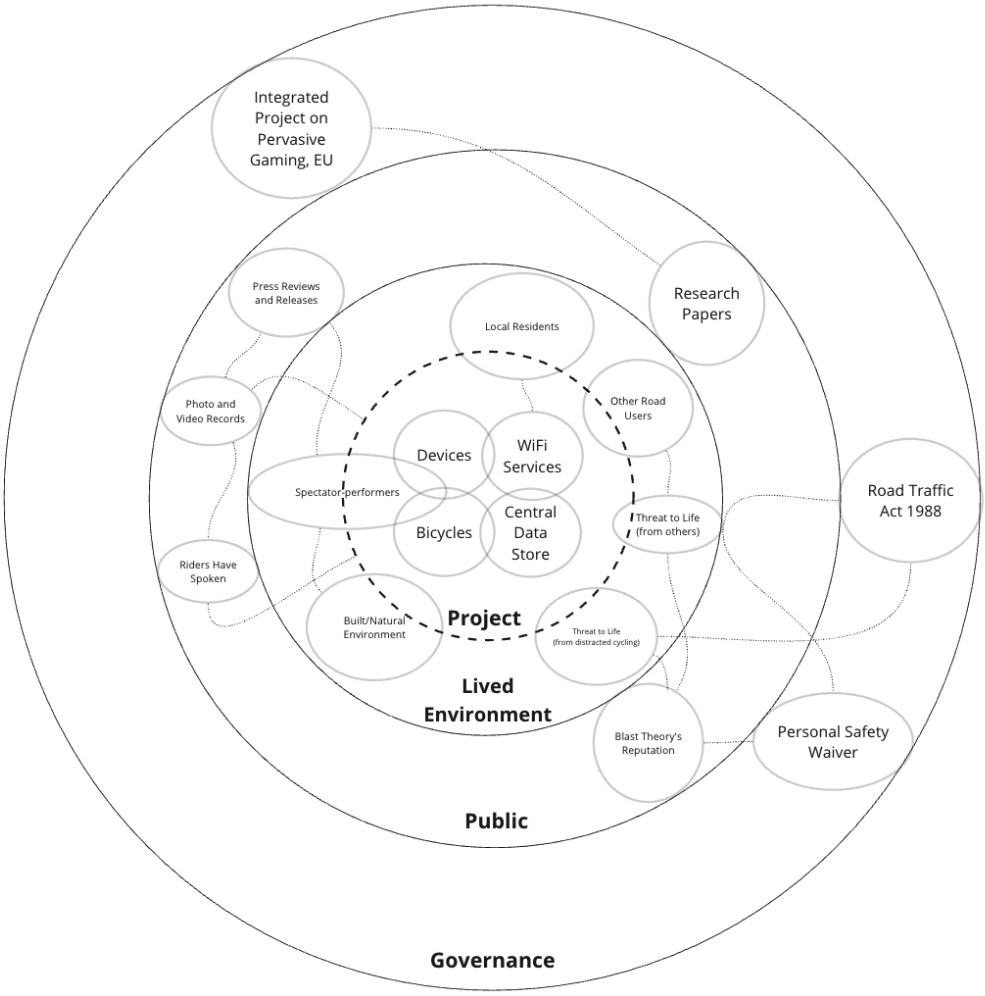
In one sense, the aim of the project is to elicit from these participants personal stories or intimate confessions (Chamberlain et al., 2011, 4, 6; Quigley 2016, 97). In another sense, as Blast Theory's Matt Adams explained in 2007, Rider Spoke challenges the traditional bounds of theatre and artwork 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 Karen Quigley (2016), we refer to the individual players as *spectator-performers*.

As will be discussed in later sections of this report, 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., 2008).

[6] 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.).

Second, the project is unique as regards its very 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.

§2.4.2. 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 (§2.1.2) 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 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 more key concerns discussed in section 3.4 which concerns the threat to life sustained from injury and 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 that have been produced because of Rider Spoke is the archiving project, Riders Have Spoken, developed by Gabriella Giannachi, Duncan Rowland, Steve Benford, and Dominic Price (discussed at length in section on archiving)

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. As will be discussed in section 3, 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);

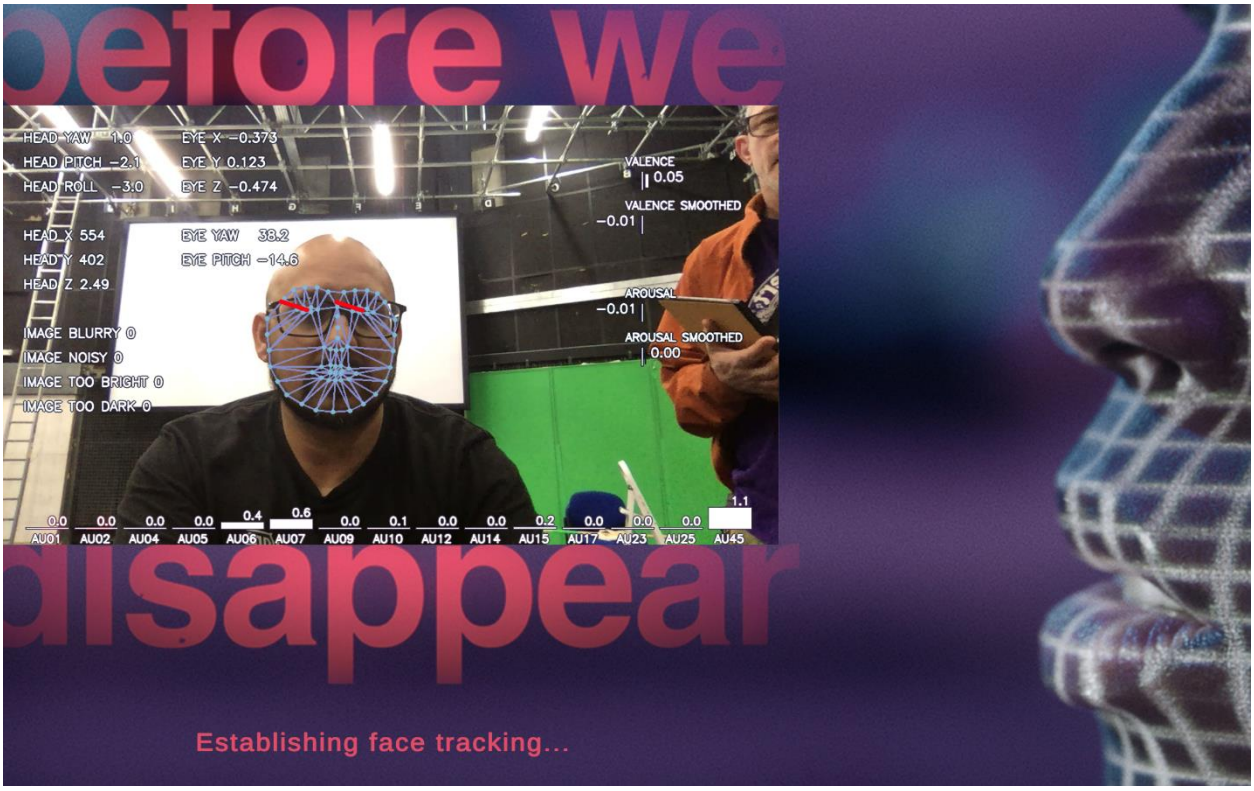
a phenomenon that Waern (2009) discusses in her chapter, “Ethics of Participatory Art”, as being all too common.

§2.5. ‘Before We Disappear’

§2.5.1. 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) BSocial 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. Implications for Dynamic Archiving

We have chosen to structure this section *thematically*. That is, each sub-heading refers to one of the seven questions used to guide discussion in the Second Workshop, with sustained discussion of each of the five projects in relation to these questions below.

§3.1. Creativity and Authenticity

§3.1.1. When the Future Comes

It can be said that the When the Future Comes Collective, in the production and performance of The Future Machine, seek to use creative endeavours to instil within participants and audiences a feeling of connection to the world to foreground the temporality of human experience. They aim to achieve this through the use of three modes: environmental (the earth itself), biological (the flora and fauna), and the cultural (human-

human interactions within cultures, politics, and so on) The Future Machine is the product of the artists' creative endeavour to circumvent the human-nature dichotomy (Jacobs et al., 2023, 4), using technology as a means to 're-connect' human beings with the changing seasons and the rhythms of the world that contemporary digital capitalism seems to have eclipsed. One participant, recorded in Jacobs et al., (2023, 10), reflected that the rituals of Future Machine are "a really lovely way of inviting people to come back together", noting that Future Machine functions more as a way of reminding us of these natural processes and encouraging a respect and enjoyment of them rather than as a spectacle that deploys the natural world as a mere gimmick.

The production of a technological artefact that is designed to reconnect audiences with the natural world as a means of facilitating experience, rather than mediating it as such (as with mobile phones, for example) is reminiscent of Heidegger's analysis of technology and his writings concerning *authenticity*. For Heidegger, the essence of technology is its function as a frame; that is to say, the essence of technology is that it mediates human relation to the world in such a way that all that exists comes to be understood purely as "standing reserve" for human intention and action rather than existing for itself. The Future Machine, then, as a technologic artefact designed explicitly with the looming irreversible climate thresholds and the fast-tech culture in which we live in, is designed to contest the means-ends relation to the world that techno-capitalism has accentuated.

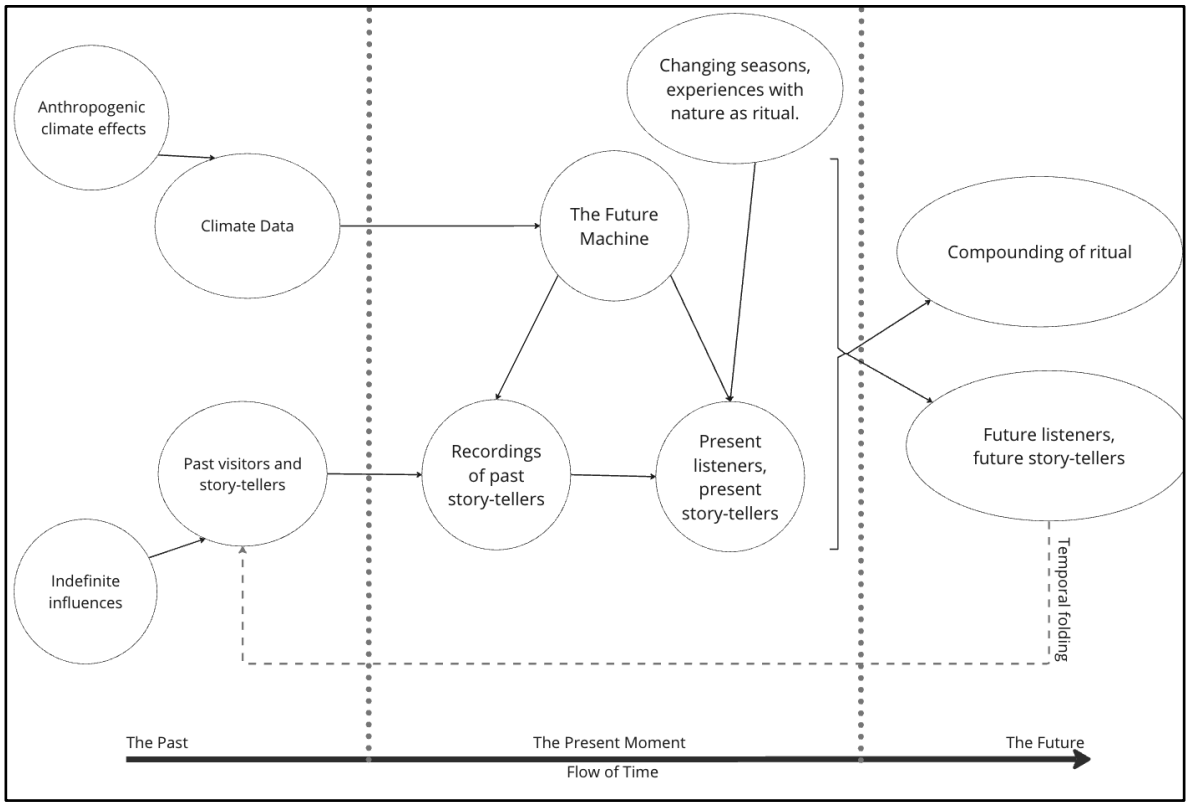
The Future Machine is not an instrumental technological product; When the Future Comes is not in the business of selling one-off experiences but in motivating participants to reflect on the embedded nature of human existence as part of the natural world. We might say, then, that Future Machine *facilitates* rather than *produces* a field of connection wherein present audience members listen to the recordings from the past and speak to the future. This experience is best summarised as a flat temporality wherein the artwork itself is constituted in the relation to future people in the specific locale (e.g., Oxford, Cumbria, Nottingham, Somerset, and London). This experience is the artwork; not the stories (Future Quests), not The Future Machine itself, but that *experiential relation that directs the audience to reflect on the entanglement of culture and nature and our collective responsibility for the environment*.

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 is 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* for conceptualising the project that foregrounds this dynamism as a departure from the diagrammatic ecology presented in §2.1.2. A dynamic archive will need to be able to reflect this embodied, temporal nature rather than freezing time and interaction out of the picture.



§3.1.2. Embodied Trust in Dance

Creativity and authenticity are synonymous with the original objectives of the 'Embodied trust in dance' project. Due to the unique way in which dance was performed with a robotic arm as partner and as intermediary though, these values perhaps take on a new, orthogonal quality, rather than being naturally complimentary; the intersection of creativity and authenticity is primarily observed in first hand stakeholder accounts of interactions and through video footage.

The wider TAS dance team revealed that the experience of 'dramaturgy' was one of authenticity, while anecdotally (the dancers were not themselves present in the later meeting), the proposition – particularly through selection of a robot in the form of an arm – had clear transferable implications for others who might want to explore dancing from the position of disability.

Authenticity relating to the experiences of the team members was evidenced in the team membership; the research agenda was directed by interaction design experts and ex-professional dancers (including one with a physical disability) whilst the performers

included expert dancers with physical disabilities. Authenticity is also evidenced in the video footage (revealing emotional reactions of dancers and audience, mistakes and so forth). The breadth of data captured, from formative ideation to more formalised scenarios, ensured that original performances were retained while also demonstrating a sense of progression.

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

Authenticity can be discussed in terms of encouraging interactive engagement with the users. In this sense authenticity can be said to be demonstrated in terms of the designed focus on maximising interactive engagement between the robot and its users, via a reinforcement learning (RL) algorithm. As a robot tour guide, LINDSEY is inherently dynamic in this regard, and thus capable of harnessing authentic interactions that extend beyond prescribed experiences. This is seen in the 'state and action specification' logic, informing the RL algorithm: there is a real consideration for the authenticity of museum interaction as expressed by a naturally emergent sequence of events, and progression through physical artefact.

Elements of creativity are demonstrated via the confirmation of the hypothesis – that the '…learned policy can…lead users to keep a more sustained engagement during the interactions in the museum…' [p6].

> *'…an increase of 22.8% [user engagement] 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'*

These results might suggest that user engagement and probability of tour completion serve as proxies for creativity inherent in the research design in this deployment context.

§3.1.4. Rider Spoke

Rider Spoke, as previously mentioned, was the product of an EU funded project exploring the potential of pervasive gaming. Blast Theory's Matt Adams cited the phenomenon of identity fragmentation as a key impetus for the project (Lavendar et al., 2007). Identity fragmentation, of course, refers to the experience of existing in and across multiple 'spaces' or locales: homes, workplaces, social media etc. Though this phenomenon is not directly a result of the proliferation of personal technologies (see Hardt and Negri 2000, page), the

digital boom has intensified the locales that we traverse in day-to-day life. There is a tendency for individuals in online spaces to fabricate a new identity for themselves, either in the pursuit of anonymity (to conceal one's thoughts and feelings whilst expressing them) or else to escape from the limitations of one's own life. Rider Spoke, in many ways, toys with this ""blurring between fiction and reality" (Lavendar et al., 2007), making Rider Spoke a distinctive case study for examining the relation between creativity and authenticity.

The spectator-performers, when they set off within the game of Rider Spoke, are free (within limits) to cycle around the city and record their own stories. These stories are, of course, prompted by Ju Row Farr's voiceover. The stories, therefore, could not be aimless, unprompted, or else wholly plucked from the mind of the player. As Chamberlain et al. (2011, 4) observe, not only must the testimony of the player respond to a particular prompt, but must also "have local relevance, [resonating] with specific locations within the city". The narration ranges from relatively innocuous prompts to more intimately probing points for the participant to reflect upon. Below we have included two such prompts, collected from Chamberlain et al. (2011), to demonstrate the variability in the prompts.

> *I want you to look for a flat or a house and find a window that you would want to go through. I want you to stare into that window and tell me what you see and tell me why you want to go through that window.*
>
> *Please will you tell me about your father. You might want to pick a particular time in your father's life or in your life. Freeze that moment and tell me about your dad: what they looked like, how they spoke and what they meant to you. And while you think about this I want you to find a place in the city that your father would like. Once you have found it stop there and record your message about your father at that moment in time.*

The testimony given by the players can be understood as needing to fulfil two criteria. First, the testimony must be an engaging, emotional, or unique response to the prompt. This is evident from the ways in which the artists and project team rank the recordings (Chamberlain et al., 2011, 16; see also §2.6). Second, and perhaps inseparable from the first, the player ought to be reflecting on genuine experiences. Rider Spoke can be understood as an unspoken challenge to participants: *when shrouded in anonymity how honest will they be about their life?* Of course, there is the possibility that one's recording will be heard by others, but when testimony is disentangled from identity, and thus from reputation or social standing, how authentic can we be? Possibilities of approval and rejection fade away, and the player is left standing at the seafront, in a side street, or in a churchyard, alone. As Tandavanitj (2023) observes, Blast Theory's "mantra" has often involved leaving the audience "responsible for what is said".

This is the crux of what has been analysed as the tension between the participant's role as a performer (giving testimony) and as a spectator (exploring the testimonies of others). Some participants engaged faithfully with the frame of the game, whilst others slipped (whether consciously or unconsciously) into a character (Chamberlain et al. 2011; Quigley 2016). Participants' reflections on the experience -- recorded in Chamberlain et al. (2011) -- show that for some, honest reflection was fundamental to 'playing the game', so much so that "some even cried while recounting [their testimonies]" (2011, 9). An in-game testimony, reported by Quigley, captures this same emotionality, which she analyses through the lens of Roland Barthes' concept of *the grain of one's voice*. This woman's voice, in addition to the content (reflection on the possibility of the loss of her husband) is what captures Quigley's attention (Quigley 2016, 185). The raw and honest confessions that Rider Spoke elicited from some players is precisely the kind of testimony that Rider Spoke sought to capture and leave for future players to find (as above).

Other players, however, engaged more playfully with the pervasive experience and toyed with the prospects that anonymity afforded them. This inauthenticity, or as Quigley describes it, *playful antagonism within the frame*, is perhaps the kind of instinct that Rider Spoke sought to challenge, or put under the microscope: *why is there an instinctive pull to misrepresent oneself when the opportunity presents itself?* In Quigley's case, her character -- Penelope Coffinfeather -- emerged as she sought to follow the aphorism, *never let the truth get in the way of a good story* (2016, 90-91, 100). Her antagonism, as she describes it, was arrested when she came across the intimate confession described above; through an encounter with authenticity-despite-anonymity, Quigley's own inauthenticity-because-of-anonymity was challenged, questioned, and subject to radical change from character to confessor.

Was Rider Spoke designed to engineer this kind of encounter with the tendency for inauthenticity when we are presented with anonymity? Is this the core tension that Blast Theory sought to play with, following their *modus operandi* of forcing audiences to reconsider the assumptions, instincts, and habits that we have developed in a fast-changing technological world? These questions add to the complexity of this artwork's design.

§3.1.5. Before we Disappear

Currently access is not available for the training data used to generate values for arousal and valence used by b-Social. It is also not currently possible to independently assess the accuracy of the facial tracking system. Before We Disappear is intended to be released

through the Steam distribution platform. The film is embedded with b-social into a single package.

A number of accessible websites detail the functionality of b-Social and how it has been used for this and other applications.

https://www.horizon.ac.uk/adaptive-interactive-movies-partnership-working-a-case-study/

https://www.youtube.com/watch?v=mCnumsXKvKY

https://www.youtube.com/watch?v=G245XO2jZK0

The creative practice of the filmmaker is intrinsic to the final production, though this can be broken down into both the film content and how the narrative is modified by audience behaviour. The metrics of valence, arousal, and shared gaze as output from b-social analysis are intended to be as "objective" as possible – derived from the machine learning training sets, and an academic appraisal of the relationship between facial behaviours and internal cognitive or emotional states. The narrative direction taken according to inferred emotional states may attempt to creatively shape the future emotional states of audience members.

§3.2. Creativity and Responsibility

§3.2.1. When the Future Comes

As regards the relationship between creativity and responsibility, When the Future Comes offers a profound case study. As noted in §3.1.1, When the Future Comes' project, The Future Machine, is designed to function as an experience facilitator, fostering a reflection on the intra-woven nature of human beings, the built environment, and the natural environment. The artists note that Future Machine, as an art/research project, is intended to bridge two existing polarised streams for artworks that engage the audience in climate-oriented emotionality and reflection, either using only climate data and anthropocentric narratives or else eschewing this data in favour of focusing solely on the qualitative or phenomenological experience of living at this pivotal stage in human history (Jacobs et al., 2023, 4).

The When the Future Comes Collective makes use of creative techniques, traditional artisanry, 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, as modelled in §3.1.2: *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 is not about listening to the past; it is about speaking for the future*. It is not about objectifying the past as an aesthetic artefact; it is about relating to the future people *qua* people in the world with their own stories and affectivity.

§3.2.2. Embodied Trust in Dance

In terms of Embodied Trust in Dance, the intersection of creativity and responsibility is again primarily observed through reflections of stakeholders, especially the dancers. It was felt that to 'trust' a robot as a dancer should be analogous to trusting a human dance partner, with important similarities and differences.

Trust with the robot arm was built through a number of 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.

Again, the ethics documentation evidences decisions taken in the design and deployment of research that most overtly link AI applications in this context and AI concerns relating to responsibility.

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

A methodology using '...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'.

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)' (p4).

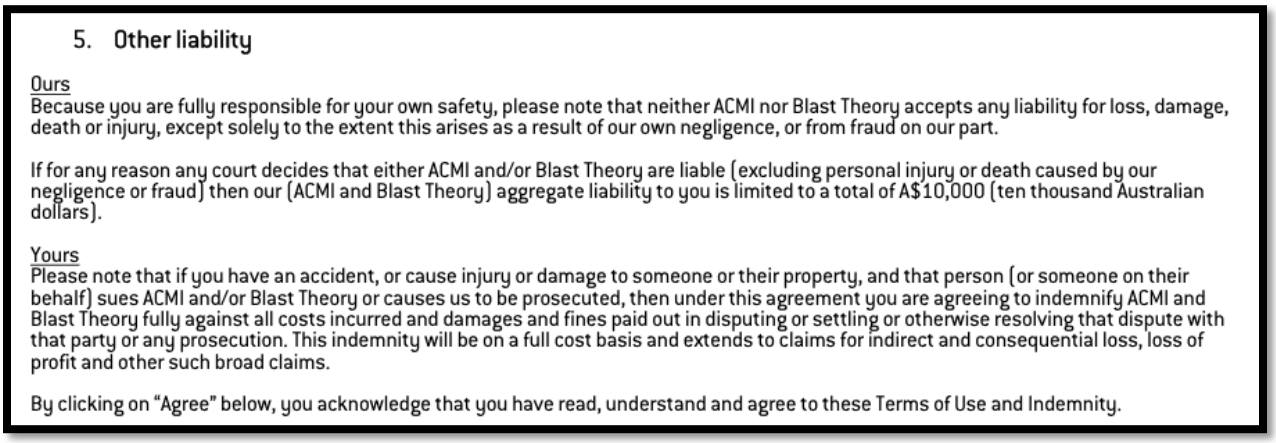
§3.2.4. Rider Spoke

In terms of the relationship between creativity and responsibility, Rider Spoke presents a unique case a the project 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). Whilst this does of course affect the participant experience – "I didn't [...] feel liberated rather in difficulty" (Lavendar et al., 2007, 11) – it also presents a responsibility issue, one that is summarised by Leo 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, that when risk is involved, consent is often – as is well known – a complex matter (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" (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, 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 have 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 therefore 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 experienced a radical shift in their responsibility as a participant.

§3.2.5. Before we Disappear

The relationship between responsibility and creativity in "Before We Disappear" has two constituents. First, any filmmaker has a degree of responsibility to produce films which meet the funder's ethical (and legal) expectations. This is especially true for films produced within the research context and public accountability of universities and their funding bodies. There may also be some responsibility in filmmaking in contentious 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 again an open question.

Secondarily, 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. However, this may be an unrealistic expectation for a commercial product. Responsibility requirements may be met through risk assessment and ethics approval by an authoritative body. In this specific case, it is hard to conceive of possible harms directly arising as a result of the facial recognition system. However, an understanding of how the 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.

§3.3. Further Insights into the Debate Concerning Responsibility

§3.3.1. When the Future Comes

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., (2023) are explicit about their desire for Future Machine to raise "awareness of the full *environmental* costs of HCI, the

internet of things and artificial intelligence" (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." (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" (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 (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 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.

§3.3.2. Embodied Trust in Dance

Embodied trust in dance contributes to a tangible understanding of practical, RAI beyond the parameters of the immediate project.

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.

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

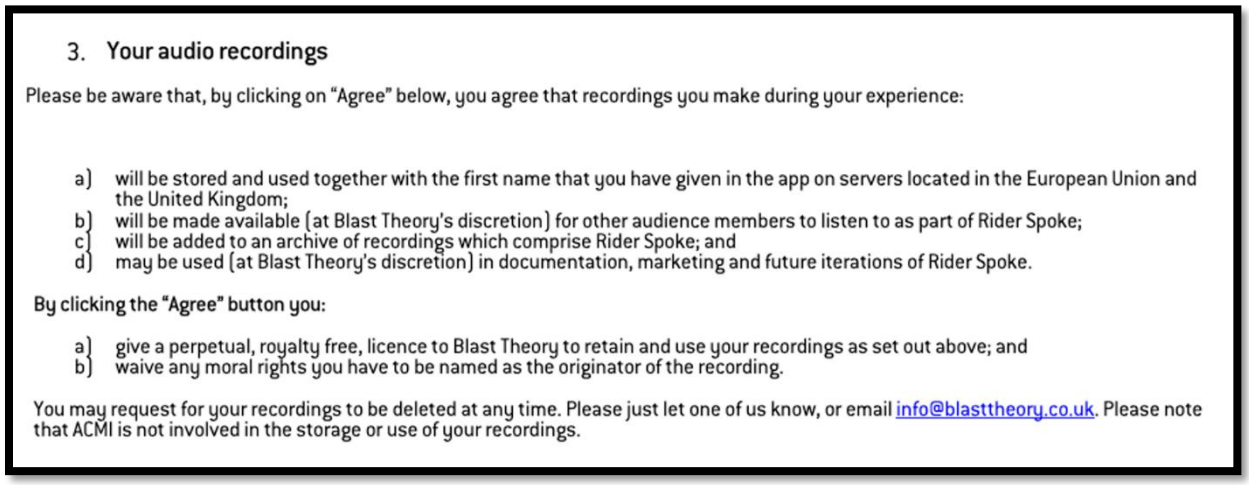
'....the aim of increasing the accessibility of the cultural experience in museums for people with learning disabilities, we plan to explore ways in which the proposed learning framework can enable the robot to behave in a more inclusive way and take into account the different user's characteristics and preferences' p7 (Paper: Learning on the job)'.

Authors took an 'experimental validation' approach, ultimately enabling conclusions to be drawn that '...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' (Paper: Learninng on the job, 1).

§3.3.4. Rider Spoke

As a pervasive gaming experience that will involve the recording and discovery of audio messages left throughout the immediate built environment by a collection of strangers, there would seem to be a duty for Blast Theory to screen recordings against a set of criteria that go beyond the categorisation system of recordings used to determine which recordings are 'left in place'. For example, the categorisation system features five categories, ranging from "I can't hear this/would never want to listen to this" to "I would love to listen to this" (Chamberlain et al., 2011, 16) though there is no mention made of what we might call *sociopolitical* or *equality, diversity, and inclusion* (EDI) criteria. To put this another way, *were recordings screened against a framework (analogous to the Equality Act 2010) that considered the recordings independently of the artwork?*

For example, whilst category 1 (the lowest rank) states that these recordings are to be discounted on account of these "criticising the artwork" or being a "repeated answer", category 2 pertains to observations that are "boring", "completely indistinct or obvious". One such example is, "some parts of the city are quite rough"; is this to be excluded on its banality or alone, or as part of perhaps a larger operation of classism within society, or as a perceived political comment about the kinds of people who live in said area? The literature on Rider Spoke does not mention socio-political responsibility to manage or handle recordings that reflect prejudices (e.g., on the basis of gender, race, class, or immigration status). It is possible that these points were handled in documentation not publicly available; the Terms of Use and Indemnity policy does not make explicit mention of the content of the recordings made, only the ownership rights.



3. Your audio recordings

Please be aware that, by clicking on "Agree" below, you agree that recordings you make during your experience:

a) will be stored and used together with the first name that you have given in the app on servers located in the European Union and the United Kingdom;
b) will be made available (at Blast Theory's discretion) for other audience members to listen to as part of Rider Spoke;
c) will be added to an archive of recordings which comprise Rider Spoke; and
d) may be used (at Blast Theory's discretion) in documentation, marketing and future iterations of Rider Spoke.

By clicking the "Agree" button you:

a) give a perpetual, royalty free, licence to Blast Theory to retain and use your recordings as set out above; and
b) waive any moral rights you have to be named as the originator of the recording.

You may request for your recordings to be deleted at any time. Please just let one of us know, or email info@blasttheory.co.uk. Please note that ACMI is not involved in the storage or use of your recordings.

Participants also 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 (2008, 39) claimed that "being mugged" is a "relatively exotic" and "unlikely" risk, the likelihood of harassment and violence against women ought not be ignored by the artists and academics contributing to the development and implementation of Rider Spoke as unfounded nor "exotic" since a majority (58%) of the participants in the first iteration were women. Moreover, 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? This possibility is discussed in §3.4.4.

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.

§3.3.5. Before we Disappear

A significant concern regarding responsibility with "Before we Disappear" regards the use and transparency 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?

§3.4. Replicability of the AI Model?

§3.4.1. When the Future Comes

The When the Future Comes Collective is committed to the promulgation of the myth of 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, 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.

Insofar as the digital system of 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.

§3.4.2. Embodied Trust in Dance

As a lab-based work, Embodied Trust in Dance offers a model for replicating research based on a well-controlled environment in which certain variables and can appear static (for example, the robot, the configurations of dancers). A significant part of the activities and data collection in Workshop 4 was recording robot arm movements. These recordings were recordable and replayable on the robot arms. The software designed to do the recording was developed by a previous researcher at the University of Nottingham and worked with (but was separate from) the robot arm. It may be possible to independently assess the accuracy of recordings (spatially and temporally), however to date this has not been completed. For instance, some of the recordings are made as a series of discrete points in space and time. it is conceivable that the robot arm moves *differently* between the points (for example, takes a shorter route through space, moves much faster) than the dancers did when recording. An analysis of the video recording could potentially reveal such characteristics of the robot control system. Analysis could also point toward potential risks for dancers and robots.

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

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 of a priori knowledge of known dimensions.

§3.4.4. Rider Spoke

Although Rider Spoke does not itself utilise Artificial Intelligence, the project does incorporate 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.

§3.4.5. Before we Disappear

The AI model used, 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 as a APK or similar, it may not be clear to filmmakers (and other interaction designers etc.) how they would incorporate b-social into their projects.

§3.5. Is the AI Model Aligned to RAI Principles?

§3.5.1. When the Future Comes

When the Future Comes produced Future Machine as part of the Trustworthy Autonomous Systems Hub's 'Responsible Innovation and Sustainability in the Digital Arts in an age of AI: Responsibility Reimagined in the age of AI' project. As such, Future Machine was held to three fundamental aims. These concern ethical acceptability, sustainability, and societal desirability (Trustworthy Autonomous Systems Hub, 2020).

Previous sections on Future Machine have explicitly engaged with the ways in which from its original conception to its production processes and continued performance, Future Machine is unequivocally designed to challenge unsustainable practices both at the individual level and the level of technologic infrastructure and culture. As regards the social desirability of the project, When the Future Comes continues to engage the public's fears and hopes, dreams and concerns about the future of human life on planet Earth, and thus operates to provide a sense of calm and returns us to the rhythms of the natural world.

§3.5.2. Embodied Trust in Dance

As a TAS hub project, alignment with RAI principles is explicit and intentional in Embodied Trust through Dance. The project is explicitly orientated toward the advancement of robotics with inclusive applications for dancers with disabilities, actively involving stakeholders from this community during and beyond the dance workshops. In this sense, the 'no decision about me, without me' principle is demonstrably adhered to.

An interesting theme emerged from Embodied Trust in Dance workshops revealing that there may be tensions between the practise of performance, in this case, the creative desires of dancers to behave authentically and take their own calculated risks, and the risk assessment by technicians and other team members who are required to protect the dancers and robots from physical harm.

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

Explicitly, the work is part of the TAS hub project and aims to advancing the experience of 'trustworthiness' among stakeholders, not least potential participants in the public space with additional learning needs and barriers to conventional museum engagement. In terms of alignment to RAI models, it aims to encourage trustworthiness by encouraging users to produce 'engagement values' via collaboration with the LINDSAY robot [11].

§3.5.4. Rider Spoke

Originally designed and implemented as part of the Integrated Project on Pervasive Gaming funded by the European Union, the project, conceivably aligns with Responsible Artificial Intelligence principles. First and foremost, though the project revolves around the recording of participant's data, the data itself "is valueless" (Lavendar et al., 16). That is to say, though it involves the tracking of location, this is done in such a way that the records only make sense relative to the framing of the game, and the recordings themselves are "not commercial information" (16).

§3.6. What Might a Dynamic Archive Look Like in These Instances?

§3.6.1. When the Future Comes

Drawing on the reports from audience members who connected with Future Machine as a friendly technology , a dynamic archive for Future Machine could be constructed on the premise that the archive is a means of conversing with Future Machine itself; that is to say, the dynamic archive would be an analogue for 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 "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 Future Machine's longitudinal study into a matter of minutes in different ways. Similarly, timelapses of the photographs taken by 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 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 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, the dynamic archive, or the pseudo-machine 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.

§3.6.2. Embodied Trust in Dance

Conceiving 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.

Conceiving a dynamic archive as a trajectory would emphasise individual stakeholder perspectives e.g. dancers, academics, audience etc. over time, more so than the concentric circle model. Using the trajectories approach 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 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.

§3.6.3. 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.

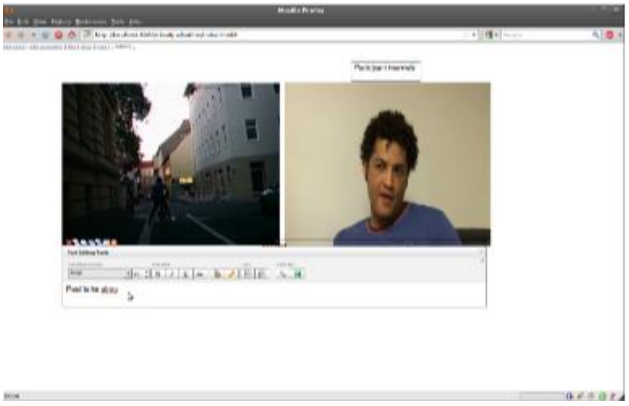
Imagining TARICS as a trajectory or 'affective cartography', we noted that this was already evident in some of the interaction frameworks present in the design of the LINDSAY robot. This is elaborated further in section 4.1

§3.6.4. 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.

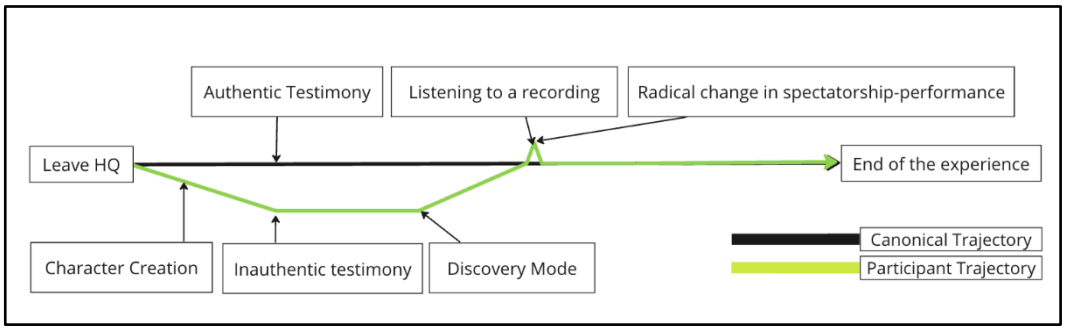
The archive produced as the Riders Have Spoken Project was presented as a hand drawn city scape, with lights on in some of the windows. Users simply had to click on a lit window in order to hear an individual's response to a prompt, whether it is about their father, their deepest hopes, or their first-time 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 because 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 synchronise 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 because 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.



§3.6.5. Before we Disappear

This work was developed within Horizon, 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 also 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 https://www.horizon.ac.uk/adaptive-interactive-movies-partnership-working-a-case-study/

§3.7. How Might These Projects Relate to an Open AI Dataset?

As a precursor to ideation, we discussed this final question about each of our projects by first addressing definitions of 'datasets'. In the generic sense of 'all data' pertaining to each of the projects discussed this has implications for 'How we might make it [archive] usable for different demographics?' More specifically, in the sense of 'open datasets', we use the UKRI description of research that is 'transparent and easily scrutinized, helping to increase public trust'; 'easy to re-use and build upon'; and 'collaborative and efficient', as indicators of required qualities https://www.ukri.org/manage-your-award/publishing-your-research-findings/making-your-research-data-open/

§3.7.1. When the Future Comes

Useful for a paradigm case of an extensive longitudinal study.

Inventive use of climate data to manifest emotionality

§3.7.2. Embodied Trust in Dance

An open AI dataset based on Embodied Trust in Dance affords reproduction of aspects of the project that are non-linear or linear based on adoption of concentric circle or trajectory models, respectively. This is true of all projects that are archived according to these developing prototypes.

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

Thematic: Focused on building an archive of the artefacts in situ, and thus, how the transcendent values of their use evolve.

Linear: Focussed on building an achieve of use cases over time, and thus, how the prescribed use case changes

§3.7.4. Rider Spoke

Demonstration of innovative use of WiFi technology to design an iterative infrastructure capable of high mobility independently of the geographical specificities of the area. Low data, high output.

§3.7.5. Before we Disappear

An open AI dataset could conceivably include the algorithms or other information to make transparent the models used by b-social. It might also include data recorded while audiences watch the film if explicit permission is sought.

§4. Prototypes of Dynamic Archives

In this section, we introduce speculative designs for prototypes based on prior discussions in workshop 1 and 2. These are fundamentally based on enabling four foundational assumptions of dynamic archive functionality, summarised as follows:

- continue to grow over time;

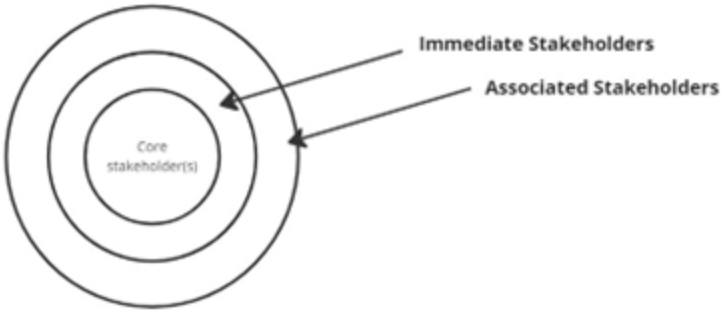
- reconstitute itself in accordance with keywords set by whoever wishes to consult it;
- reconstitute itself by making visible different time-based versions of itself;
- be used curatorially and creatively as a live (generative) archive.

In previous workshops, the mapping of ecosystems in terms of object-interactions (stakeholders, artefacts, etc.,) and trajectories from the researcher perspective came to exemplify two human computer interaction (HCI) approaches for generating dynamic archives. In the first instance, the concentric circles drew heavily on preexisting notions of boundaries and domains of interaction, while in the second, the linear representation of interaction stages emphasised the passage of time and emergent pathways taken by individual participants or artefacts.

We draw on these frameworks, critically examining their affordances when conceptualised as interactive, dynamic archives, capable of becoming aligned to the above assumptions as well as RAI principles more broadly. Ultimately this section concludes with a specific instantiation of a dynamic archive prototype, based on the combined content of the creative AI works reviewed.

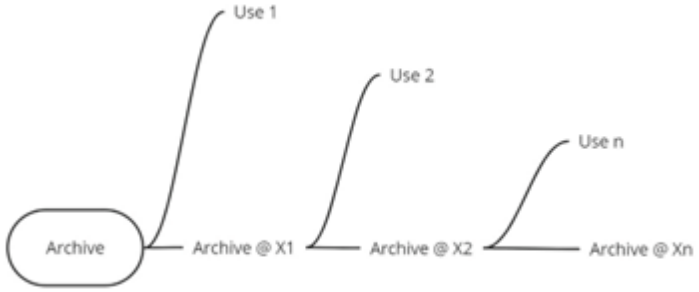
§4.1. From Concentric Circles to Trajectories

All previously outlined projects demonstrate the four assumptions of dynamic archive functionality to various extents, and to illustrate this, we used two frameworks for conceptually representing their components: the concentric circle and trajectory diagrams. In the first instance, the concentric circle has been instrumental in setting out components of a project as a static archive, depicting human and technological actors, as data repositories or interactive touchpoints; crucially within the boundaries of defined stakeholder groups and contexts. In the second instance, the trajectory – 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.

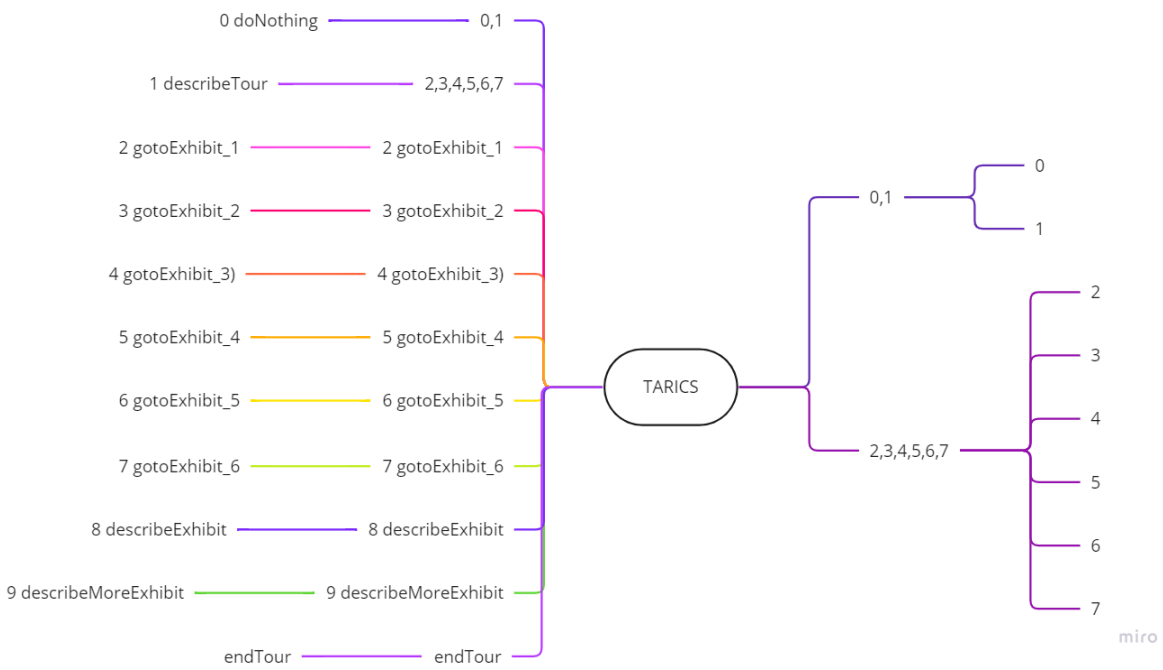


The above concentric circle diagram has been used since WS2 to illustrate AI ecosystems, centred on the primary artefacts, stakeholders, and interaction between them as a core feature. Immediate stakeholders - those present and influencing the interaction(s) such as researchers and technicians, or present 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 (Benford et al.) 2009.

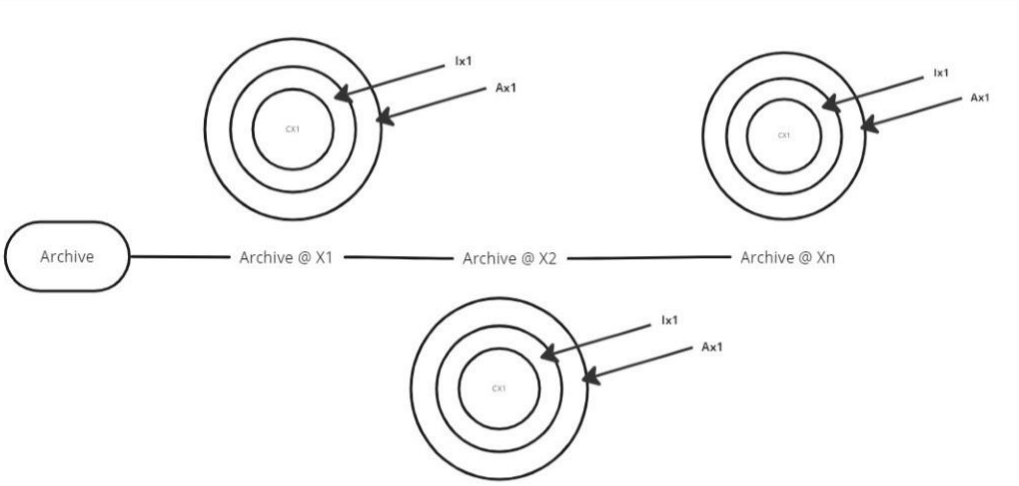
This model exemplifies a static approach to archiving the ecosystem, and while dynamic aspects such as instances of interaction, or transcendent themes such as shared stakeholder values, can be depicted as such, the implication of the model is one of a fixed delineation of boundaries asserted by an assigned 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. 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': 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 [REF]. 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' (REF.PAPER) or progression to 'describeTour', where successors are the sum of interactions with LINDSEY [REF]. Moreover, 'gotoExhibit' might be substantiated by all instances of user response to the exhibits in situ '2-7' [REF].



Finally, we discussed the combination of the concentric circle and trajectory models. In the above illustration, the concept of the trajectory is applied to the concentric circle view of the AI ecosystem, such that multiple trajectories of use cases can be captured for different users at each stage of interaction X_1, X_2... Xn. The inspiration for this work in progress was 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.

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.